Temporal and Spatio-temporal Point Processes

Jakob G. Rasmussen
Department of Mathematics
Aalborg University
Denmark

Contents

Spatial vs (Spatio-)temporal point processes

- In this lecture we will focus on point processes on the time line: temporal point processes.
- All the methods you have learned for spatial point processes can be used on the time line, but the fact that the time line has a natural direction opens up other possibilities.
- We will explore one such possibility: the conditional intensity function.
- We will later generalize this to marked points process, and obtain spatio-temporal point processes by letting the marks represent locations.

Temporal point pattern data

- An observed point pattern on the time line typically represents a set of events observed occuring at various points in time.
- Examples of temporal point patterns:
  - Earthquakes or other disasters
  - Visits at a server
  - Accidents at a road junction
- Temporal point processes can also have marks:
  - Magnitudes or locations of earthquakes
  - Time spent by a visitor at a server
  - Severity, cost or locations of an accident

Definition of an unmarked temporal point process

- **Definitions:**
  1. a random sequence T on $\mathbb{R}$
  2. a random counting measure N on $\mathbb{R}$
  3. a random sequence of intervals $\mathcal{T}$ on $[0, \infty)$
- **Interpretations:**
  1. Event times: $T = (\ldots, t_1, t_2, \ldots)$ is a sequence of event times.
  2. Number of events: $N(A)$ counts the number of events falling in any Borel set $A \subset \mathbb{R}$.
  3. Interevent times: $\mathcal{T} = (\ldots, \tau_1, \tau_2, \ldots)$ is a sequence of non-negative random variables interevent times.
- All three definitions are equivalent.
- 3. is rather specific to temporal point processes while 1 and 2 is similar to definitions of spatial point processes.
- Simple (or orderly) point process: We only consider cases where no events occur at the same time with probability one.
- In practice we often consider a point process defined on an interval of finite length.

The Poisson process - the simplest point process

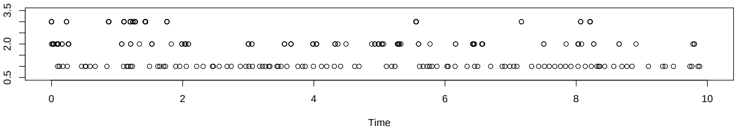
- The Poisson process can be defined just like the spatial case.
- Assume Λ locally finite measure on $\mathbb{R}$ with density λ ($\Lambda(B) = \int_B \lambda(s) \mathrm{d}s$).
- **Definition:**
  T is a Poisson process with intensity measure Λ (or intensity function λ) if for any bounded region B with $\Lambda(B) > 0$:
  1. $N(B) \sim \text{po}(\Lambda(B))$
  2. Given $N(B)$, events in B are i.i.d. with density $\propto \lambda(u)$, $u \in B$.
- If $\lambda(s) = \lambda$ is constant, the Poisson process is called homogeneous.
- Interpretation: The Poisson process is a model for events occurring along the time line independently of each other
- This process has limited practical usefulness - we need models where events can depend on each other.

Evolutionarity, history and interevent times

- *Evolutionarity*: what happens in the present depends only on the past, not the future.
- *History*: $\mathcal{H}_t = (..., t_1, t_2, ..., t_n)$ is a vector of all past events $t_i \leq t$.
- We can define a point process by the distribution of all interevent times, $\tau_n = t_n - t_{n-1}$, one at a time.
- Given the events up to event n, $\mathcal{H}_{t_n}$, denote the density function for the time of the next event by $f(t|\mathcal{H}_{t_n})$ for $t > t_n$.

Simple examples using interevent time distributions

- *Renewal process*: A point process where all the interevent times have i.i.d. distributions, i.e. $f(t_n|\mathcal{H}_{t_{n-1}}) = g(t_n - t_{n-1})$.
- For renewal processes f only depends on the last event.
- *Homogeneous Poisson process*: Special case of renewal process with exponential distributed interevent time.
- Simulations of renewal processes shown below:
  - Upper process is clustered.
  - Middle process is Poisson.
  - Lower process is regular.



