Pruscha:

Semiparametric Point Process and Time Series Models for Series of Events

Sonderforschungsbereich 386, Paper 114 (1998)

Online unter: http://epub.ub.uni-muenchen.de/

Semiparametric Point Process and Time Series Models for Series of Events

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 series of events occurring at random times τ_n and carrying further quantitive information ξ_n. Examples are sequences of extrasystoles in ECG-records. We will present two approaches for analyzing such (typically long) sequences (τ_n, ξ_n), $n = 1, 2, \ldots$. (i) A point process model is based on an intensity of the form $\alpha(t) \cdot b_t(\theta)$, $t \geq 0$, with b_t a stochastic intensity of the self-exciting type. (ii) A time series approach is based on a transitional GLM. The conditional expectation of the waiting time $\sigma_{n+1} = \tau_{n+1} - \tau_n$ is set to be $\nu(\tau_n) \cdot h(\eta_n(\theta))$, with h a response function and η_n a regression term. The deterministic functions α and ν, respectively, describe the long-term trend of the process.

Keywords: Semiparametric estimation; Trend function; Penalized least squares; Point process; Intensity process; GLM-based time series; ECG-data

1 Introduction

Series of events can be analyzed in a dual way, namely by establishing

- point process models for the sequence $\tau_1, \tau_2, \ldots$ of occurrence times
- time series models for the sequence $\sigma_1, \sigma_2, \ldots$ of the lengths of time intervals between events ($\sigma_n = \tau_n - \tau_{n-1}$, $\tau_0 = 0$)

(see Cox & Lewis, 1966). In the present paper, we employ

- intensity based point process models of the self-exciting type (Hawkes, 1971)
- GLM-based time series models of the transitional type (Zeger & Qaqish, 1988; Fahrmeir & Tutz, 1994)

In both cases the model equation is factorized into a deterministic trend function and into a stochastic part describing the short-term oscillation. As in many semiparametric problems a major task is to separate the two factors by the estimation procedure. In our situation of one single realization over a long time interval we will need a kind of ergodic behaviour of the process. The trend function is estimated by penalized least squares (l.s.) methods based on aggregated data, gained by an averaging procedure over longer time intervals (preserving enough short-term information). For the parameter of the stochastic component we will then employ likelihood methods, thus reversing the

order of estimation known from partial likelihood approaches in survival time analysis. In addition to the occurrence times τ_n metrically scaled marks ξ_n are observed, forming p-dimensional vectors and playing the role of covariates.
The results will show that both approaches, the time series and the point process approach, yield satisfactory trend estimations and give insight into the inner dynamics of the process.

2 Semiparametric Point Process Model

The sequence τ_n, $n \geq 1$, of occurrence times is described by an intensity process λ_t, $t \geq 0$, in the multiplicative form

$$\lambda_t = \alpha(t) b_t(\theta), \quad \theta \in \Theta \subset \mathbb{R}^d, \tag{1}$$

with a deterministic and smooth function α and a stochastic intensity process b. The latter is modelled in a self-exciting form by

$$b_t = b^{(n)}(t,\theta) = f(w_n, t - \tau_n) \quad \text{for } t \in \ (\tau_n, \tau_{n+1}], \tag{2}$$

with a regression term w_n being iteratively defined by

$$w_n = u(w_{n-1}, (\sigma_n, \xi_n)).$$

Ex. 1. $f(w,s) = w, \quad u(w,(s,x)) = \rho e^{-\gamma s} w + \kappa e^{\beta^\top x}$;
we are faced with a piecewise constant intensity function b_t (Pruscha, 1983). With $w_0 = \kappa$ equation (2) can be written in the closed form

$$b^{(n)}(t,\theta) = \kappa \sum_{i=0}^{n} \rho^{n-i} e^{-\gamma(\tau_n - \tau_i)} e^{\beta^\top \xi_i}, \quad \xi_0 = 0,\ \tau_0 = 0.$$

Ex. 2. $f(w,s) = \rho + e^{-\gamma s} w, \quad u(w,(s,x)) = e^{-\gamma s} w + \kappa e^{\beta^\top x}$;
the intensity function b_t is of the form of Hawkes's self-exciting point process (Hawkes, 1971). In fact, with $w_0 = 0$ we can write equation (2) as

$$b^{(n)}(t,\theta) = \rho + \kappa \sum_{i=1}^{n} e^{-\gamma(t-\tau_i)} e^{\beta^\top \xi_i}.$$

