Thinning-stable point processes as a model for bursty spatial data

Sergei Zuyev

Chalmers University of Technology, Gothenburg, Sweden

Paris, Jan 14th 2015
Fixed line telephony

- Scientific language of telecommunications since the start of XX century has been Queueing Theory (Erlang, Palm, Kleinrock, et al.)
Fixed line telephony

- Scientific language of telecommunications since the start of XX century has been Queueing Theory (Erlang, Palm, Kleinrock, et al.)
- Basic model: Poisson arrivals temporal process (1D point process).
Why Poisson?

**Poisson limit theorem:** If $\Phi_n$ are i.i.d. point processes with $E \Phi_i(B) = \mu(B) < \infty$ for any bounded $B$ and $t \circ \Phi_i$, $t \in (0, 1]$ denotes independent $t$-thinning of its points, then

$$\frac{1}{n} \circ (\Phi_1 + \cdots + \Phi_n) \rightarrow \Pi,$$

where $\Pi$ is a Poisson PP with intensity measure $\mu$. 
Burstiness!

Crucial assumption: $E \Phi_i(B) = \mu(B) < \infty$ roughly means workload associated with points (duration of calls) is fairly constant.
Burstiness!

- **Crucial assumption**: \( E \Phi_i(B) = \mu(B) < \infty \) roughly means **workload** associated with points (duration of calls) is fairly constant.

- **SMS message** \( \sim 10^2 \) bytes of data, **video download** \( \sim 10^{10} \) bytes: 8-order magnitude difference!

- **Addressing burstiness** in time: Heavy-tailed traffic queueing, Fractional BM, etc.
Performance of modern telecommunications systems is strongly affected by their spatial structure. Spatial Poisson PP as a model for structuring elements of telecom networks: E.N. Gilbert, Salai, Baccelli, Klein, Lebourges & Z
What is *random* in stations’ position?
What is *random* in stations’ position?
Challenge: spatial burstiness
**Definition**

A random vector $\xi$ (generally, a random element on a convex cone) is called **strictly $\alpha$-stable** (notation: $\text{St}_\alpha S$) if for any $t \in [0, 1]$

\[
t^{1/\alpha} \xi' + (1 - t)^{1/\alpha} \xi'' \overset{D}{=} \xi,
\]

where $\xi'$ and $\xi''$ are independent copies of $\xi$.

**Stability and CLT**

Only $\text{St}_\alpha S$ vectors $\xi$ can appear as a weak limit

\[
n^{-1/\alpha} (\zeta_1 + \cdots + \zeta_n) \Rightarrow \xi.
\]
A point process $\Phi$ (or its probability distribution) is called **discrete $\alpha$-stable or $\alpha$-stable with respect to thinning** (notation $D\alpha S$), if for any $0 \leq t \leq 1$

$$t^{1/\alpha} \circ \Phi' + (1 - t)^{1/\alpha} \circ \Phi'' \overset{D}{=} \Phi,$$

where $\Phi'$ and $\Phi''$ are independent copies of $\Phi$ and $t \circ \Phi$ is independent thinning of its points with retention probability $t$. 

**Definition**
Let $\Psi_1, \Psi_2, \ldots$ be a sequence of i.i.d. point processes and $S_n = \sum_{i=1}^{n} \Psi_i$. If there exists a PP $\Phi$ such that for some $\alpha$ we have

$$n^{-1/\alpha} \circ S_n \Longrightarrow \Phi \quad \text{as } n \to \infty$$

then $\Phi$ is $D_{\alpha}S$.

**CLT**

When intensity measure of $\Psi$ is $\sigma$-finite, then $\alpha = 1$ and $\Phi$ is a Poisson processes. Otherwise, $\Phi$ has infinite intensity measure – bursty.
Cox process

Let $\xi$ be a random measure on the space $X$. A point process $\Phi$ on $X$ is a **Cox process** directed by $\xi$, when, conditional on $\xi$, realisations of $\Phi$ are those of a Poisson process with intensity measure $\xi$. 
Characterisation of $D\alpha S$ PP

Theorem

A PP $\Phi$ is a (regular) $D\alpha S$ iff it is a Cox process $\Pi_\xi$ with a $St\alpha S$ intensity measure $\xi$, i.e. a random measure satisfying

$$t^{1/\alpha}\xi' + (1 - t)^{1/\alpha}\xi'' \overset{D}{=} \xi.$$ 

Its p.g.f.l. is given by

$$G_\Phi[u] = \mathbb{E} \prod_{x_i \in \Phi} u(x_i) = \exp - \int_{\mathbb{B} M_1} \langle 1 - u, \mu \rangle^\alpha \sigma(d\mu), \quad 1 - u \in \mathbb{B} M$$

for some locally finite spectral measure $\sigma$ on the set $\mathbb{M}_1$ of probability measures.

