

Statistics 222, Spatial Statistics.

Outline for the day:

1. Old projects and exam on the course website

<http://www.stat.ucla.edu/~frederic/222/F23> .

2. Fitting Hawkes models with covariates. `fithawkeswithcovariates.r`

3. Integral term in loglikelihood for Hawkes processes.

4. Analytic nonparametric estimation.

Next week we will do review and Stoyan-Grabarnik statistic and Stoyan-Grabarnik estimation.

3. The integral term in the loglikelihood for Hawkes processes.

$$\text{loglikelihood} = \sum_i \log(\lambda(t_i, x_i, y_i)) - \iiint \lambda(t, x, y) dx dy dt.$$

The space-time region is $B = [0, T] \times S$.

For a Hawkes process, $\lambda(t, x, y) = \mu \rho(x, y) + K \sum_{i: t_i < t} g(t - t_i, x - x_i, y - y_i)$, where ρ and g are densities.

$$\begin{aligned} \int_0^T \iint \lambda(t, x, y) dx dy dt &= \int_0^T \iint \mu \rho(x, y) dx dy dt + \int_0^T \iint K \sum_{i: t_i < t} g(t - t_i, x - x_i, y - y_i) dx dy dt \\ &= \mu T + \int_0^T \iint K \int_B 1_{\{t' < t\}} g(t - t', x - x', y - y') dN(t', x', y') dx dy dt \end{aligned}$$

interchanging the integrals

$$= \mu T + K \int_B \int_0^T \iint 1_{\{t' < t\}} g(t - t', x - x', y - y') dx dy dt dN(t', x', y')$$

changing coordinates, letting $u = t - t'$, $v = x - x'$, $w = y - y'$,

$$= \mu T + K \int_B \int_0^{T-t'} \iint_{S-(x', y')} g(u, v, w) du dv dw dN(t', x', y')$$

$$\sim \mu T + K \int_B (1) dN(t', x', y')$$

$$= \mu T + KN(B).$$

This is approximate because typically $\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty g(u, v, w) du dv dw = 1$, but instead, we have

$$\int_0^{T-t'} \iint_{S-(x', y')} g(u, v, w) du dv dw \text{ which is often close to } 1.$$

4. Nonparametric triggering function estimation.

Marsan and Lengliné (2008) assume g is a step function, and estimate steps β_k as parameters.

$$\ell(\theta) = \sum_i \log(\lambda(\tau_i, \mathbf{x}_i | \mathcal{H}_{\tau_i})) - \int_0^T \int_S \lambda(t, \mathbf{x} | \mathcal{H}_t) dx dt$$

Setting the partial derivatives of this loglikelihood with respect to the steps β_k to zero yields

$$0 = \partial \ell(\theta) / \partial \beta_k = \sum_{(i,j): \tau_i - \tau_j \in U_k} K / \lambda(\tau_i) - Kn |U_k|,$$

where $|U_k|$ is the width of step k , for $k = 1, 2, \dots, p$. This is a system of p equations in p unknowns.

However, the equations are nonlinear. They depend on $1/\lambda(\tau_i)$.

Gradient descent methods: way too slow for large p .

Marsan and Lengliné (2008) find *approximate* maximum likelihood estimates using the E-M method for point processes. You pick initial values of the parameters, then given those, you know the probability event i triggered event j . Using these, you can weight each pair of points by its probability and re-estimate the parameters, and repeat until convergence.

This method works well but is iterative and time-consuming.

Analytic solution.

Set $p = n$. ($p =$ number of steps in the step function, g , and $n =$ # of observed points.)

Setting the derivatives of the loglikelihood to zero we have the p equations

$$0 = \partial \ell(\theta) / \partial \beta_k = \sum_{(i,j): \tau_i - \tau_j \in U_k} K / \lambda(\tau_i) - K n |U_k|,$$

which are p linear equations in terms of $1/\lambda(\tau_i)$, for $i = 1, 2, \dots, n$. (!)

So, if $p=n$, then we can use these equations to solve for $1/\lambda(\tau_i)$,

and if we know $1/\lambda(\tau_i)$, then we know $\lambda(\tau_i)$,

and if we know $\lambda(\tau_i)$, then we can solve for β_i because the def. of a Hawkes process is

$$\lambda(\tau_j) = \mu + K \sum_{i < j} g(\tau_j - \tau_i),$$

which results in n linear equations in the p unknowns $\beta_1, \beta_2, \dots, \beta_p$, when g is a step function.

Analytic solution.

We can write the resulting estimator in very condensed form.