We will close this section by deriving the conditional expectation of the waiting times $\sigma_n = \tau_n - \tau_{n-1}$. In terms of the intensity λ, the transition probabilities for σ_{n+1}, given the information $\mathcal{F}_n = (\tau_1, \xi_1, \ldots, \tau_n, \xi_n)$ of the past, can be written by

$$\mathbb{P}(\sigma_{n+1} \leq s \mid \mathcal{F}_n) = 1 - \exp(-\Lambda_{\tau_n, \tau_n + s}),$$

where $\Lambda_{s,t} = \int_s^t \lambda_u du$. Conseqently, the conditional expectation of σ_{n+1} is

$$\mathbb{E}(\sigma_{n+1} \mid \mathcal{F}_n) = \int_0^\infty \exp(-\Lambda_{\tau_n, \tau_n + s}) ds \ . \tag{3}$$

Assuming (for the moment) a piecewise constant intensity on $(\tau_n, \tau_{n+1}]$, say $\lambda^{(n)}$, we get $1/\lambda^{(n)}$ on the ride hand side of (3): the factors in the following model equation (4) will stand in a reciprocal relation to those in (1).

3 Semiparametric Time Series Model

The sequence σ_n, $n \geq 1$, of the waiting times between events is now analyzed within the framework of time series. Let $\mathcal{F}_n = (\sigma_1, \xi_1, \ldots, \sigma_n, \xi_n)$ comprise the information up to time τ_n as above. Then the conditional expectation of σ_{n+1}, given $\mathcal{F}_n$, is modelled by a transitional GLM of the form

$$\mathbb{E}(\sigma_{n+1} \mid \mathcal{F}_n) = \nu(\tau_n) h(\eta_n(\theta)), \tag{4}$$

with a smooth, deterministic trend component ν, a suitable response function h, and a regression term η_n iteratively defined by

$$\eta_n = \rho \eta_{n-1} + \gamma \sigma_n + G(\beta^\top \xi_n).$$

Ex. 1. $h(\eta) = \eta$ or $h(\eta) = 1/\eta$, with $\eta_0 = 1$, ρ and γ nonnegative, $G(x) = \exp(x)$ or $G(x) = x$, in the latter case with a side condition $\beta^\top \xi > 0$

Ex. 2. $h(\eta) = e^\eta$, with $\eta_0 = 0$, $G(x) = x$.

Observe that in both cases positivity of $h(\eta)$ is guaranteed. Zeger and Qaqish (1988, 2.2(iv)) used $h(\eta) = 1/\eta$ and put $1/\sigma_n$ (instead σ_n) in the regression term η_n. These types of response functions h are in fact suggested by the two examples in sec. 2. Assuming for simplicity $\alpha = 1$ then in the case of Ex. 1 in sec. 2 equation (3) yields $1/w_n$, leading to the choice $h(\eta) = 1/\eta$; in the case of Ex. 2 we get $c_1 \exp(-c_2 w_n)$ after some calculations, proposing the response function $h(\eta) = \exp(\eta)$.

4 Nonparametric Estimation of the Trend

In order to estimate α or ν a method of rough averaging is suggested. To this end, we divide the observation period $(0, T]$ into K subintervals $I_j = (u_{j-1}, u_j]$ of length $\Delta_j = u_j - u_{j-1}$, $j = 1, \ldots, K$, where K is assumed to be considerably smaller than the observed number N_T of events in $(0, T]$. We are going to base the l.s. function on the difference $Y_j - \mathbb{E}Y_j$, where Y_j is the rate of occurrence of events or the averaged length of time intervals between events, respectively, in the interval I_j. Let for the following $\Delta N_j = N_{u_j} - N_{u_{j-1}}$ be the number of events in I_j