$D\alpha S$ PPs exist only for $0 < \alpha \leq 1$ and for $\alpha = 1$ these are Poisson.
A r.v. $\gamma$ has Sibuya distribution, $Sib(\alpha)$, if

$$g_{\gamma}(s) = 1 - (1 - s)^{\alpha}, \quad \alpha \in (0, 1).$$

It corresponds to the number of trials to get the first success in a series of Bernoulli trials with probability of success in the $k$th trial being $\alpha/k$. 

Sibuya point processes

Definition

Sergei Zuyev

Thinning-stable point processes as a model for bursty spatial data
A r.v. $\gamma$ has **Sibuya distribution**, $\text{Sib}(\alpha)$, if

$$g_\gamma(s) = 1 - (1 - s)^\alpha, \quad \alpha \in (0, 1).$$

It corresponds to the number of trials to get the first success in a series of Bernoulli trials with probability of success in the $k$th trial being $\alpha/k$.

Let $\mu$ be a probability measure on $X$. The point process $\Upsilon$ on $X$ is called the **Sibuya point process** with exponent $\alpha$ and parameter measure $\mu$ if $\Upsilon(X) \sim \text{Sib}(\alpha)$ and each point is $\mu$-distributed independently of the other points. Its distribution is denoted by $\text{Sib}(\alpha, \mu)$.
Examples of Sibuya point processes

Figure: Sibuya processes: $\alpha = 0.4$, $\mu \sim \mathcal{N}(0, 0.3^2 I)$
**Theorem** Davydov, Molchanov & Z’11

Let $\mathcal{M}_1$ be the set of all probability measures on $X$. A regular $D\alpha S$ point process $\Phi$ can be represented as a **cluster process** with

- **Poisson centre process** on $\mathcal{M}_1$ driven by intensity measure $\sigma$;
- **Component processes** being Sibuya processes $\text{Sib}(\alpha, \mu)$, $\mu \in \mathcal{M}_1$. 

**D$_{\alpha}$S point processes as cluster processes**
Statistical Inference for $D\alpha S$ processes

We assume the observed realisation comes from a stationary and ergodic $D\alpha S$ process without multiple points.
We assume the observed realisation comes from a stationary and ergodic D$\alpha$S process without multiple points.

Such processes are characterised by:

- $\lambda$ – the Poisson parameter: mean number of clusters per unit volume
- $\alpha$ – the stability parameter
We assume the observed realisation comes from a stationary and ergodic $D_{\alpha}S$ process without multiple points.

Such processes are characterised by:

- $\lambda$ – the Poisson parameter: mean number of clusters per unit volume
- $\alpha$ – the stability parameter
- A probability distribution $\sigma_0(d\mu)$ on $M_1$ (the distribution of the Sibuya parameter measure)
Construction

1. Generate a homogeneous Poisson PP $\sum_{i} \delta_{y_i}$ of centres of intensity $\lambda$;
2. For each $y_i$ generate independently a probability measure $\mu_i$ from distribution $\sigma_0$;
3. Take the union of independent Sibuya clusters $\text{Sib}(\alpha, \mu_i(\cdot - y_i))$.
Example of $D_\alpha S$ point process

Figure: $\lambda = 0.4$, $\alpha = 0.6$, $\sigma_0 = \delta\mu$, where $\mu \sim \mathcal{N}(0, 0.3^2 I)$
Parameters to estimate

Consider the case when all the clusters have the same distribution, so that $\sigma_0 = \delta_\mu$ for some $\mu \in \mathbb{M}_1$.

We always need to estimate $\lambda$ and $\alpha$, often also $\mu$. 
Parameters to estimate

Consider the case when all the clusters have the same distribution, so that $\sigma_0 = \delta\mu$ for some $\mu \in \mathbb{M}_1$.

We always need to estimate $\lambda$ and $\alpha$, often also $\mu$.

We consider three possible cases for $\mu$:

- $\mu$ is already known
- $\mu$ is unknown but lies in a parametric class (e.g. $\mu \sim \mathcal{N}(0, \sigma^2I)$ or $\mu \sim U(B_r(0)))$
- $\mu$ is totally unknown
Estimation of $\mu$

Idea

Identifying a big cluster in the dataset and using it to estimate $\mu$. 
Estimation of $\mu$

Idea

Identifying a big cluster in the dataset and using it to estimate $\mu$.

How to distinguish clusters in the configuration? How to identify at least the biggest clusters?
Estimation of $\mu$

Idea

Identifying a big cluster in the dataset and using it to estimate $\mu$.

How to distinguish clusters in the configuration? How to identify at least the biggest clusters?

