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

Sergei Zuyev

Chalmers University of Technology, Gothenburg, Sweden

Paris, Jan 14th 2015

Communications Science. XXth Century

Fixed line telephony

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

Communications Science. XXth Century

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 Φ_n are i. i. d. point processes with $\mathbf{E}\,\Phi_i(B) = \mu(B) < \infty$ for any bounded B and $t \circ \Phi_i,\ t \in (0,1]$ denotes independent t-thinning of its points, then

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

where Π is a Poisson PP with indensity measure μ.

Limitation of Poisson framework

Burstiness!

- Crucial assumption: $\mathbf{E}\,\Phi_i(B) = \mu(B) < \infty$ roughly means workload associated with points (duration of calls) is fairy constant.

Limitation of Poisson framework

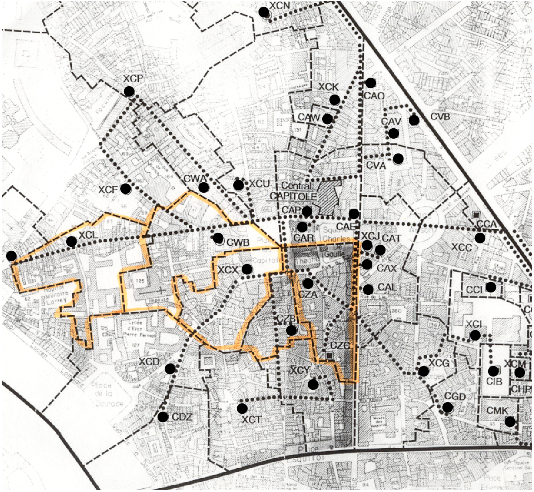
Burstiness!

- Crucial assumption: $\mathbf{E}\,\Phi_i(B) = \mu(B) < \infty$ roughly means workload associated with points (duration of calls) is fairy 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.

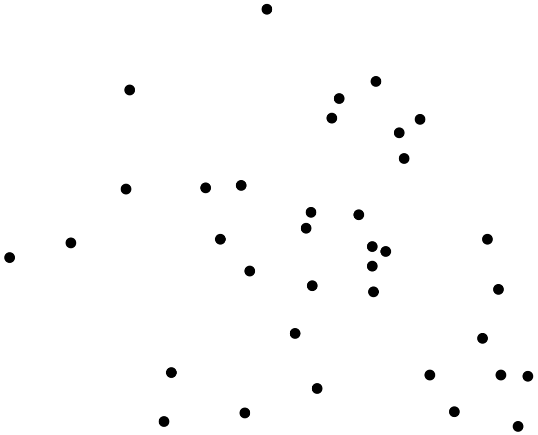
Late XXth Century

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

Stability

Definition

A random vector ξ (generally, a random element on a convex cone) is called strictly α-stable (notation: StαS) if for any $t \in [0, 1]$

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

where ξ' and ξ'' are independent copies of ξ.

Stability and CLT

Only StαS vectors ξ can appear as a weak limit $n^{-1/\alpha}(\zeta_1 + \cdots + \zeta_n) \Longrightarrow \xi$.

DαS point processes

Definition

A point process Φ (or its probability distribution) is called discrete α-stable or α-stable with respect to thinning (notation DαS), if for any $0 \le t \le 1$

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

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

Discrete stability and limit theorems

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 Φ such that for some α we have

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

then Φ is DαS.

CLT

When intensity measure of Ψ is σ-finite, then $\alpha = 1$ and Φ is a Poisson processes. Otherwise, Φ has infinite intensity measure – bursty

DαS point processes and StαS random measures

Cox process

Let ξ be a random measure on the space X. A point process Φ on X is a Cox process directed by ξ, when, conditional on ξ, realisations of Φ are those of a Poisson process with intensity measure ξ.

Characterisation of DαS PP

Theorem

A PP Φ is a (regular) DαS iff it is a Cox process Π_ξ with a StαS intensity measure ξ, i.e. a random measure satisfying

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

Its p.g.fl. is given by

$$G_\Phi[u] = \mathbf{E}\prod_{x_i\in\Phi} u(x_i) = \exp\Big\{-\int_{\mathbb{M}_1} \langle 1-u,\mu\rangle^\alpha \sigma(d\mu)\Big\}\,, \quad 1-u\in \mathrm{BM}$$

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

DαS PPs exist only for $0<\alpha\le 1$ and for $\alpha=1$ these are Poisson.

