State estimation for temporal point processes

Marie-Colette van Lieshout

colette@cwi.nl

CWI & University of Twente
The Netherlands

Includes joint work with Robin Markwitz.

Overview

Inference for temporal point processes is usually based on the **stochastic intensity**. This is

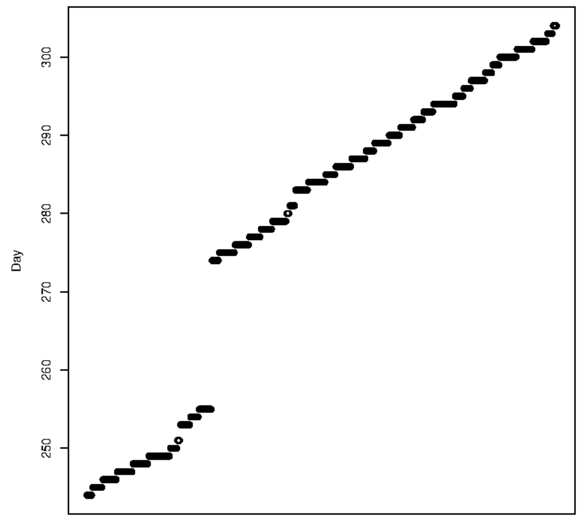
- quite natural;
- fits in with powerful tools from martingale theory;
- allows a closed form likelihood;

but does not seem able to deal with an interrupted flow of time.

Aim of this talk: develop inference for

- **imputation** of missing data;
- **aoristically censored** data.

Missing data



Dates of calls during Sept–Oct 2001 to NHS Direct in days from 1/1/2001 against index. Part of data covering 2001–2002. (Diggle and Hawtin)

Salient feature: calls during Sept 13–30 not registered due to technical failure.

Goal: estimate how many calls were lost.

Model

Label days from September 1st. Set

- $N(i)$: number of calls during day i, **Poisson** distributed with intensity

$$\mu_0(i) \exp\left[S(i)\right], \quad i = 0, \ldots, 60;$$

- $m_0(i)$: seasonal and weekly effects;
- $S(i)$: a discrete **Ornstein–Uhlenbeck** process

$$S(i) = -\frac{\sigma^2}{2} + \sigma \sum_{j=0}^{i} e^{-\beta(i-j)} \Gamma(j)$$

for independent

- $\Gamma(i) \sim \mathcal{N}(0, 1 - e^{-2\beta})$, $i > 0$,
- $\Gamma(0) \sim \mathcal{N}(0, 1)$.

Baseline model

$$\begin{aligned}\log \mu_0(i) &= \delta_{d(i)} \\ &+ a_1 \cos\left(\frac{2i\pi}{365}\right) + b_1 \sin\left(\frac{2i\pi}{365}\right) + a_2 \cos\left(\frac{4i\pi}{365}\right) + b_2 \sin\left(\frac{4i\pi}{365}\right) \\ &+ g \times i.\end{aligned}$$