Exercises

1. What kind of interevent distributions leads to renewal processes that are more clustered than a Poisson process?
2. How about more regular than a Poisson process?
3. Make a quick implementation in R of renewal processes using either of the above interevent distributions. Can you visually see the regularity/clustering?

The conditional intensity function

- More complicated processes than renewal processes are not handled well by specifying the interevent time distribution.
- Conditional intensity (/risk/rate/hazard) function:

$$\lambda^*(t) = \frac{f(t|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})}$$

- $f(t|\mathcal{H}_{t_n})$ and $F(t|\mathcal{H}_{t_n})$ are the distribution and density functions of the interevent times conditioned on the history.
- Interpretation:

$$\lambda^*(t)dt \approx \mathbb{E}[N(\mathrm{d}t)|\mathcal{H}_{t-}] \approx \mathbb{P}(N(\mathrm{d}t) > 0|\mathcal{H}_{t-})$$

- Integrated conditional intensity function: $\Lambda^*(t) = \int_0^t \lambda^*(s)\mathrm{d}s$.
- The examples so far formulated using conditional intensity:
  - Renewal process: λ^* is found directly from the above formula.
  - Poisson process: $\lambda^*(t) = \lambda(t)$ is independent of $\mathcal{H}_t$.

Exercises

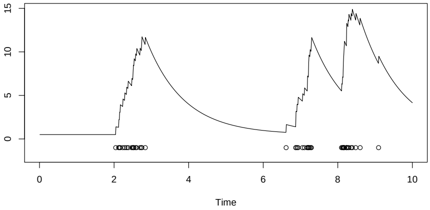
1. Prove that the homogeneous Poisson process defined by i.i.d. exponentially distributed waiting times with parameter λ has $\lambda^*(t) = \lambda$.
2. Find the intensity function for a renewal process with a uniform distribution on the interval [0,1] as interevent time distribution. Note that this increases to infinity within finite time, when no new points appear; does this intuitively make sense?

A proper example: The Hawkes process

- Hawkes process:

$$\lambda^*(t) = \mu + \alpha \sum_{t_i < t} \gamma(t - t_i)$$

where $\gamma(t)$ is a density, e.g. $\gamma(t) = \beta \exp(-\beta t)$



- μ is called immigration rate, γ is called offspring intensity, α is the mean number of offsprings, β controls the times of offsprings.
- This produces clustered point patterns.
- This is a model for reproducing populations, e.g. plants or viruses, but also reproducing in a more abstract sense, e.g. earthquakes, crimes, or terrorist attacks.

Interpreting the Hawkes process

- The Hawkes process can also be defined using a clustering and branching structure, one generation at a time.
- Define the Hawkes process as the union of the following:
  - Generation 0: Homogeneous Poisson process with intensity μ.
  - Generation 1: Each generation 0 point, say t_i, generates independent Poisson processes with intensity $\alpha\gamma(t - t_i)$
  - Generation 2: Each generation 1 point, say t_j, generates independent Poisson processes with intensity $\alpha\gamma(t - t_j)$
  - Etc. until there are no points in a generation (if $\alpha < 1$ this happens eventually with probability 1).
- Note that the intensity of each of these processes add up thus giving the correct conditional intensity $\mu + \alpha \sum_{t_i < t} \gamma(t - t_i)$.
- On the next slide, we show that the conditional intensity function uniquely defines a process, thus the two definitions are equivalent.
- The definition using generations helps us interpreting the Hawkes process, but in general it may not be easy to interpret a specific conditional intensity function.

Can we choose any function as conditional intensity?

- **Proposition:**

$$f(t|\mathcal{H}_{t_n}) = \lambda^*(t) \exp\left(-\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s\right)$$

$$F(t|\mathcal{H}_{t_n}) = 1 - \exp\left(-\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s\right)$$

- **Proposition:** A conditional intensity function $\lambda^*(t)$ uniquely defines a point process if it satisfies the following conditions for any point pattern $(\ldots, t_1, \ldots, t_n)$ and any $t > t_n$:
  1. $\lambda^*(t)$ is non-negative and integrable on any interval starting at t_n, and
  2. $\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s \to \infty$ for $t \to \infty$.