- Interpreting data as a mixture model
Estimation of $\mu$

Idea
Identifying a big cluster in the dataset and using it to estimate $\mu$.

How to distinguish clusters in the configuration? How to identify at least the biggest clusters?

- Interpreting data as a mixture model
- Expectation-Maximisation algorithm
Estimation of $\mu$

Idea

Identifying a big cluster in the dataset and using it to estimate $\mu$.

How to distinguish clusters in the configuration? How to identify at least the biggest clusters?

- Interpreting data as a mixture model
- Expectation-Maximisation algorithm
- Bayesian Information Criterion
Example: gaussian spherical clusters, 2D case

(a) Original process
(b) Clustered process

Figure: \( D_{\alpha}S \) process with Gaussian clusters: \( \lambda = 0.5, \alpha = 0.6 \), covariance matrix \( 0.1^2 I \). \texttt{mclust} R-procedure with Poisson noise.
After we single out one big cluster:

- we estimate $\mu$ using kernel density or we just use the sample measure
After we single out one big cluster:

- we estimate $\mu$ using kernel density or we just use the sample measure
- if $\mu$ is in a parametric class we estimate the parameters
Overlapping clusters - heavy thinning approach

Figure: \( \lambda = 0.4, \alpha = 0.6, \mu_x \sim \mathcal{N}(x, 0.5^2 I) \)
Estimation of $\lambda$ and $\alpha$

When $\mu$ is known or have already been estimated, we suggest these

Estimation methods for $\lambda$ and $\alpha$

1. via void probabilities
Estimation of $\lambda$ and $\alpha$

When $\mu$ is known or have already been estimated, we suggest these

- **Estimation methods for $\lambda$ and $\alpha$**
  - 1. via void probabilities
  - 2. via the p.g.f. of the counts distribution
The void probabilities (which characterise the distribution of a simple point process) are given by

\[
\mathbb{P}\{\Phi(B) = 0\} = \exp\left\{-\lambda \int_A \mu(B)^\alpha \, dx\right\}.
\]
Unbiased estimator for the void probability function

Let \( \{x_i\}_{i=1}^n \subseteq A \) a sequence of test points and \( r_i = \text{dist}(x_i, \text{supp } \Phi) \), then

\[
\hat{p}(r) = \frac{1}{n} \sum_{i=1}^n 1_{\{r_i > r\}}
\]

is an unbiased estimator for \( P\{\Phi(B_r(0)) = 0\} \).

Then \( \alpha \) and \( \lambda \) are estimated by the best fit to this curve.
Example: uniformly distributed clusters, 1D case

Figure: $\lambda = 0.3$, $\alpha = 0.7$, $\mu \sim U(B_1(0))$, $|A| = 3000$

Estimated values: $\hat{\lambda} = 0.29$, $\hat{\alpha} = 0.68$. Requires big data!
Void probabilities for thinned processes

**p.g.fl. of $D\alpha S$ processes**

$$G_{\Phi}[h] = \exp \left\{ - \int_{\mathbb{R}} \langle 1-h, \mu \rangle^\alpha \sigma(d\mu) \right\}, \quad 1-h \in \text{BM}(X).$$

**p.g.fl. of a $p$-thinned point process**

$$G_{p\circ \Phi}[h] = \exp \left\{ - p^\alpha \int_{\mathbb{R}} \langle 1-h, \mu \rangle^\alpha \sigma(d\mu) \right\}, \quad p \in [0, 1], \quad 1-h \in \text{BM}(X).$$

$$\sigma\left(\{\mu(\cdot - x), \, x \in B\}\right) = \lambda \cdot |B| \quad \Rightarrow \quad \alpha_{\text{new}} = \alpha, \quad \lambda_{\text{new}} = \lambda \cdot p^\alpha.$$
There is no need to simulate $p$-thinning!
Let $r_k$ be the distance from 0 to the $k$-th closest point in the configuration.
Estimation via thinned process

\[ \mathbb{P}\{(p \circ \Phi)(B_r(0)) = 0\} \]

\[ \chi^\Phi \]

\[ = \mathbb{P}\{ \text{"the closest survived point is the k-th"} \} \mathbb{P}\{ r_k > r \} \]

\[ = \sum_{k=1}^1 \mathbb{P}\{ r_k > r \} \]

\[ \chi^\Phi \]

\[ = p(1 - p)^{k-1} \mathbb{P}\{ r_k > r \} \]

\[ = \sum_{k=1}^1 \mathbb{P}\{ r_k > r \} \]