4.1 Point process model

Putting $Y_j = \frac{\Delta N_j}{\Delta_j}$ we have on the basis of model (1)

$$\mathbb{E}Y_j = \frac{1}{\Delta_j} \int_{u_{j-1}}^{u_j} \mathbb{E}\lambda_s \, ds = \alpha(t_j^*) \cdot \frac{1}{\Delta_j} \int_{u_{j-1}}^{u_j} \mathbb{E} b_s(\theta) \, ds$$

for some $t_j^* \in [u_{j-1}, u_j]$. We assume, that the approximation

$$\alpha(t_j^*) \approx \alpha(t_j), \quad t_j = \frac{1}{2}(u_{j-1} + u_j), \tag{5}$$

is possible, as well as, for larger $\Delta_j = u_j - u_{j-1}$,

$$\frac{1}{\Delta_j}\int_{u_{j-1}}^{u_j} \mathbb{E} b_s(\theta)\, ds \;\approx\; b^{(\infty)}(\theta), \tag{6}$$

where $b^{(\infty)}(\theta)$ is independent of j. Putting $a(t) = \alpha(t)\cdot b^{(\infty)}(\theta)$, and using the approximations (5) and (6), we can write $\mathbb{E}Y_j \approx a(t_j)$.

In Ex.2 of sec.2 the limit intensity value is identified as (Pruscha, 1997)

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

where $\sigma^{(\infty)} =$ a.s.- $\lim(\frac{1}{n}\sum_{i=1}^n \sigma_i)$. In Ex.1 of sec. 2 the limit intensity is $\frac{\kappa}{1-\rho\exp(-\sigma^{(\infty)}\cdot\gamma)}$.

4.2 Time series model

We divide the interval $(0,T]$, $T=\tau_N$, as above and define $Y_j = \frac{\Delta_j}{\Delta N_j}$. Assuming for the moment that $I_j = (\tau_k, \tau_{k+n(j)}]$, i.e. $\Delta N_j = n(j)$, for some $k = k(j)$, we have on the basis of model (4)

$$\mathbb{E}Y_j = \frac{1}{n(j)}\sum_{i=1}^{n(j)} \mathbb{E}(\sigma_{k+i}) = \nu(t_j^*)\cdot\frac{1}{n(j)}\sum_{i=1}^{n(j)} \mathbb{E}h(\eta_{k+i-1}(\theta))$$

for some $t_j^* \in \bar{I}_j$. By means of approximations similar to (5) and (6) above, namely $\nu(t_j^*) \approx \nu(t_j)$, $t_j = \frac{1}{2}(u_{j-1}+u_j)$, and, for larger $n(j)$,

$$\frac{1}{n(j)}\sum_{i=1}^{n(j)} \mathbb{E}h(\eta_{k+i-1}(\theta)) \;\approx\; h^{(\infty)}(\theta), \tag{7}$$

we arrive at $\mathbb{E}Y_j \approx a(t_j)$, where we now have $a(t) = \nu(t)\cdot h^{(\infty)}(\theta)$. For the regression term η_n and the examples given in sec. 3, explicit formulas for $h^{(\infty)}(\theta)$ can be given, assuming the existence of the Cesaro limits $\sigma^{(\infty)}$, $\xi^{(\infty)}$, $\eta^{(\infty)}$, $G^{(\infty)}(\beta)$ of the sequences σ_n, ξ_n, η_n, $G(\beta^\top\xi_n)$. Indeed, we first have for $|\rho|<1$

$$\eta^{(\infty)}(\theta) = \frac{1}{1-\rho}\left(\gamma\sigma^{(\infty)} + G^{(\infty)}(\beta)\right), \tag{8}$$

where $G^{(\infty)}(\beta) = \beta^\top\xi^{(\infty)}$ in the case $G(x)=x$. Then for the examples in sec. 3

Ex. 1 (with $h(\eta)=\eta$): $h^{(\infty)}(\theta) = \eta^{(\infty)}(\theta)$

Ex. 2 (with $h(\eta)=e^\eta$): If centered variables σ_i and ξ_i are used, we get $\eta^{(\infty)} = 0$ from (8), such that we have $h^{(\infty)}(\theta) \approx 1+\eta^{(\infty)} = 1$ for all θ; i.e. we can assume that the side condition (9) below is automatically fulfilled.