Let $\lambda = \{\lambda(\tau_1), \lambda(\tau_2), \dots, \lambda(\tau_n)\}$.

Suppose the steps of g have equal widths, $|U_1| = |U_2|$, etc. Call this width U .

Let A_{ij} = the number points τ_k such that $\tau_j - \tau_k$ is in U_i , for i, j in $\{1, 2, \dots, p\}$.

Then the loglikelihood derivatives equalling zero can be rewritten

$$0 = KA(I/\lambda) - Kb,$$

where $\mathbf{b} = nU\mathbf{1}$, with $\mathbf{1} = \{1, 1, \dots, 1\}$.

This has solution $I/\lambda = A^{-1}b$, if A is invertible.

Similarly, the Hawkes equation can be rewritten $\lambda = \mu + KA^T\beta$, whose solution is

$$\underline{\hat{\mathbf{b}}} = (KA^T)^{-1}(\lambda - \mu).$$

Combining these two underlined formulas yields the estimates

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

This is very simple, trivial to program, and rapid to compute.

Analytic solution.

There are problems, however.

1. Estimating $n=p$ steps. High variance.

However, if we can assume g is smooth, then we can smooth our estimates for stability.

2. Need to estimate K and μ too.

We can use Marsan and Lengliné's method or take derivatives for these as well.

3. What about spatially-varying steps and unequally sized steps for g ?

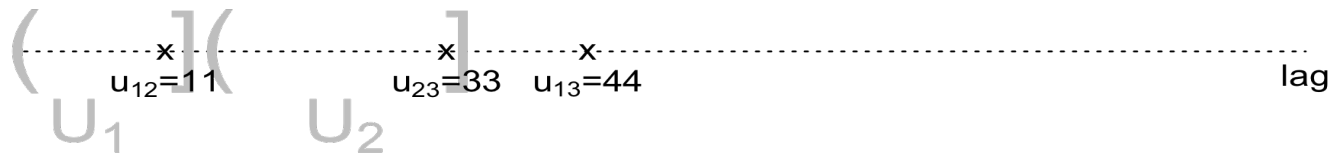
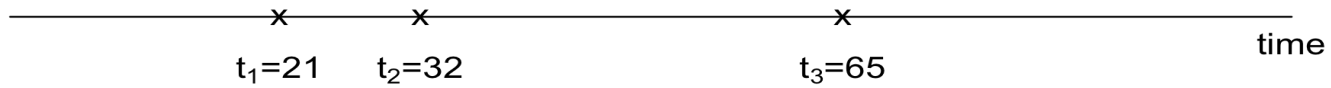
No problem. The estimation generalizes in a completely obvious way.

4. A can be singular.

We may need better solutions for this.

I let $u_j = \tau_j - \tau_{j-1}$, sorted the u_j values, and then used $[u_{(1)}, u_{(2)}]$, etc. as my binwidths, so each row and column of A would have at least one non-zero entry. If it still isn't singular, adding in a few random 1's into A often helps.

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$



4. A can be singular.

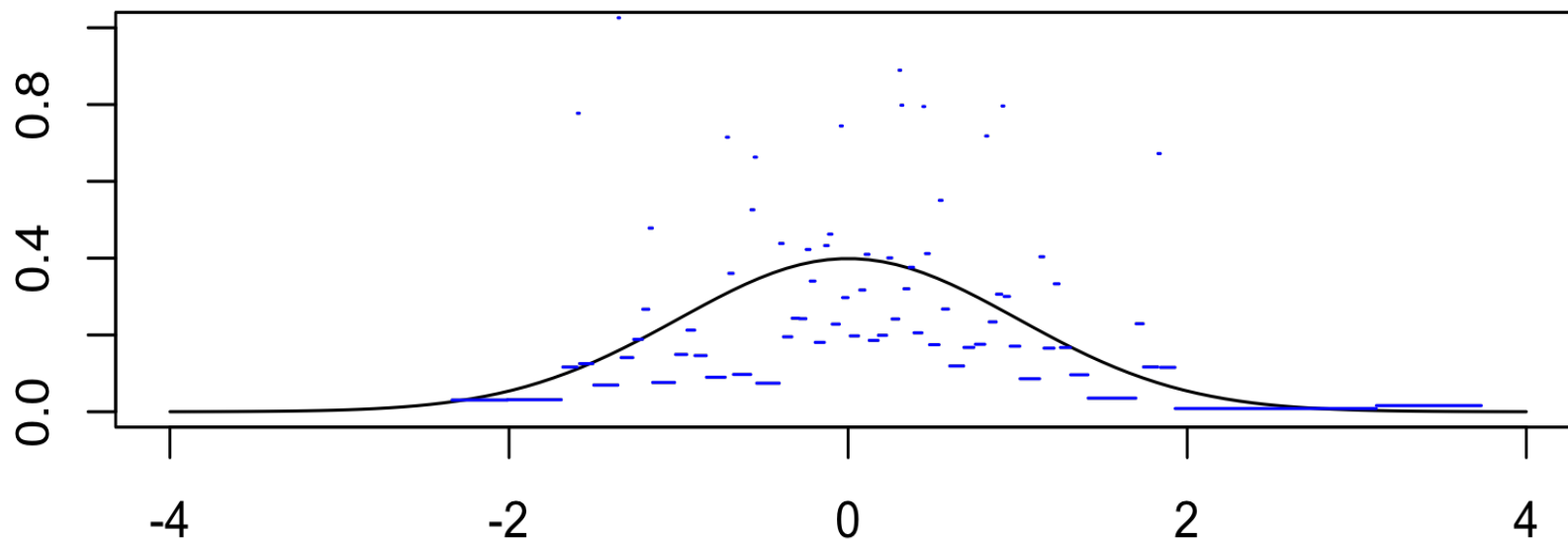
We may need better solutions for this.