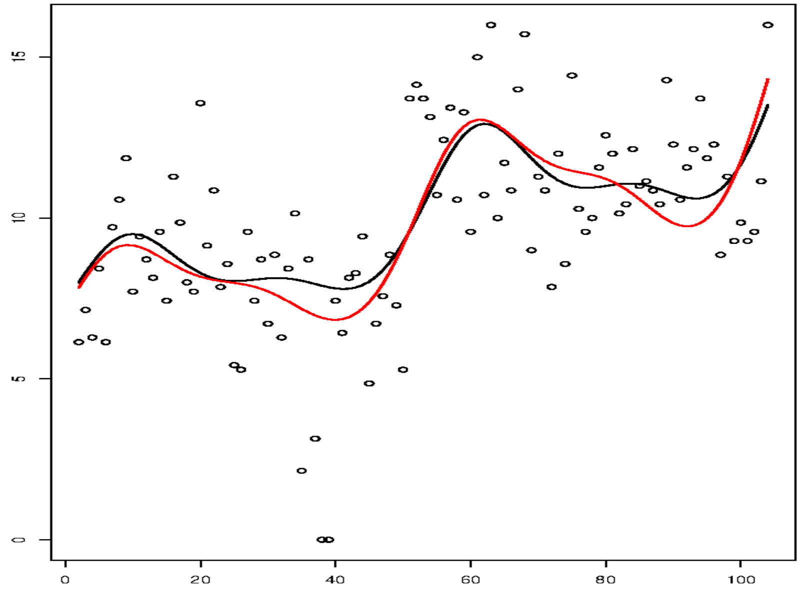
Here

- $\delta_{d(i)}$: more calls during the week-end;
- second order harmonics for seasonal effects;
- g: trend parameter for gradual acceptance of new service.

CWI

UNIVERSITY OF TWENTE.

Baseline parameters



Validate Poisson log-linear baseline model by weekly averages during 2001–2002.

- **Red** line: naive approach plugging in 0 for missing observations.
- **Black** line: taking into account the gap.

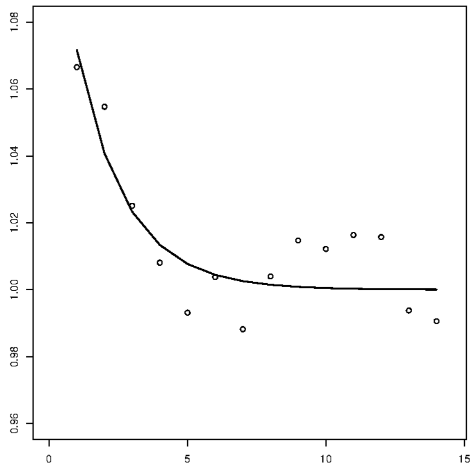
Note: trough because of missing data clearly visible, also a year later.

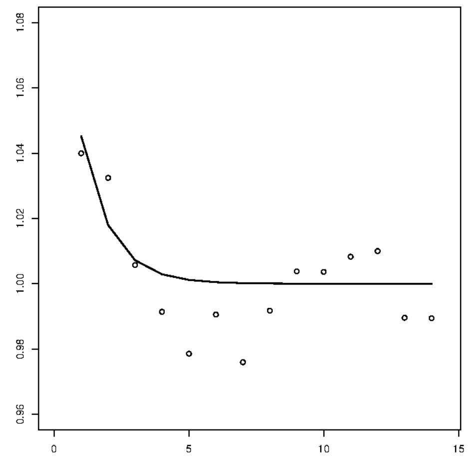
Minimum contrast estimator

Minimise (with $m = 14$)

$$\sum_{\nu=1}^{m} \left[\frac{1}{I-\nu+1} \sum_{i=\nu}^{I} \frac{N(i)N(i-\nu)}{\hat{\mu}_0(i)\hat{\mu}_0(i-\nu)} - \exp\left(\sigma^2 e^{-\beta\nu}\right) \right]^2$$

over β and σ^2 where N_i, $i = 0, \ldots, I = 729$, are the daily counts.





- **Left**: naive approach; $\hat{\sigma}^2 = 0.12$ and $\hat{\beta} = 0.55$.
- **Right**: for lag ν, consider only pairs $(N_i, N_{i-\nu})$ for which both i and $i - \nu$ are not spurious; $\hat{\sigma}^2 = 0.11$ and $\hat{\beta} = 0.91$.

Conclusions

In the naive approach

- $\hat{\mu}_0$ is under-estimated;
- therefore correlations are over-estimated;
- $\hat{\beta}$ strongly affected by gap;
- $\hat{\sigma}^2$ less so.

Moreover

- correlations vanish in about a week.

State estimation

Write $M = \{T_1 + 1, \ldots, T_2 - 1\}$ for the unrecorded days.

Goal: sample from $\mathcal{L}(N(m)_{m \in M} | n(i), i \notin M)$.