4.3 An asymptotic set-up justifying the various approximations will be given below in sec 6. We now define for both models the penalized l.s. criterion $\Psi(a) = \frac{1}{K}SSE(a) + \lambda H^{(2)}(a)$, with $\lambda > 0$ a smoothing parameter and with

$$SSE(a) = (Y-a)^\top\cdot W\cdot(Y-a), \quad H^{(2)}(a) = \int_0^T \left(a''(s)\right)^2\, ds,$$

where $(Y - a)$ is a $K \times 1$–vector having components $Y_j - a(t_j)$ and where W is an appropriate $K \times K$–weight matrix. It is well known, that $\Psi(a) = \min$ is solved by natural cubic spline functions a (Green and Silverman, 1994). Note that α or ν can be estimated only in the form $a = \alpha \cdot b^{(\infty)}$ or $a = \nu \cdot h^{(\infty)}$, where the second factor does not depend on t, but on θ.

5 Parametric Estimation

For the parametric part of the problem we apply likelihood methods. Maximization of the log-likelihood function $l_T(\theta)$, $\theta \in \Theta \subset \mathbb{R}^d$, $T = \tau_N$, is done by plugging into $l_T(\theta)$ the spline solution $\hat{a}$ from sec. 4, leading to a function $\hat{l}_T(\theta)$, and then solving

$$\hat{l}_T(\theta) = \max, \quad \text{under } b^{(\infty)}(\theta) = 1 \text{ or } h^{(\infty)}(\theta) = 1. \tag{9}$$

The side condition should guarantee that the two factors α and b in (1) or ν and h in (4) can be separated by the estimation procedure.

Point process model: The log-likelihood function of a realization $(\tau_1, \ldots, \tau_N)$ of the point process has the form

$$l_T(\theta) = \sum_{i=1}^{N} \left(\log(\alpha(\tau_i) b^{(i-1)}(\tau_i, \theta)) - \int_{\tau_{i-1}}^{\tau_i} \alpha(s) b^{(i-1)}(s, \theta)\, ds \right),$$

where we have to insert $\hat{\alpha}(t) = \hat{a}(t)/b^{(\infty)}(\theta)$.

Time series model: Assuming that the σ_i's are (conditionally) exponentially distributed, the log-likelihood function of a time series realization $(\sigma_1, \ldots, \sigma_N)$ is given by $\sum_{i=1}^{N} \log(\lambda_{i-1} \exp(-\lambda_{i-1} \cdot \sigma_i))$, where $\lambda_i = 1/(\nu(\tau_i) \cdot h(\eta_i))$ cf (4). We arrive at

$$l_T(\theta) = \sum_{i=1}^{N} \left(-\log \nu(\tau_{i-1}) h(\eta_{i-1}(\theta)) - \frac{\sigma_i}{\nu(\tau_{i-1}) h(\eta_{i-1})} \right).$$

Plugging in the solution $\hat{\nu}(t) = \hat{a}(t)/h^{(\infty)}(\theta)$ from sec. 4 and neglecting irrelevant terms we are led to a function $\hat{l}_T(\theta)$ which can (under the side condition $h^{(\infty)}(\theta) = 1$) be interpreted as the log-likelihood function of a realization $(\hat{\sigma}_1, \ldots, \hat{\sigma}_N)$, with $\hat{\sigma}_n = \sigma_n / \hat{a}(\tau_{n-1})$, fulfilling the (purely parametric) model equation

$$\mathbb{E}(\hat{\sigma}_{n+1} \mid \mathcal{F}_n) = h(\eta_n(\theta)). \tag{10}$$

For the transitional GLM (10) quasi-likelihood methods from GLM software can be employed.

6 Asymptotic Set-up

6.1 Trend function

The following set-up is known from nonparametric regression (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; define for $T > 0$

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

Let further $I_{j,T} = (u_{j-1,T}, u_{j,T}]$, $j = 1, 2, \ldots$, be a division of $(0,T]$ into subintervals of length $\Delta_{j,T}$. Putting $\Delta_T^{(0)} = \min_j \Delta_{j,T}$, $\Delta_T^{(1)} = \max_j \Delta_{j,T}$, then we consider limits of the kind

$$T \to \infty, \quad \Delta_T^{(0)} \to \infty, \quad \frac{\Delta_T^{(1)}}{T} \to 0. \tag{11}$$

Since for $s, s' \in I_{j,T}$

$$|\alpha_T(s) - \alpha_T(s')| \le \Delta_{j,T} \max_{I_{j,T}} |\alpha_T'(t)| \le \frac{\Delta_T^{(1)}}{T} \max_{[0,1]} |\alpha_1'(t)|,$$

which tends to 0 under (11), the setting (5) is established. An analogous device is used for the trend function $\nu_T(t) = \nu_1(t/T)$, $T = \tau_N$, in the time series case.