I let $u_j = \tau_j - \tau_{j-1}$, sorted the u_j values, and then used $[u_{(1)}, u_{(2)}]$, etc. as my binwidths, so each row and column of A would have at least one non-zero entry. If it still isn't singular, adding in a few random 1's into A often helps.

$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Note: take the simple case of a dataset where point i is only influenced by point $i-1$. This is basically a renewal process, and we are just estimating a renewal density.

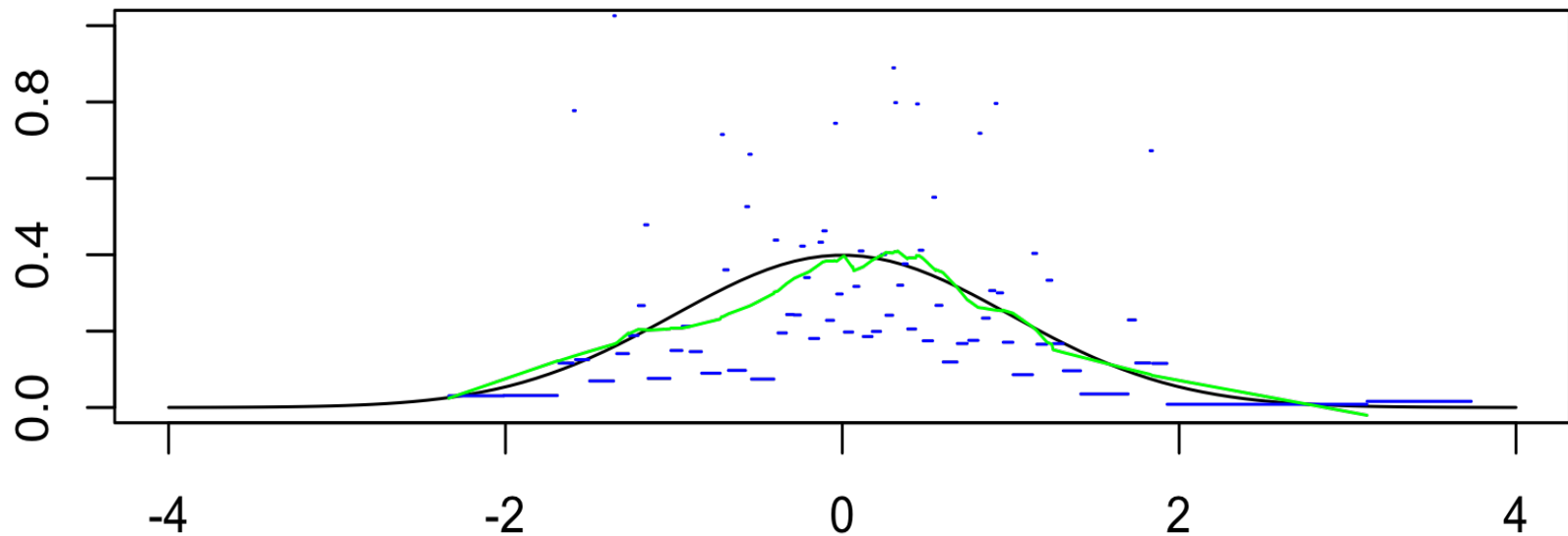
Here $A = I$, $K = 1$, and we get the density estimator $1/\{n(x_i - x_{i-1})\}$.