Approach:

- condition on $n(i)$ for i within a week of the missing days and sample from the posterior of $\Gamma(i)$, $i = 0, \ldots, I$;
- conditionally on the $\Gamma(i)$, sample from independent Poisson distributions.

The **log posterior pdf** $l(\gamma(i)_i \mid n(i)_{i \notin M})$ is, up to $c(\beta, \sigma^2, \mu_0, n(i)_{i \notin M})$,

$$-\frac{1}{2}\gamma(0)^2 - \sum_{i=1}^{60} \frac{\gamma(i)^2}{2(1 - e^{-2\beta})} + \sum_{i \in \{T_1 - 6, \ldots, T_1, T_2, \ldots, T_2 + 6\}} \left[n(i) S(i) - \mu_0(i) e^{S(i)} \right].$$

Metropolis–Hastings algorithm

Problem: $c(\beta, \sigma^2, \mu_0, n(i)_{i \notin M})$ is intractable.

Idea: construct a Markov chain with $\exp[l(\gamma(i)_i \mid n(i)_{i \notin M})]$ as equilibrium distribution.

In current state $\gamma(i)_i$,

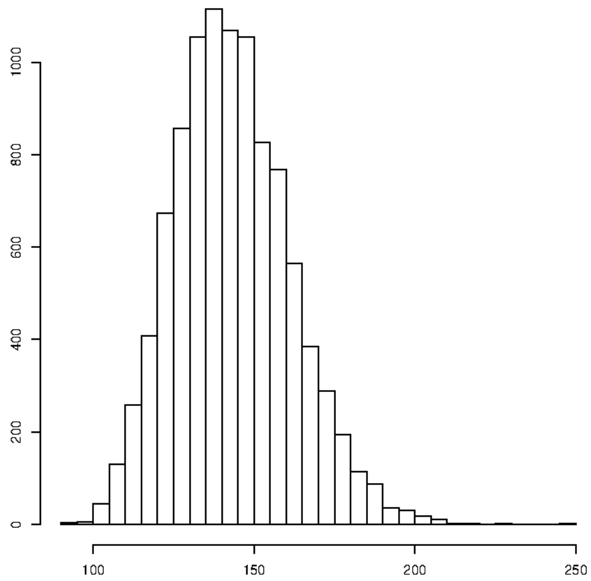
- update all $\gamma(i)$ to $\tilde{\gamma}(i) = \gamma(i) + e_i$ where e_i is a sample from

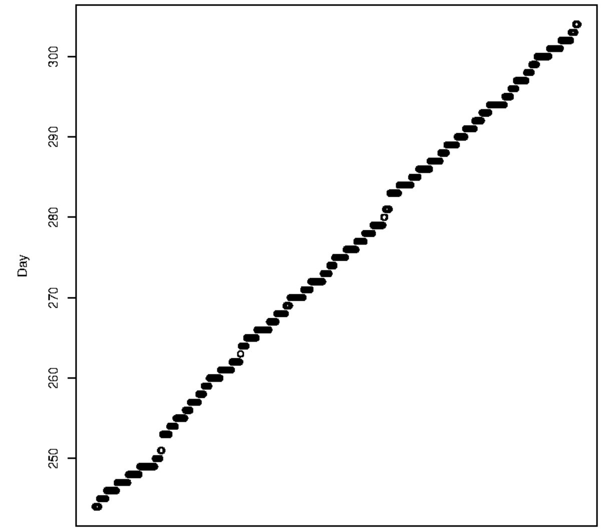
$$E_i \sim \mathcal{N}\left(\frac{h}{2}\frac{\partial}{\partial \gamma(i)} l(\gamma(j)_j \mid n(j)_{j \notin M}), \quad h\right)$$

  and write $q(\gamma(i)_i, \tilde{\gamma}(i)_i)$ for the corresponding pdf;
- accept with probability

$$\min\left(1, \frac{q(\tilde{\gamma}(i)_i, \gamma(i)_i)\, e^{l(\tilde{\gamma}(i)_i \mid n(i)_{i \notin M})}}{q(\gamma(i)_i, \tilde{\gamma}(i)_i)\, e^{l(\gamma(i)_i \mid n(i)_{i \notin M})}}\right).$$

Results





- **Left**: Histogram of expected number of missed calls ($h = 0.5$, burn-in of $10,000$ steps, sub-sampling every $1,000$ steps.
- **Right**: Imputation of missed calls (realisation after burn-in).