6.2 Ergodic properties

Point process model: For each $T > 0$, the counting process N_t, $t \in [0,T]$, with intensity process $\lambda_{t,T} = \alpha_T(t) \cdot b_t(\theta), t \in [0,T]$, may fulfil the ergodic law

$$\frac{1}{\Delta_T} \int_{I_T} \mathbb{E}_T b_s(\theta)\, ds \to b^{(\infty)}(\theta)$$

for limits of the kind (11), where I_T denotes an interval of length Δ_T. Then the approximation (6) is justified.

Time series model: For each $N \in \mathbb{N}$ the variables σ_n, $n \in \{1, \ldots, N\}$, satisfying the equation

$$\mathbb{E}_N(\sigma_{n+1} \mid \mathcal{F}_n) = \nu_T(\tau_n) h(\eta_n(\theta))$$

$(T = \tau_N)$, may fulfil the ergodic law

$$\frac{1}{n_N} \sum_{i \in I_N} \mathbb{E}_N h(\eta_i(\theta)) \;\to\; h^{(\infty)}(\theta),$$

where I_N comprises n_N adjacent integers. Then the approximation (7) is justified. A more detailed analysis will be given elsewhere.

7 Application

Data sets on long-term ECGs of patients suffering from heart arrhythmias consist of the occurrence times τ_n of ventricular extrasystoles. The covariates ξ_n are the strength of these events measured as the relative deviation from the normal beat (see Pruscha, Ulm & Schmid, 1997). As an example we choose a patient with 714 extrasystoles within a 20-hours observation period. The reciprocal relationship of the trend functions $\alpha(t)$ and $\nu(t)$ becomes apparent in the plot of *Fig. 1*.

Parameters were estimated from the data of 48 patients; 15 persons died of a later sudden heart attack, while 33 survived. We used

- *point process model* (1), Ex. 2, with $\kappa = 1$ and normalized σ_i's; the mean cluster size m_c is also calculated from the estimated parameters, according to a cluster process representation by Hawkes & Oakes (1974) (γ will be denoted by γ_{PP})
- *time series model* (4), Ex. 2., employing `S+,glm,quasi(link=log),dispersion` ϕ `estimated` for equation (10), with standardized σ_i's, (γ will here be denoted by γ_{TS})

The *Table* shows the tendency, that larger positive γ_{PP}-values and smaller mean clustering sizes m_c correspond to smaller γ_{TS}-values and smaller autocorrelations $r_{\hat{\sigma}}(1)$, where $\hat{\sigma}_i = \sigma_i / \hat{a}(\tau_{i-1})$ as in (10).

Patient	N	γ_{PP}	m_c	γ_{TS}	$r_{\hat{\sigma}}(1)$	$\sqrt{\hat{\phi}}$	$\hat{h}^{(\infty)}$
OTDJFZ	482	125.0	1.008	-0.043	-0.036	0.955	1.001
OVXSBH	714	2.532	1.653	0.043	0.092	5.872	1.001
PFQWGU	1697	1.981	2.019	0.049	0.151	1.878	1.012
OTEMQL	1788	1.802	2.248	0.151	0.194	1.619	1.014

The data of the 48 patients, labelled by their status *1=survival* or *2=death*, were submitted to a binary logistic regression, with $x = Sz$, $y = \sqrt{m_c - 1}$ and $z = VES/h$ as regressors. Hereby, *VES/h* is the average rate of extrasystoles per hour, and *Sz* is the standard deviation of the time interval lengths between those two heart beats being qualified as normal beats (not being designated as extrasystoles), vgl. *Tab. 1* of Pruscha, Ulm & Schmid (1997). Then in the scatter diagram of *Fig. 2*, the (x', y') values of the 48 patients were plotted, with x', y' being the x and y values, respectively, corrected by the variable $z = VES/h$ on the basis of the logistic regression equation. We have three distinct outlier cases from the group *2* lying in the "area" of group *1*. The two most extremes of them are also outliers in the evaluation system of *nonlinear dynamics* (Schmid et al, 1996), where the variables $x = \alpha_{sin}$ and $y = \alpha_{VES}$ were used. With respect to the quality of prediction both systems perform nearly equally well.