Relaxing condition 2 for the finite case

- For finite point processes (i.e. a finite number of points with probability one), we drop item 2., i.e. $\int_{t_n}^{t} \lambda^*(s)\mathrm{d}s \to \infty$ for $t \to \infty$ in the proposition.
- This leads to improper distribution functions F for the interevent time, i.e. $\lim_{t\to\infty} F(t|\mathcal{H}_{t_n}) = p < 1$.
- Interpretation: with probability $1 - p$ we get no more points.
- Example: A unit rate Poisson process on the interval has conditional intensity function $\lambda^*(t) = 1[t \in [0,1]]$, leading to the following distribution function for the first point

$$F(t|\mathcal{H}_0) = 1 - \exp\left(-\int_0^t 1[s \in [0,1]]\mathrm{d}s\right) = 1 - \exp\left(-\min\{t, 1\}\right).$$

  which has maximum value of $1 - \exp(-1) \approx 0.63$. I.e. with probability 0.63 we get one (or more) points, and probability 0.37, the process terminates with no points.

Closed form calculations with λ^*

- Some quantities predicting the future can be derived on closed form.
- For example, in the unmarked case, the probability of getting no events in the interval $(s, t]$, given that we know what happened up to time s:

$$p = 1 - F(t|\mathcal{H}_s) = \exp\left(-\int_s^t \lambda^*(u)du\right)$$

- For a Hawkes process with $\gamma(t) = \beta \exp(-\beta t)$:

$$p = \exp\left(\mu(s-t) + \alpha \sum_{t_i < s}\left(e^{-\beta(t-t_i)} - e^{-\beta(s-t_i)}\right)\right)$$

- Other quantities may be hard to calculate, such as the mean number of events within some interval.

Exercise

1. Create a conditional intensity function for a point process, where an event will reduce the chance of having events immediately after (i.e. events reduce the conditional intensity function). Check that it fulfill the conditions for being a proper conditional intensity function.
2. What kind of point pattern will your process produce (i.e. clustered, regular or neither)
3. Calculate the probability that an event will fall within the interval $(s, t]$ given that we have observed the process up and including time time s.
4. Modify your process such that it terminates after some time or condition has been met.

The marked case

- *History:* In the marked case the marks are also included in the history $\mathcal{H}_t = (\ldots, (t_1, \kappa_1), (t_2, \kappa_2), \ldots, (t_n, \kappa_n))$.
- **Definition:** Conditional intensity function:

$$\lambda^*(t, \kappa) = \lambda^*(t) f^*(\kappa | t)$$

  where $f^*(\kappa|t)$ is the density of the mark of an event at time t, possibly depending on the history and t, and $\lambda^*(t)$ is now called the ground intensity.
- Note that the ground intensity may depend on previous marks.
- Thus

$$\lambda^*(t, \kappa) = \lambda^*(t) f^*(\kappa | t) = \frac{f(t|\mathcal{H}_{t_n}) f^*(\kappa|t)}{1 - F(t|\mathcal{H}_{t_n})} = \frac{f(t, \kappa|\mathcal{H}_{t_n})}{1 - F(t|\mathcal{H}_{t_n})}$$

  Note: $F(t|\mathcal{H}_{t_n})$ is only the distribution function for time (not mark), but may still depend on previous marks.

Types of marks

- Dependence structure:
  - *Unpredictable marks:* $f^*(\kappa|t)$ does not depend on the past.
  - *Independent marks:* $f^*(\kappa|t)$ does not depend on the past or future ($=$ future does not depend on marks).
  - Note that independent marks are also unpredictable.
- The marks κ_i can belong to any probability space $\mathbb{M}$:
  - $\mathbb{M}$ is finite, corresponding to different types of points (a so-called multitype process). Here $f^*(\kappa|t)$ is the probability function.
  - $\mathbb{M} \subseteq \mathbb{N}$, again $f^*(\kappa|t)$ is the probability function.
  - $\mathbb{M} \subseteq \mathbb{R}^d$, here $f^*(\kappa|t)$ is the density function.
  - More complicated spaces...