Aoristic data - Motivating example

When was Scrooge McDuck's lucky dime stolen?

Alternating renewal process

Let $C_i = (Y_i, Z_i)$, $i \in \mathbb{N}$, be i.i.d., and $T_i = Y_i + Z_i$. Assume that T_1 is absolutely continuous and $0 < \mathbb{E}T_1 < \infty$.

Set $S_0 = 0$ and, for $n \in \mathbb{N}$, $S_n = \sum_{i=1}^{n} T_i$. Then

$$N(t) = \sup\left\{n \in \mathbb{N}_0 : S_n \le t\right\}, \quad t \ge 0,$$

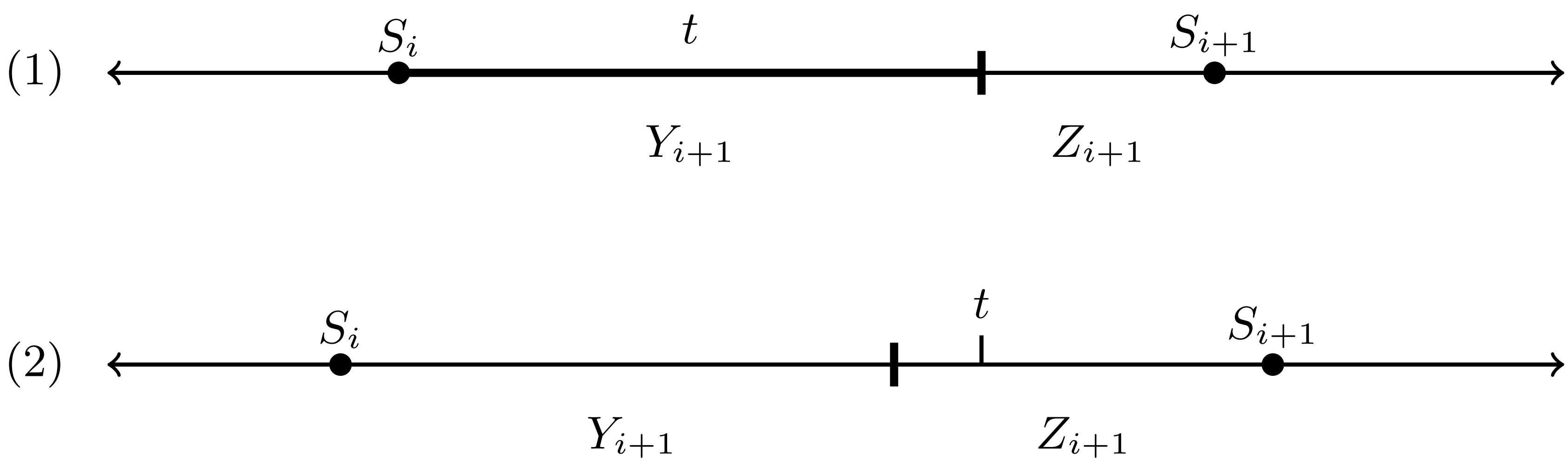
is well-defined and the supremum is attained with probability one.

The **renewal function**

$$M(t) = \mathbb{E}N(t) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \le t), \quad t \ge 0,$$

is finite and absolutely continuous.

Aoristic censoring



(1): a point $t \in X$ falls in a Y-phase. The whole interval is recorded.

(2): a point $t \in X$ falls in a Z-phase. The exact point is recorded.