Acknowledgments

This work is supported by the Deutsche Forschungsgemeinschaft (SFB 386). The ECG data were kindly placed at my disposal by K. Ulm (Technical University, Munich). The help of dipl. math. S. Siegle with respect to `Fortran` programming and `S+` data analysis is thankfully acknowledged.

References

Cox, D. R. & Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

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

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* M. Decker, N.Y.

Fahrmeir, L. & Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer-Verlag, N.Y.

Green, P. J. & 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. & Oakes, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.*, **11**, 493-503.

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

Pruscha, H. (1997). Semiparametric estimation in regression models for point processes based on one realization. *Proceedings of the 12th International Workshop on Statistical Modelling, Biel 1997*, 260-264.

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

Schmid, G., Morfill, G. E. et al. (1996). Variability of ventricular premature complexes and mortality risk. *PACE*, **19**, 976-980.

Zeger, S. L. & Qaqish, B. (1988). Markov regression models for time series: a quasi likelihood approach. *Biometrics*, **44**, 1019-1031.

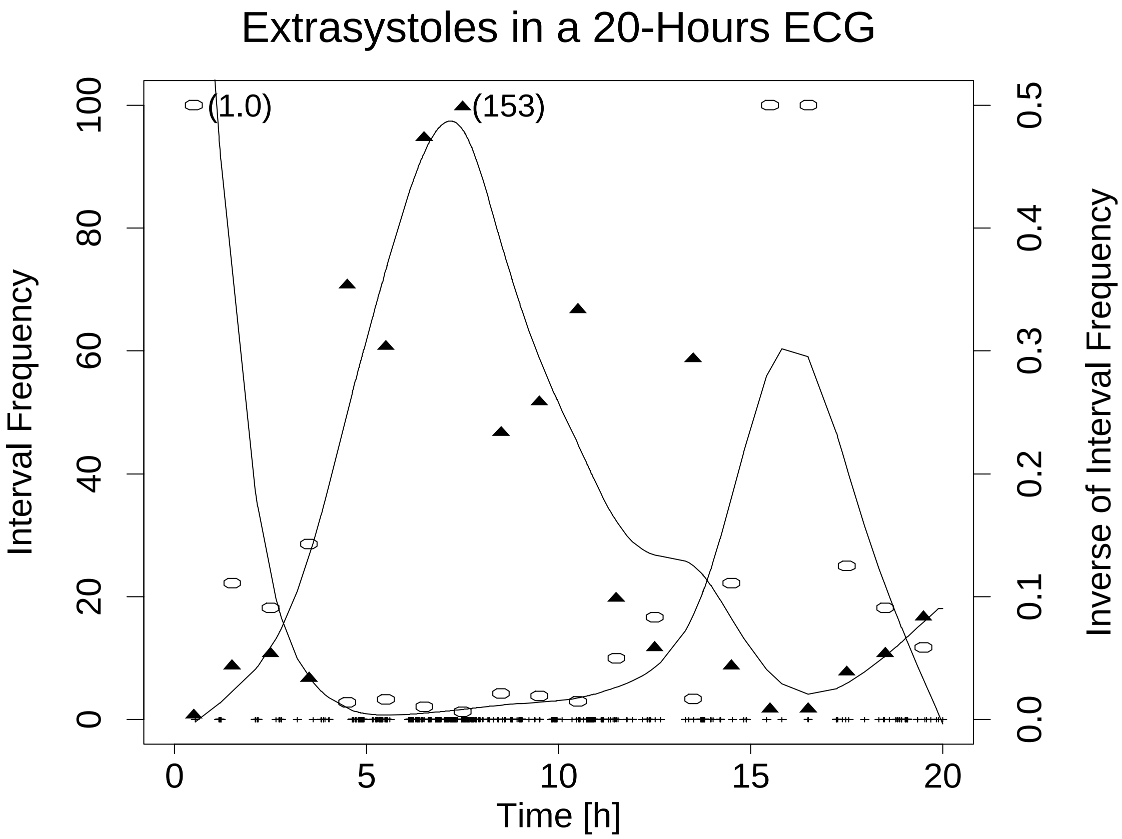
Legends to Figs.