Example: ETAS model

- ETAS: Epidemic Type Aftershock Sequence
- Data: Earthquakes with times t_i and magnitudes $\kappa_i > \kappa_0$.
- The ETAS model is a marked Hawkes process with ground intensity (this is just one version)

$$\lambda^*(t) = \mu + A \sum_{t_i < t} e^{\alpha(\kappa_i - \kappa_0)} \left(1 + \frac{t - t_i}{c}\right)^{-p}$$

  and (conditional) mark density $f^*(\kappa|t) = \delta e^{-\delta(\kappa - \kappa_0)}$.
- Parameters:
  - $\mu > 0$ controls number of main earthquakes
  - $A > 0$ controls number of aftershocks
  - $\alpha > 0$ controls the relation between magnitude and number of aftershocks
  - $c > 0$ and $p > 0$ controls times of aftershocks
  - $\delta > 0$ controls magnitudes

Another example: Stress-release model

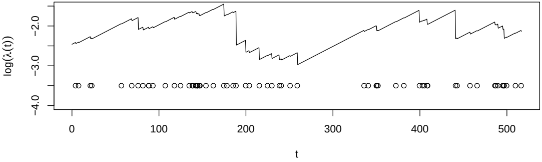
- Another model for earthquakes focuses on the underground tension, which is built-up and then released as an earthquake.
- The stress-release model (again one version) has ground intensity

$$\lambda^*(t) = e^{a+b(t-cS(t))}$$

where

$$S(t) = \sum_{t_i<t} 10^{d(\kappa_i-\kappa_0)}$$

and $a \in \mathbb{R}$ and $b, c, d > 0$ are parameters, and mark density is the same as for the ETAS model.



Exercises

1. Are the marks unpredictable, independent or neither in the ETAS and stress-release models?
2. Does the magnitude of an earthquakes influence the number and/or magnitudes of later earthquakes in the two models?
3. When in the future is the intensity at its maximum after the last observed earthquake for the two models?
4. Are the models clustered, regular or neither?
5. Would you expect both models to fit the same earthquake dataset well?

R/PtProcess

- Spatstat does not contain temporal point process modeling using the conditional intensity function, so we will use another, much smaller, package PtProcess.
- R-demo, part 1: Conditional intensity functions and models in R/PtProcess.

Spatio-temporal processes

- If we include locations as marks, i.e. $\mathbb{M} \subseteq \mathbb{R}^d$ (typically $d = 2$ or $d = 3$), we get a spatio-temporal point process.
- We can also have other marks than locations (e.g. magnitude) - then we call the process a marked spatio-temporal point process.
- The reason for treating times as the point and locations as marks is that we can then use all the theory for the conditional intensity function for point process modelling.
- Unfortunately none of the examples in PtProcess include spatial coordinates, so I do not have any practical examples to show.

Example: Spatial ETAS model

- We can include locations of earthquakes into the ETAS model.
- Earthquake i: time t_i, magnitude m_i and position (x_i, y_i).
- Conditional intensity:

$$\lambda^*(t,x,y,m) = f(m-m_0)\Big(\mu(x,y) + \sum_{t_i<t} \nu(t-t_i) g((x,y)-(x_i,y_i)) h(m_i-m_0)\Big)$$

- ν controls the temporal decay, fx. $\nu(t) = \left(1+\frac{t}{c}\right)^{-p}$
- μ controls the location of main shocks
- g controls the location of aftershocks, fx. density of a normal distribution
- h controls the effect of the magnitude on the number of aftershocks, $h(m) = e^{\alpha m}$
- f is the density of magnitudes, fx an exponential density

Likelihood function

- Observed data:
  - event times $(t_1, \ldots, t_n) \in [0, b)$
  - maybe with marks $(\kappa_1, \ldots, \kappa_n) \in \mathbb{M}$
- One way of estimating parameters in a model from real data is obtaining and maximizing the likelihood function.
- Likelihood function, unmarked case:

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i) \right) \exp(-\Lambda^*(b))$$

- Likelihood function, marked case:

$$L = \left(\prod_{i=1}^{n} \lambda^*(t_i, \kappa_i) \right) \exp\left(-\Lambda^*(b)\right).$$

Exercise

1. Prove that for the homogeneous Poisson process with intensity λ on the interval $[0, b)$, the likelihood function is given by

$$L = \lambda^n \mathrm{e}^{-\lambda b}.$$

2. Prove that the maximum of L is achieved at

$$\hat{\lambda} = \frac{n}{b}.$$

(Hint: the logarithm simplifies the problem...)