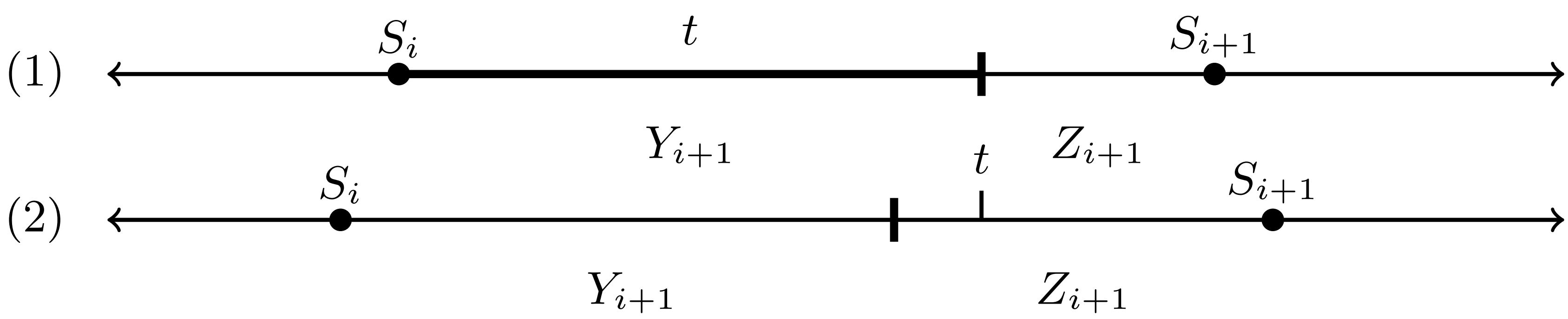
Age and excess

The **age** with respect to the Y-process is defined as

$$A(t) = (t - S_{N(t)})\,\mathbf{1}\{S_{N(t)} + Y_{N(t)+1} > t\}, \quad t \geq 0,$$

the **excess** with respect to the Y-process as

$$B(t) = (S_{N(t)+1} - Z_{N(t)+1} - t)\,\mathbf{1}\{S_{N(t)} + Y_{N(t)+1} > t\}, \quad t \geq 0.$$



The recorded interval of $t \in X$ is $t + [-A(t), B(t)]$,
mark $[-A(t), B(t)]$ parametrised by $I(t) = (-A(t), A(t) + B(t)) \in \mathbb{R} \times \mathbb{R}^+$.

Age and excess: joint distribution

Goal: limit distribution of $I(t)$ as $t \to \infty$.

Lemma

For $t \geq 0$, the joint distribution of $(A(t), B(t))$ has an atom at $(0,0)$ of size

$$c(t) = F_Y(t) - \int_0^t [1 - F_Y(t-s)]\, dM(s)$$

and, for $0 \leq u \leq t, v \geq 0$,

$$\begin{aligned} \mathbb{P}(A(t) \leq u; B(t) \leq v) &= c(t) \\ &+ [F_Y(t+v) - F_Y(t)]\, \mathbf{1}\{u = t\} \\ &+ \int_{t-u}^t [F_Y(t+v-s) - F_Y(t-s)]\, dM(s). \end{aligned}$$

Here F_Y is the cdf of Y_1.

Age and excess: long-term behaviour

Theorem

As $t \to \infty$,

$$(-A(t), A(t) + B(t)) \to^d \nu$$

where ν is the mixture of an atom at $(0, 0)$ and an absolutely continuous component that has probability density function

$$\frac{f_Y(l)}{\mathbb{E}Y_1}$$

on $\{(a, l) \in \mathbb{R} \times \mathbb{R}^+ : a \leq 0 \leq a + l\}$.

The mixture weights are $\mathbb{E}Z_1/\mathbb{E}T_1$ for the atom and $\mathbb{E}Y_1/\mathbb{E}T_1$.

Complete model formulation

Let X be a point process on open set $\mathcal{X} \subset \mathbb{R}^+$ with pdf p_X wrt a unit rate Poisson process on $\mathcal{X}$ (e.g. to model **near-repeats**).