Fig. 1 (above)

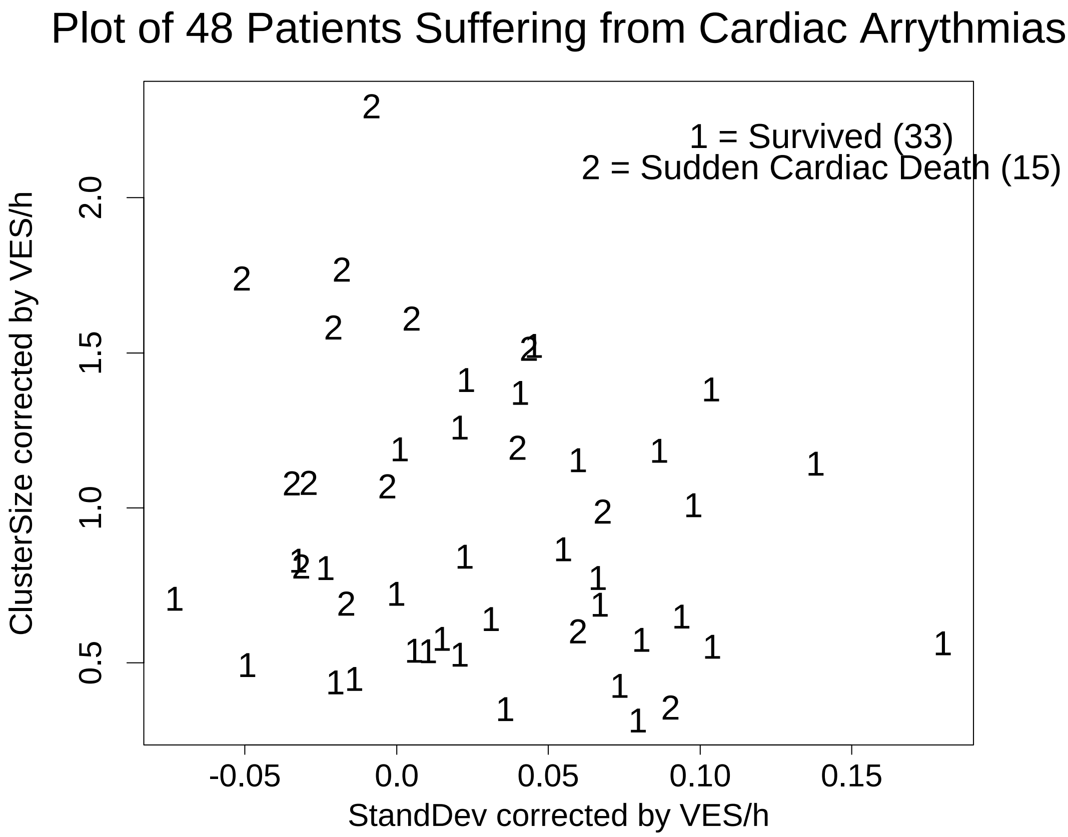
Extrasystoles within a 20-hours ECG record of a patient (OVXSBH). The $n = 714$ time points of occurrence are marked on the time scale. The occurrence frequencies $n(j)$ (triangles) and the mean waiting times $\frac{1}{n(j)}$ (circles) were plotted over the $j = 1, \ldots, 20$ hours, together with the estimated trend functions α and ν, respectively.

Fig. 2 (below)

Scatterplot of the variable *Cluster Size* over the variable *Standard Deviation*, both corrected by VES/h on the basis of a logistic regression equation, for 48 patients.



Extrasystoles in a 20-Hours ECG
Interval Frequency
Inverse of Interval Frequency
Time [h]
(1.0)
(153)
0
20
40
60
80
100
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20



Plot of 48 Patients Suffering from Cardiac Arrythmias
1 = Survived (33)
2 = Sudden Cardiac Death (15)
ClusterSize corrected by VES/h
StandDev corrected by VES/h
0.5
1.0
1.5
2.0
-0.05
0.0
0.05
0.10
0.15