$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Note: take the simple case of a dataset where point i is only influenced by point $i-1$. This is basically a renewal process, and we are just estimating a renewal density.

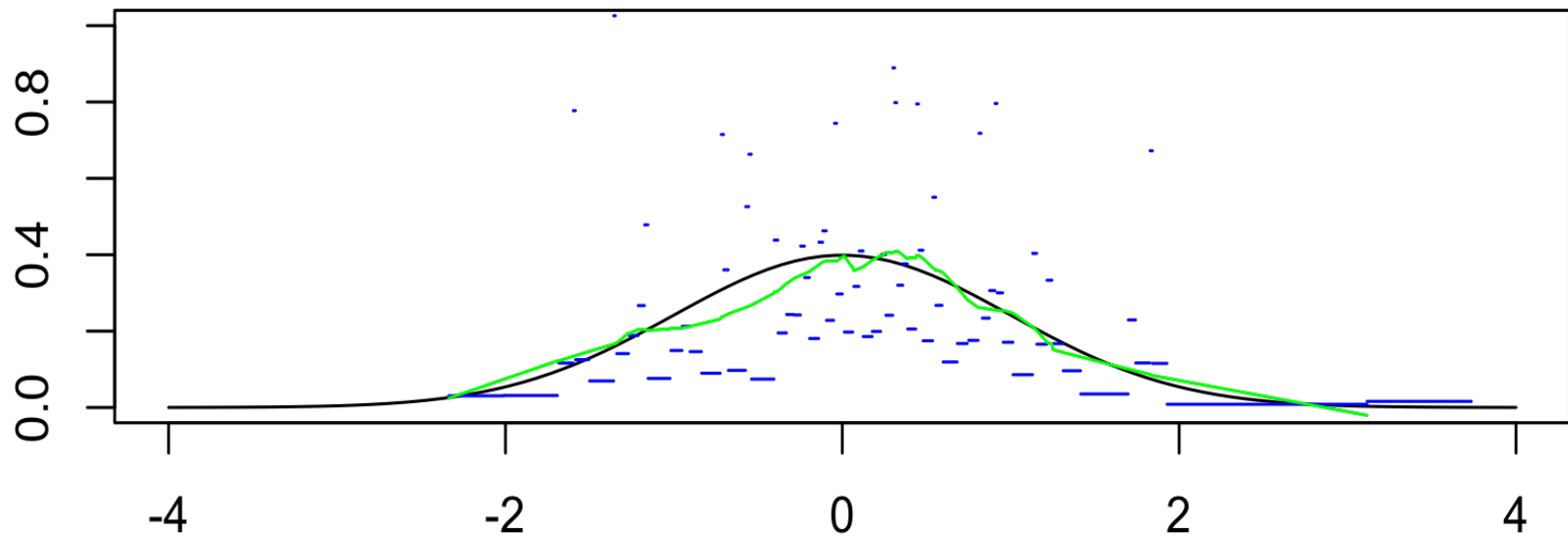
Here $A = I$, $K = 1$, and we get the density estimator $1/\{n(x_i - x_{i-1})\}$.



$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Note: take the simple case of a dataset where point i is only influenced by point $i-1$. This is basically a renewal process, and we are just estimating a renewal density.