Label the points of X independently according to ν, to obtain the **complete model** W with realisations

$$\{(t_1, I_1), \ldots, (t_n, I_n)\} \subset \mathcal{X} \times (\mathbb{R} \times \mathbb{R}^+).$$

Each $(t_j, I_j = (a_j, l_j))$ defines an interval $[t_j + a_j, t_j + a_j + l_j]$.

Censoring: observe

$$U = \bigcup_{(t,I)\in W} ((t, 0) + I).$$

Aim: reconstruct X or W from U.

CWI

UNIVERSITY OF TWENTE.

Posterior distribution

Theorem

Let $\mathbf{u}$ be a realisation of U with atomic part $\{(a_1, 0), \ldots, (a_m, 0)\}$, $m \in \mathbb{N}_0$, and non-atomic part $\{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, $n \geq m$. Then

$$\begin{aligned}
\mathbb{P}(X \in A \mid U = \mathbf{u}) \quad = \quad & c(\mathbf{u}) \int_{\mathcal{X}^{n-m}} p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \\
& \left(\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\} \right) \\
& \mathbf{1}_A(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}) \prod_{i=1}^{n-m} dx_i
\end{aligned}$$

provided that $c(\mathbf{u})$ exists in $(0, \infty)$.

Point to interval assignments

Let $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0)\} \cup \{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$.

Interpretation: the $D_1, \ldots D_{n-m}$ assign points x_{D_i} to intervals $[a_{m+i}, a_{m+i} + l_{m+i}]$.

Corollary: for $d_1, \ldots d_{n-m} \in \{1, \ldots, n-m\}$ such that $\{d_1, \ldots, d_{n-m}\} = \{1, \ldots, n-m\}$,

$$\mathbb{P}(D_1 = d_1, \ldots, D_{n-m} = d_{n-m} | X = \{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\}, U = \mathbf{u})$$

$$= \frac{\prod_{i=1}^{n-m} \mathbf{1}\{x_{d_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}}{\sum_{\substack{C_1, \ldots, C_{n-m} \\ \cup_i \{C_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{C_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}}$$

provided that $x_i \in [a_{m+i}, a_{m+i} + l_{m+i}]$ for $i = 1, \ldots, n-m$, 0 otherwise.

Example: inhomogeneous Poisson process

Let $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0)\} \cup \{(a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$.

Let X be an inhomogeneous Poisson process with intensity function $\lambda : \mathcal{X} \to \mathbb{R}^+$.

Lemma

Given $\mathbf{u}$, the points of X are independent. Each interval of $\mathbf{u}$ contains a single point with pdf

$$\frac{\lambda(x)}{\int_{[a_i, a_i + l_i] \cap \mathcal{X}} \lambda(s) ds}$$

on $[a_i, a_i + l_i] \cap \mathcal{X}$ for intervals with $l_i > 0$.

CWI

UNIVERSITY OF TWENTE.

Forward model parameters

Observe $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$, where $a_i \in \mathbb{R}$, $l_i > 0$ and $n \neq 0$.

Aim: estimate the parameters η of ν, that is,

- the parameters ζ of f_Y;
- other parameters θ involved in the joint distribution of $C_1 = (Y_1, Z_1)$.

The **log likelihood function** reads

$$L(\eta; \mathbf{u}) = m \log \left(\frac{\mathbb{E}[Z_1; \zeta, \theta]}{\mathbb{E}[T_1; \zeta, \theta]} \right) + (n - m) \log \left(\frac{\mathbb{E}[Y_1; \zeta]}{\mathbb{E} T_1 [\zeta, \theta]} \right) + \sum_{i=1}^{n-m} \log \left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]} \right)$$

upon ignoring terms that do not depend on η.

Special case

Assumption: $p = \mathbb{E}[Z_1; \zeta, \theta] / \mathbb{E}[T_1; \zeta, \theta]$ does not depend on ζ.

Then $\eta = (p, \zeta)$ and

$$L(p, \zeta; \mathbf{u}) = m \log p + (n-m) \log(1-p) + \sum_{i=1}^{n-m} \log \left(\frac{f_Y(l_i; \zeta)}{\mathbb{E}[Y_1; \zeta]} \right),$$

so $\hat{p} = m/n$.