Likelihood function for the Hawkes process

- Hawkes process (with exponential offspring rate):

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \left(\mu + \sum_{t_i < t} \alpha \beta e^{-\beta(t - t_i)} \right) \times \\
&\quad \exp\left(\mu b + \alpha \left(n - e^{-\beta b} \sum_{t_i} e^{\beta t_i} \right) \right)
\end{aligned}
$$

- This cannot be maximized wrt. (μ, α, β) analytically.
- Maximum likelihood estimation: The MLE for the homogeneous Poisson process is easily found, but in general it is hard to maximise the likelihood function analytically. However, the maximum can be approximate using e.g. Newton-Raphson, since L is known and easy to calculate.

R/PtProcess

- R-demo, part 2: Maximum likelihood estimation in R/PtProcess.

Simulation of a point process - why?

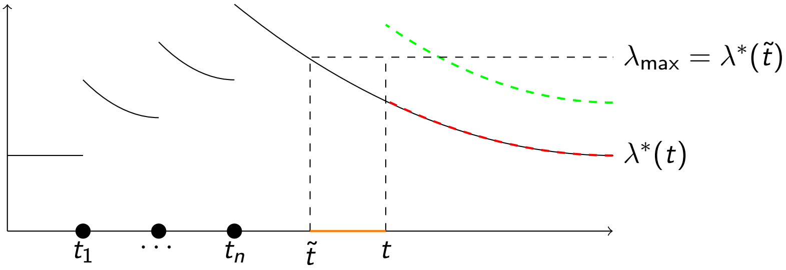
- What does a point pattern typically look like?
- Prediction.
- Model checking.
- Approximation of quantities that are hard to obtain.

Simulation

- Homogeneous Poisson process: Simulate the interevent times as independent Exp(λ) variables.
- Simulation of general point processes specified by a conditional intensity function can also be done by simulating exponential variables one at a time, and afterwards modify these this using the conditional intensity function.
- Two approaches:
  1. Thinning: A newly simulated point is removed with some probability.
  2. Inversion: A newly simulation point is moved according to some distribution.

Simulation using Ogata's modified thinning algorithm

- Unmarked case: We use dependent thinning recursively:
  1. Generate next potential point $\tilde{t}$ from $t - \tilde{t} \sim$Exp($\lambda_{\max}$).
  2. Generate $U \sim$ Unif$(0, 1)$.
  3. If $U \leq \lambda^*(t)/\lambda_{\max}$ keep point, i.e. $t_{n+1} = t$; otherwise throw point away.
  4. Let $\tilde{t} := t$, and start over.



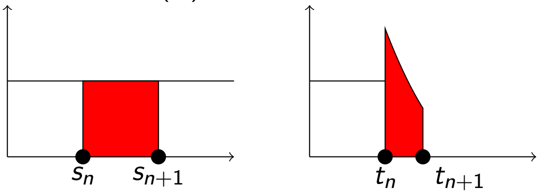
- Marked case: every time a new point t_i is kept, the mark should be simulated using the mark density $f^*(\kappa_i|t_i)$.

Exercises

1. Consider an unmarked version of the stress-release model with conditional intensity function $\lambda^*(t) = \exp(a + bt - cN(t))$ where $a \in \mathbb{R}$ and $b, c > 0$ are parameters, and $N(t) = \sum_{t_i < t} 1$ is the number of points strictly before time t. Make an illustration of the conditional intensity function for some given point pattern.
2. Note that the conditional intensity rises unboundedly in intervals with no points; why is this problematic for the Ogata's modified thinning algorithm?
3. One solution is to include a maximum distance forward in time that we will go from time t, say $l(t)$, so that the conditional intensity is bounded in this interval, and we can define a $\lambda_{\max}$ here. Explain how to simulate the above process using this (hint: if the simulated exponential variable goes beyond $t + l(t)$ there simply is no point in the interval $[t, t + l(t)]$ and the algorithm starts again at $t + l(t)$).