Unbiased estimator for the void probability function

Let \( \{x_i\}_{n_i=1} \subseteq A \) a sequence of test points and \( r_i, k \) be the distance from \( x_i \) to its \( k \)-closest point of \( \text{supp} \Phi \). Then

\[ \hat{G}(r) = \frac{1}{n} \sum_{i=1}^n \sum_{k=0}^{\infty} p(1 - p)^{k-1} \mathbb{P}\{ r_i, k > r \} \]

is an unbiased estimator for \( \mathbb{P}\{ \Phi(B_r(0)) = 0 \} \).
Estimation via thinned process

\[
\mathbb{P}\{(p \circ \Phi)(B_r(0)) = 0\} \\
= \sum_{k=1}^{\infty} \left\{ \text{“the closest survived point is the k-th”} \right\} \mathbb{P}\{r_k > r\} \\
= p(1 - p)^{k-1} \mathbb{P}\{r_k > r\} \\
\]

Unbiased estimator for the void probability function

Let \( \{x_i\}_{i=1}^{n} \subseteq A \) a sequence of test points and \( r_{i,k} \) be the distance from \( x_i \) to its \( k \)-closest point of \( \text{supp} \Phi \). Then

\[
\hat{\mathbb{G}}(r) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=0}^{\infty} p(1 - p)^{k-1} \mathbb{1}_{\{r_{i,k} > r\}}
\]
Example: uniform clusters, 1D case

**Figure**: Estimation of v.p. of the thinned process for a process generated with $\lambda = 0.3$, $\alpha = 0.7$, $\mu \sim U(B_1(0))$, $|A| = 1000$

Estimated values: $\hat{\lambda} = 0.29$, $\hat{\alpha} = 0.72$
Counts distribution

Putting \( u(x) = 1 - (1 - s) \mathbb{I}_B(x) \) with \( s \in [0, 1] \), in the p.g.f.l. expression, we get the p.g.f. of the counts \( \Phi(B) \) for any set \( B \):

\[
\psi_{\Phi(B)}(s) := \mathbb{E}[s^{\Phi(B)}] = \exp \left\{ - (1 - s)^\alpha \int_S \mu(B)^\alpha \sigma(d\mu) \right\}.
\]  

(2)

It is a heavy-tailed distribution with \( \mathbb{P}\{\Phi(B) > x\} = L(x) x^{-\alpha} \), where \( L \) is slowly-varying.
Estimation via counts distribution

The empirical p.g.f. is then

$$\hat{\psi}_{\Phi(B)}(s) := \frac{1}{n} \sum_{i=1}^{n} s^{\Phi(B_i)} \quad \forall s \in [0, 1],$$

where $B_i, i = 1, \ldots, n$, are translates of a fixed reference set $B$ and it is an unbiased estimator of $\psi_{\Phi(B)}$. It is then fitted to (2) for a range of $s$ estimating $\lambda$ and $\alpha$.

We also tried the Hill plot from extremal distributions inference to estimate $\alpha$, but the results were poor!
Simulation studies looked at the bias and variance in the estimation of $\alpha$, $\lambda$ in different situations:

- Big sample – moderate sample
- Overlapping clusters (large $\lambda$) – separate clusters (small $\lambda$)
- Heavy clusters (small $\alpha$) – moderate clusters ($\alpha$ close to 1)
Best methods

- The simplest void probabilities method is preferred for large datasets or for moderate datasets with separated clusters. It best estimates $\alpha$, but in the latter case $\lambda$ is best estimated by counts p.g.f. fitting.
Best methods

- The simplest void probabilities method is preferred for large datasets or for moderate datasets with separated clusters. It best estimates $\alpha$, but in the latter case $\lambda$ is best estimated by counts p.g.f. fitting.

- $\lambda$ is best estimated by void probabilities with thinning method which produces best estimates in all the situations apart from moderate separated clusters. But it is also more computationally expensive.
The simplest void probabilities method is preferred for large datasets or for moderate datasets with separated clusters. It best estimates $\alpha$, but in the latter case $\lambda$ is best estimated by counts p.g.f. fitting.

$\lambda$ is best estimated by void probabilities with thinning method which produces best estimates in all the situations apart from moderate separated clusters. But it is also more computationally expensive.

As common in modern Statistics, all methods should be tried and consistency in estimated values gives more trust to the model.
Fête de la Musique data

Figure: Estimated $\hat{\alpha} = 0.17 - 0.28$ depending on the way base stations records are extrapolated to spatial positions of callers
Generalisations

For the Paris data we observed a bad fit of cluster size to Sibuya distribution. Possible cure:

**F-stable point processes** when thinning is replaced by more general subcritical branching operation. Multiple points are now also allowed.
References


2. S. Crespi, B. Spinelli and SZ *Inference for discrete stable point processes* (under preparation)

3. G. Zanella and SZ *F-stable point processes* (under preparation)
Thank you!

Questions?