Property: the lengths of non-degenerate intervals have pdf

$$f(l) = \frac{l f_Y(l)}{\mathbb{E} Y_1}.$$

Given $L = l$, the left-most points are uniformly distributed on $[-l, 0]$.

State estimation

Sampling from $\mathcal{L}(X|U = \mathbf{u})$ for $\mathbf{u} = \{(a_1, 0), \ldots, (a_m, 0), (a_{m+1}, l_{m+1}), \ldots, (a_n, l_n)\}$ is difficult because

- $c(\mathbf{u})$ cannot be calculated;
- the sum

$$\sum_{\substack{D_1, \ldots, D_{n-m} \\ \cup_i \{D_i\} = \{1, \ldots, n-m\}}} \prod_{i=1}^{n-m} \mathbf{1}\{x_{D_i} \in [a_{m+i}, a_{m+i} + l_{m+i}]\}$$

  is cumbersome.

Solution:

- use Metropolis–Hastings techniques
- for W rather than X
- and project on ground process X.

Metropolis–Hastings algorithm

Order points by index of $\mathbf{u}$. Assume $p_X > 0$ and $n > m$.

Recall: given $\mathbf{u}$, W has pdf proportional to $p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})$ on $\mathcal{X}^{n-m}$.

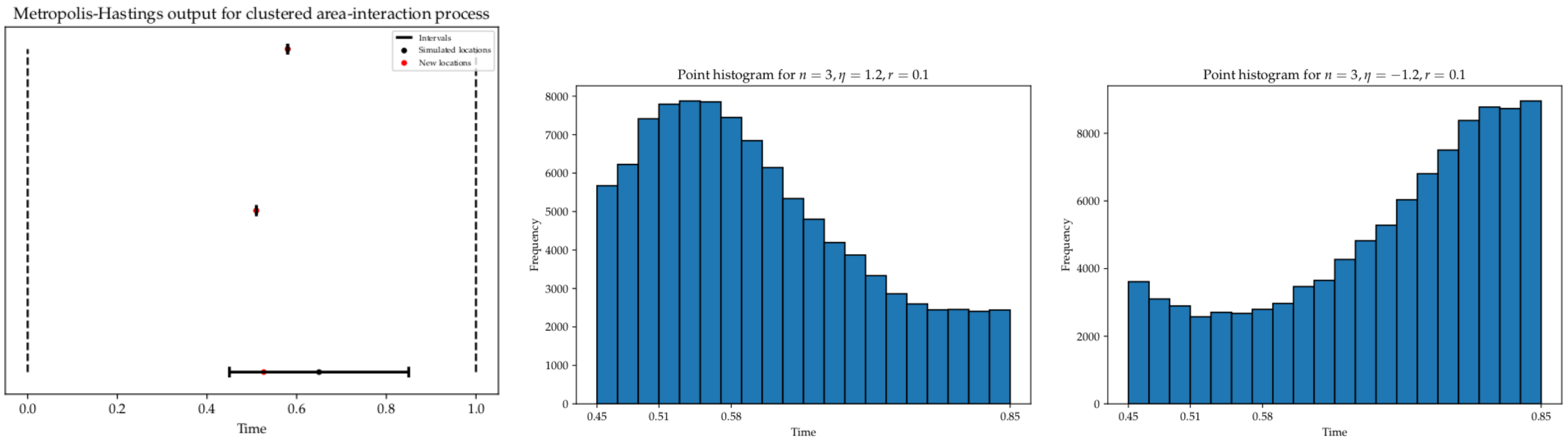
In current state $(x_1, \ldots, x_{n-m})$,

- pick an interval $[a_{m+i}, a_{m+i} + l_{m+i}]$, $i = 1, \ldots, n-m$, uniformly;
- generate y_i uniformly on $\mathcal{X} \cap [a_{m+i}, a_{m+i} + l_{m+i}]$ and propose to change x_i to y_i ;
- accept with probability

$$\min\left(1, \frac{p_X((\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\} \setminus \{x_i\}) \cup \{y_i\})}{p_X(\{a_1, \ldots, a_m, x_1, \ldots, x_{n-m}\})}\right).$$