Sibuya point processes

Definition

A r.v. γ has Sibuya distribution, Sib(α), if

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

It corresponds to the number of trials to get the first success in a series of Bernoulli trials with probability of success in the kth trial being α/k.

Sibuya point processes

Definition

A r.v. γ has Sibuya distribution, Sib(α), if

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

It corresponds to the number of trials to get the first success in a series of Bernoulli trials with probability of success in the kth trial being α/k.

Sibuya point processes

Let μ be a probability measure on X. The point process Υ on X is called the Sibuya point process with exponent α and parameter measure μ if $\Upsilon(X) \sim$ Sib(α) and each point is μ-distributed independently of the other points. Its distribution is denoted by Sib(α, μ).

Examples of Sibuya point processes

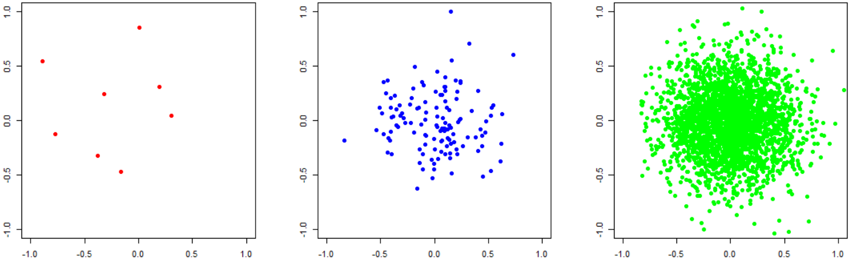


Figure : Sibuya processes: $\alpha = 0.4$, $\mu \sim \mathcal{N}(0, 0.3^2 \mathrm{I})$

DαS point processes as cluster processes

Theorem Davydov, Molchanov & Z'11

Let $\mathbb{M}_1$ be the set of all probability measures on X. A regular DαS point process Φ can be represented as a cluster process with

- Poisson centre process on $\mathbb{M}_1$ driven by intensity measure σ;
- Component processes being Sibuya processes Sib(α, μ), $\mu \in \mathbb{M}_1$.

Statistical Inference for DαS processes

We assume the observed realisation comes from a stationary and ergodic DαS process without multiple points.

Statistical Inference for DαS processes

We assume the observed realisation comes from a stationary and ergodic DαS process without multiple points.

Such processes are characterised by:

- λ – the Poisson parameter: mean number of clusters per unit volume
- α – the stability parameter

Statistical Inference for DαS processes

We assume the observed realisation comes from a stationary and ergodic DαS process without multiple points.

Such processes are characterised by:

- λ – the Poisson parameter: mean number of clusters per unit volume
- α – the stability parameter
- A probability distribution $\sigma_0(d\mu)$ on $\mathbb{M}_1$ (the distribution of the Sibuya parameter measure)

Construction

1. Generate a homogeneous Poisson PP $\sum_i \delta_{y_i}$ of centres of intensity λ;
2. For each y_i generate independently a probability measure μ_i from distribution σ_0;
3. Take the union of independent Sibuya clusters Sib$(\alpha, \mu_i(\bullet - y_i))$.

Example of DαS point process

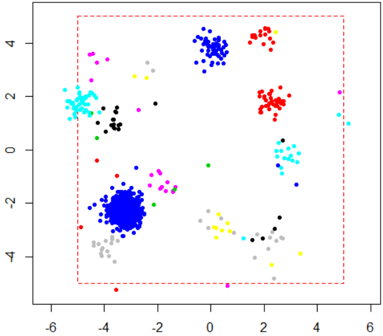


Figure : $\lambda = 0.4$, $\alpha = 0.6$, $\sigma_0 = \delta_\mu$, where $\mu \sim \mathcal{N}(0, 0.3^2\mathrm{I})$

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 λ and α, often also μ.

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 λ and α, often also μ.

We consider three possible cases for μ:

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

Estimation of μ

Idea

Identifying a big cluster in the dataset and using it to estimate μ.

Estimation of μ

Idea

Identifying a big cluster in the dataset and using it to estimate μ.

