

Computational details for ppstat

Niels Richard Hansen¹, *University of Copenhagen*

1 The likelihood and derivatives

The core computational problem in the use of point processes for statistical modeling is the optimization of the minus-log-likelihood function, which is given as

$$l(\theta) = \int_0^T \lambda_\theta(s) ds - \sum_{j=1}^n \log \lambda_\theta(t_j)$$

where $0 < t_1 < t_2 < \dots < t_n < T$ are observations and λ_θ is a parameterized family of intensities. Typically $\theta \in \Theta \subseteq \mathbb{R}^p$. For the Hawkes family of generalized linear point process models in `ppstat` we consider situations where

$$\lambda_\theta(t) = \phi \left(\alpha^T X(t) + \sum_{m=1}^K \sum_{i=1}^{n(m)} h_{\beta^m}^m(t - s_i^m) \right)$$

where $\phi : I \rightarrow [0, \infty)$, $I \subseteq \mathbb{R}$, is a given function,

$$\theta = \begin{pmatrix} \alpha \\ \beta^1 \\ \vdots \\ \beta^K \end{pmatrix}$$

and $s_1^m < \dots < s_{n(m)}^m < t$ for $m = 1, \dots, K$ are observations of point processes (one of these sets of points could be the t_i observations above). The process $X(t)$ is an auxiliary, $d(0)$ -dimensional observed processes – observed at least discretely. The processes

$$\sum_{i=1}^{n(m)} h_{\beta^m}^m(t - s_i^m)$$

are linear filters using the (parameterized) filter function $h_{\beta^m}^m$, which are given via a basis expansion

$$h_{\beta^m}^m(t) = (\beta^m)^T B(t) = \sum_{l=1}^{d(m)} \beta_l^m B_l(t),$$

¹Postal address: Department of Mathematical Sciences, University of Copenhagen Universitetsparken 5, 2100 Copenhagen Ø, Denmark.
Email address: richard@math.ku.dk

and $\beta^m \in \mathbb{R}^{d(m)}$. Collecting these ingredients – and interchanging two sums – the intensity function can be written as

$$\lambda_\theta(t) = \phi \left(\alpha^T X(t) + \sum_{m=1}^K \sum_{l=1}^{d(m)} \beta_l^m \sum_{i=1}^{n(m)} B_l(t - s_i^m) \right) = \phi(\theta^T Z(t))$$

with $Z(t)$ a process of dimension $p = d(0) + d(1) + \dots + d(K)$. Each of the linear filter components

$$\sum_{i=1}^{n(m)} B_l(t - s_i^m)$$

are computable from the observations and the fixed choice of basis. The minus-log-likelihood function that we want to minimize reads

$$l(\theta) = \int_0^T \phi(\theta^T Z(s)) ds - \sum_{j=1}^n \log \phi(\theta^T Z(t_j)).$$

The integral is not in general analytically computable. We discretize time to have a total of N time points and let Z denote the $N \times p$ matrix of the $Z(t)$ -process values at the discretization points. With Δ the N -dimensional vector of interdistances from the discretization we arrive at the approximation of the minus-log-likelihood function that we seek to minimize:

$$l(\theta) \simeq \Delta^T \phi(Z\theta) - \sum_{j=1}^n \log \phi(\theta^T Z(t_j)).$$

We have used the convention that ϕ applied to a vector means coordinate-wise applications of ϕ . Using this expression a precomputation of the Z matrix will allow for a rapid computation of (the approximation to) l . The derivatives are likewise approximated as

$$Dl(\theta) \simeq [\Delta \circ \phi'(Z\theta)]^T Z - \sum_{j=1}^n \frac{\phi'(\theta^T Z(t_j))}{\phi(\theta^T Z(t_j))} Z(t_j)^T$$

with \circ the Hadamard (or coordinate-wise) matrix product and

$$D^2l(\theta) \simeq Z^T [\Delta \circ \phi''(\theta^T Z) \circ Z] - \sum_{j=1}^n \frac{\phi''(\theta^T Z(t_j)) \phi(\theta^T Z(t_j)) - \phi'(\theta^T Z(t_j))^2}{\phi(\theta^T Z(t_j))^2} Z(t_j) Z(t_j)^T.$$

The formulas above are used for generic computations, but one particular case can be somewhat simplified. If $\phi = \text{Id}$ is the identity function then

$$\Delta^T \phi(Z\theta) = (\Delta^T Z)\theta$$

and the $z = \Delta^T Z$ can be precomputed. This is similar to computing the time integrals of the B_l -basis functions *before* the expansion of the linear filter.

2 Algorithms for optimization

The BFGS and L-BFGS-B algorithms as implemented in R.

The iterative weighted least squares algorithm.