Simulation using inversion (or transformation)

- **Inversion Theorem (first half):**
  If $\{s_i\}_{i\in\mathbb{Z}}$ is a unit rate Poisson process on $\mathbb{R}$, and $t_i = \Lambda^{*-1}(s_i)$, then $\{t_i\}_{i\in\mathbb{Z}}$ is a point process with cif $\lambda^*(t_i)$.
- Finite process: Poisson process on $[0, b)$ transforms into point process on $[0, \Lambda^{*-1}(b))$ with intensity $\lambda^*(t_i)$.
- *Simulation by inversion:* Simulate $s_i \sim \text{Exp}(1)$ and transform to $t_i = \Lambda^{*-1}(s_i)$ one at a time until the end is reached.



- Note: Λ^{*-1} is often not available on closed form - numerical approximation.
- Marked case: every time a new point t_i is simulated, the mark should be simulated using the mark density $f^*(\kappa_i|t_i)$.

Exercises

1. Consider again the process with conditional intensity function $\lambda^*(t) = \exp(a + bt - cN(t))$. Given the point pattern before t, say $t_1, \ldots, t_n$, show that the integrated conditional intensity function is given by

$$\Lambda^*(t) = \sum_{i=1}^{n+1} \mathrm{e}^{a-c(i-1)} \frac{1}{b} \left(\mathrm{e}^{bt_i} - \mathrm{e}^{bt_{i-1}} \right)$$

   where $t_0 = 0$ and $t_{n+1} = t$.

2. Note that t is only present in the last term of the sum, and find Λ^{*-1}. Explain how to simulate the process using simulation by inversion.

R/PtProcess

- R-demo, part 3: Simulation in R/PtProcess.

Model checking: Residual analysis

- **Inversion Theorem (second half):**
  If $\{t_i\}_{i\in\mathbb{Z}}$ is a point process with intensity $\lambda^*(t_i)$, and $s_i = \Lambda^*(t_i)$, then $\{s_i\}_{i\in\mathbb{Z}}$ is a unit rate Poisson process.
- *Residual analysis:* transform data using $\hat{\Lambda}^*$ (estimated Λ^*) - if the model fits well, then $\{s_i\}_{i\in\mathbb{Z}} \approx$ unit-rate Poisson.
- Algorithm: Calculate $s_i = \hat{\Lambda}^*(t_i)$ where $\{t_i\}_{i\in\mathbb{Z}}$ is the data, and check whether $\tau_i = s_i - s_{i-1}$ looks like i.i.d. Exp(1) variables, e.g.:
  - QQ-plot or histogram to see whether they look exponential
  - Plot τ_i vs. τ_{i-1} to see if there is any pattern (i.e. dependence)
- Any discrepancy from a unit-rate Poisson process may give information about how the model does not fit the data, e.g. if $\{t_i\}_{i\in\mathbb{Z}}$ is too clustered for a chosen model, then $\{s_i\}_{i\in\mathbb{Z}}$ is too clustered for a Poisson process.

R/PtProcess

- R-demo, part 4: Residual analysis in R/PtProcess.

Everything needed for a statistical analysis seems to be covered now...

Typical steps in a statistical analysis:

- Preliminary analysis (simulation of potential models)
- Model specification and interpretation
- Parameter estimation (maximum likelihood)
- Model checking (residual analysis or simulation based approaches)
- Prediction - simulation or calculation using λ^*

Analysis of data

In the remaining part of this lecture, you should analyse a dataset.

1. Fit point process models (e.g. `etas_gif`, `srm_gif`, `simple_gif`) to datasets (e.g. `Ogata`, `NtChina`, `Tangshan`) in `PtProcess`.
2. The analysis should include model specification, parameter estimation, residual analysis, simulation, and conclusions about the data.