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

Estimation of μ

Idea

Identifying a big cluster in the dataset and using it to estimate μ.

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

- Interpreting data as a mixture model

Estimation of μ

Idea

Identifying a big cluster in the dataset and using it to estimate μ.

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 μ

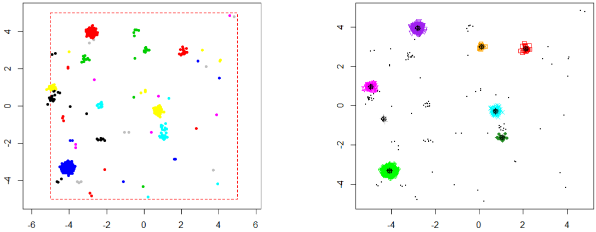
Idea

Identifying a big cluster in the dataset and using it to estimate μ.

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 : DαS process with Gaussian clusters: $\lambda = 0.5$, $\alpha = 0.6$, covariance matrix $0.1^2\mathrm{I}$. `mclust` R-procedure with Poisson noise.

Estimation of μ

After we single out one big cluster:

- we estimate μ using kernel density or we just use the sample measure

Estimation of μ

After we single out one big cluster:

- we estimate μ using kernel density or we just use the sample measure
- if μ is in a parametric class we estimate the parameters

Overlaping clusters - heavy thinning approach

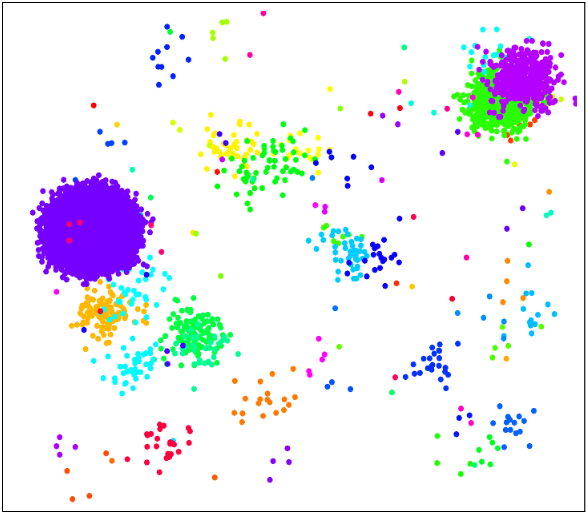


Figure : $\lambda = 0.4$, $\alpha = 0.6$, $\mu_x \sim \mathcal{N}(x, 0.5^2 \mathrm{I})$

Estimation of λ and α

When μ is known or have already been estimated, we suggest these

Estimation methods for λ and α

1. via void probabilities

Estimation of λ and α

When μ is known or have already been estimated, we suggest these

Estimation methods for λ and α

1. via void probabilities
2. via the p.g.f. of the counts distribution

Void probabilities for DαS point processes

The void probabilities (which characterise the distribution of a simple point process) are given by

$$\mathbf{P}\{\Phi(B) = 0\} = \exp\Big\{ - \lambda \int_A \mu(B)^\alpha \,\mathrm{d}x\Big\}.$$

Estimation of void probabilities

Unbiased estimator for the void probability function

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

$$\widehat{G}(r) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{r_i > r\}}$$

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

Then α and λ are estimated by the best fit to this curve.

Example: uniformly distributed clusters, 1D case

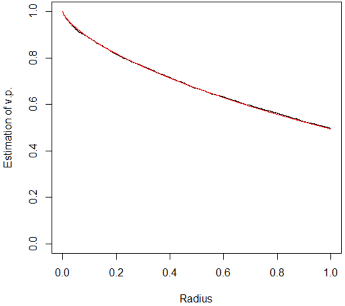


Figure : $\lambda = 0.3$, $\alpha = 0.7$, $\mu \sim U(B_1(0))$, $|A| = 3000$

Estimated values: $\widehat{\lambda} = 0.29$, $\widehat{\alpha} = 0.68$. Requires big data!