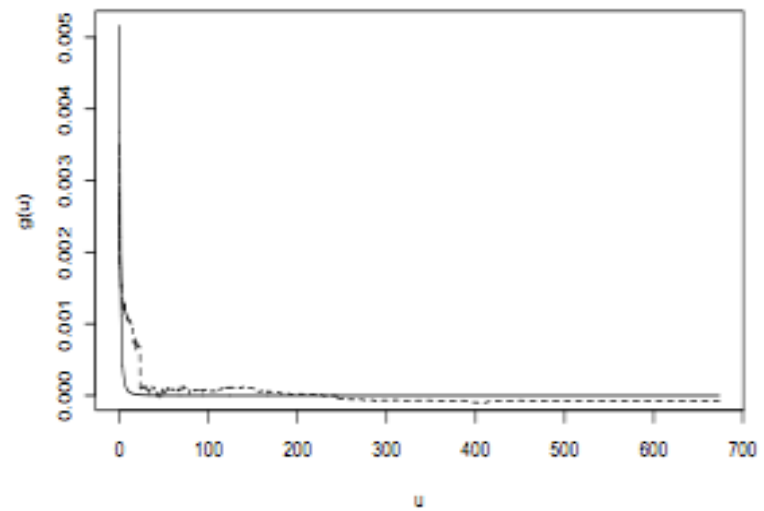
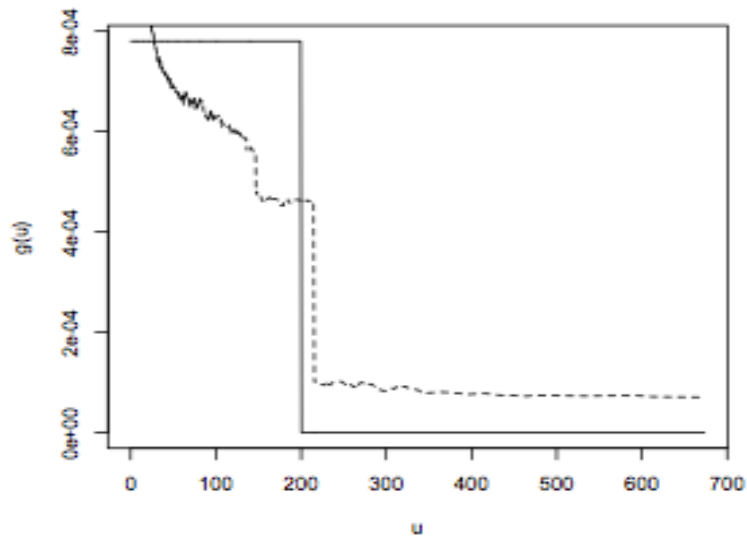
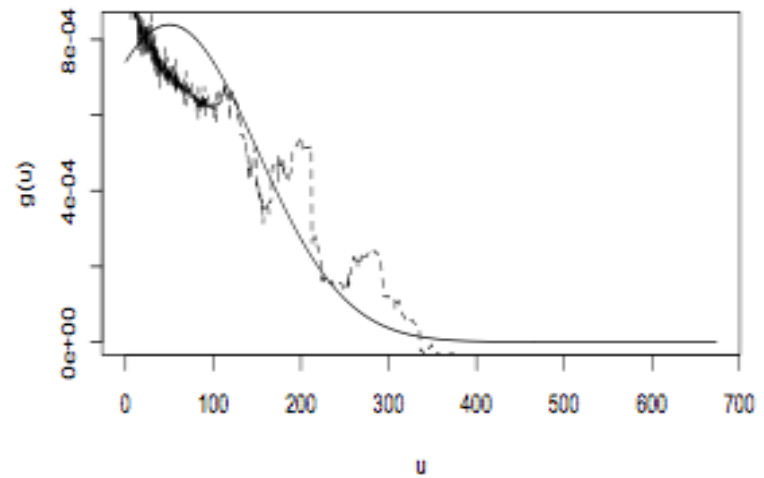
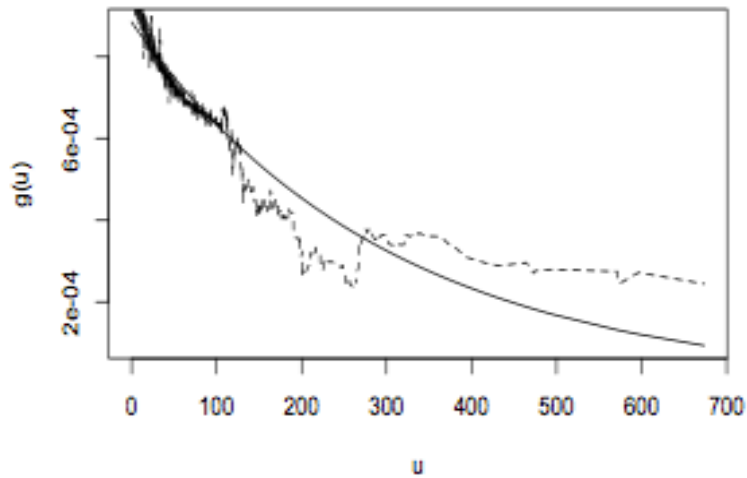
Here $A = I$, $K = 1$, and we get the density estimator $1/\{n(x_i - x_{i-1})\}$.



$$\hat{\beta} = (KA^T)^{-1}[1/(A^{-1}b) - \mu]$$

Computation time and performance comparison.

Test of concept. Examples of exponential, truncated normal, uniform, and Pareto g .



Computation time and performance comparison.

Triangles = Marsan and Lengliné (2008) method. Circles = analytic method.