Effect of prior

$\mathbf{u} = \{(0.51, 0), (0.58, 0), (0.45, 0.4)\}.$



Posterior distribution of X for **area-interaction model**

$$p_X(\mathbf{x}) \propto \beta^{n(\mathbf{x})} \exp\left(-\log\gamma |\cup b(x_i, r) \cap \mathcal{X}|\right)$$

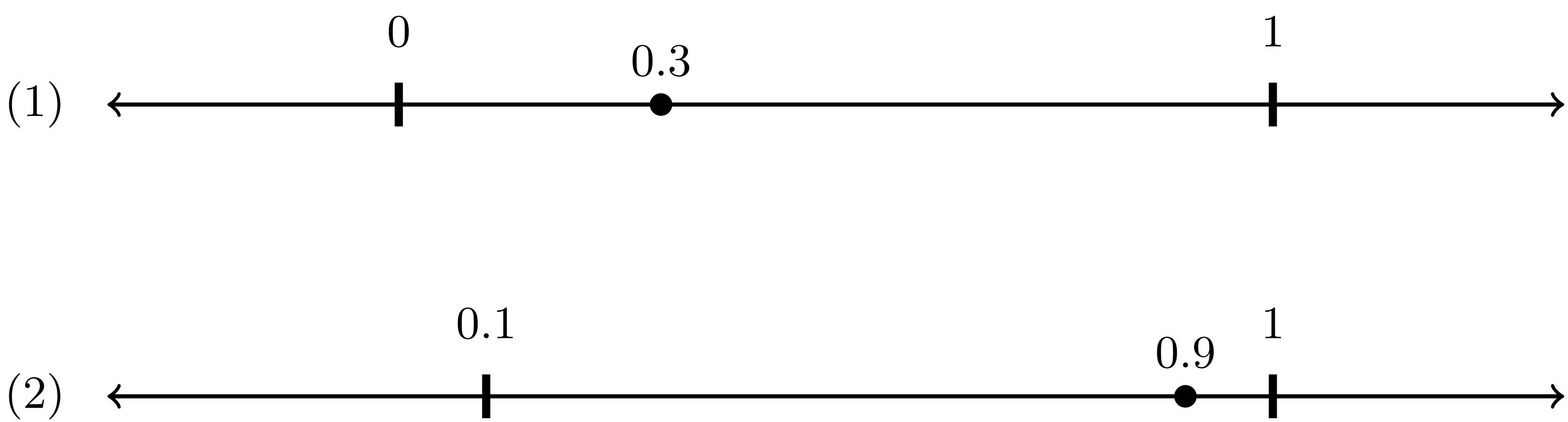
with $|\eta| = 2r \log\gamma = 1.2$ and $r = 0.1$.

Blocking states

Notes: When p_X has zeroes,

- proposals may have p_X-value zero;
- changing one component at a time might lead to non-irreducibility.

Example: if $p_X(\mathbf{x}) = 0$ when $\mathbf{x}$ contains 0.55-close points, from $(0.3, 0.9)$ one cannot reach $= (0.9, 0.3)$.



Solution: update point-interval assignments.

Summary

We

- studied inference for temporal point processes with missing information due to:
  - broken observation windows, and
  - aoristic censoring;
- combined state estimation and parametric inference;
- employed Monte Carlo methods;
- focussed on alternating renewal processes, Markov models and Cox processes.

In future, we will

- include spatial information;
- allow time-dependent home-away patterns.

References

M.N.M. van Lieshout.
Likelihood based inference for partially observed renewal processes.
Statistics and Probability Letters, 118:190–196, 2016.

M.N.M. van Lieshout and R.L. Markwitz.
State estimation for aoristic models.
Arxiv 2108.10584, August 2021.