Void probabilities for thinned processes

p.g.fl. of DαS processes

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

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

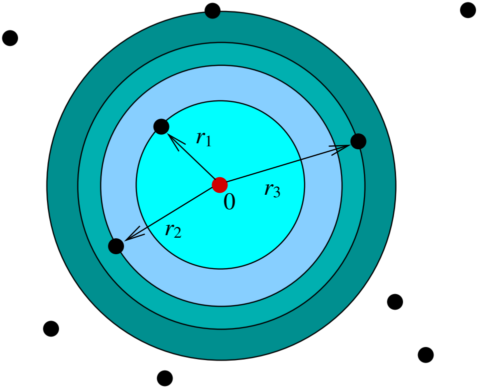
$$G_{p\circ\Phi}[h] = \exp\left\{ -p^\alpha \int_{\mathbb{S}} \langle 1-h,\mu\rangle^\alpha \sigma(d\mu)\right\}, \quad p\in[0,1], \quad 1-h \in \mathrm{BM}(X).$$

$$\sigma(\{\mu(\cdot - x),\ x\in B\}) = \lambda\cdot|B| \implies \alpha_{new} = \alpha,\ \lambda_{new} = \lambda\cdot p^\alpha.$$

Estimation via thinned process

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

$$\mathbf{P}\{(p \circ \Phi)(B_r(0)) = 0\}$$

$$= \sum_{k=1}^{\Phi} \mathbf{P}\{\text{“the closest survived point is the k-th”}\}\mathbf{P}\{r_k > r\}$$

$$= \sum_{k=1}^{\Phi} p(1-p)^{k-1}\mathbf{P}\{r_k > r\}$$

Estimation via thinned process

$$\mathbf{P}\{(p\circ\Phi)(B_r(0))=0\}$$

$$=\sum_{k=1}^{\Phi}\mathbf{P}\{\text{“the closest survived point is the k-th”}\}\mathbf{P}\{r_k>r\}$$

$$=\sum_{k=1}^{\Phi}p(1-p)^{k-1}\mathbf{P}\{r_k>r\}$$

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 $\operatorname{supp}\Phi$. Then

$$\widehat{G}(r)=\frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}p(1-p)^{k-1}\,\mathbb{1}_{\{r_{i,k}>r\}}$$

Example: uniform clusters, 1D case

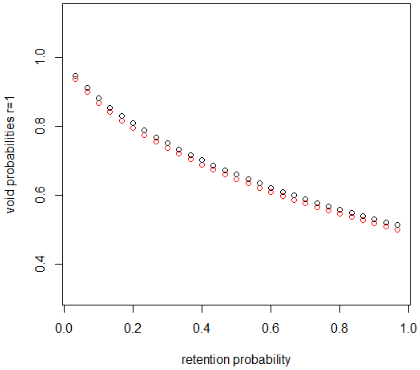


Figure : Estimation of v.p. of the thinned process for a process generated with $\lambda = 0.3$, $\alpha = 0.7$, $\mu \sim U(B_1(0))$, $|A| = 1000$

Estimated values: $\widehat{\lambda} = 0.29$, $\widehat{\alpha} = 0.72$

Counts distribution

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

$$\psi_{\Phi(B)}(s) := \mathbb{E}[s^{\Phi(B)}] = \exp\Big\{ -(1-s)^{\alpha} \int_{\mathbb{S}} \mu(B)^{\alpha}\sigma(d\mu)\Big\}. \qquad (2)$$

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

Estimation via counts distribution

The empirical p.g.f. is then

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

where B_i, $i = 1, \ldots, n$, are translates of a fixed referece set B and it is an unbiased estimator of $\psi_{\Phi(B)}$. It is then fitted to (2) for a range of s estimating λ and α.
We also tried the Hill plot from extremal distributions inference to estimate α, but the results were poor!

Conclusions

Simulation studies looked at the bias and variance in the extimation of $\alpha,\ \lambda$ in different situations:

- Big sample – moderate sample
- Overlapping clusters (large λ) – separate clusters (small λ)
- Heavy clusters (small α) – moderate clusters (α close to 1)

Best methods

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

Best methods

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

Best methods

- The simplest void probabilities method is prefered for large datasets or for moderate datasets with separated clusters. It best estimates α, but in the latter case λ is best estimated by counts p.g.f. fitting.
- λ 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 $\widehat{\alpha} = 0.17 - 0.28$ 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

1. Yu. Davydov, I. Molchanov and SZ ***Stability for random measures, point processes and discrete semigroups***, *Bernoulli*, **17**(3), 1015-1043, 2011
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?