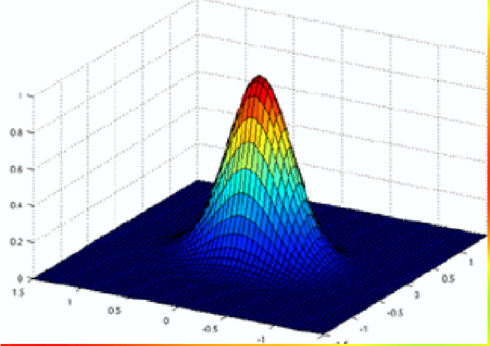

Scan Statistics for Independently Marked Point Processes

Joint work with Z. Kabluchko

Overview

- Introduction
- Motivation: CSA hypothesis testing for marked finite point processes
- Scan statistics for Lévy measures and point processes
- Limit theorem for a multiscale scan statistics for Lévy measures
- Ideas of the proof
- Outlook
- References

Introduction

Scan statistic

Let $\Phi = \{X_i\}$ be an independently marked point process in $\mathbb{R}^d$ with iid marks $\{M_i\}$ observed within a cube W. For a (cubic) subwindow $W_o \subset W$, define $S(W_o) = \sum_{i:X_i \in W_o} M_i$.

Scan statistic: $T = \sup_{W_o \in \mathcal{W}} S(W_o)$

- Usual scan statistic of fixed size $r > 0$: $\mathcal{W} = \{W_1 = x + r[0,1]^d, x \in \mathbb{R}^d : \; W_1 \subset W\}$.
- Multiscale scan statistic: $\mathcal{W} = \{\text{all cubes } W_1 \subset W\}$

Limit theorems: $T = T_n \xrightarrow{d} ?$ as $W = W_n = n[0,1]^d$, $n \to \infty$

Motivation

CSR hypothesis tests for independently marked (binomial) processes

Multiscale scan statistic T is a likelihood ratio test statistic for the following hypotheses:

- H_0: $\Phi = \{X_i\}$ is an independently marked binomial point process in W with iid marks $\{V_i\}$ having distribution F_0
- H_1: $\Phi = \{X_i\}$ is an independently marked binomial point process in W with marks $\{V_i\}$ having distribution F_1 if $X_i \in W_1$ and F_0 if $X_i \in W \setminus W_1$ where $W_1 \subset W$ is a certain subwindow of W.

with $M_i = \log p(V_i)$ where $p = dF_1/dF_0$ is the density of F_1 w.r.t. F_0.

Scan statistics for point processes

- Scan statistics in $\mathbb{R}^1$ and $\mathbb{R}^2$: Glaz, Balakrishnan (1999), Glaz, Naus, Wallenstein (2001)
- LT for the usual scan statistic in $\mathbb{R}^d$: $\Phi =$ stationary compound Poisson process (Chan (2009))
- LT for the multiscale scan statistic in $\mathbb{R}^1$: Cohen (1968), Iglehart (1972), Karlin, Dembo (1992), Doney, Maller (2005)
- LT for the multiscale scan statistic in $\mathbb{R}^d$: independently marked empirical processes and independently scattered Lévy measures (Kabluchko, S. (2009))

Scan statistic for Lévy noise

- Lévy noise: Let $\xi = \{\xi(t), t \geq 0\}$ be a Lévy process with $\xi(0) = 0$, $\mathbb{E}\,\xi(1) = \mu$, $\sigma^2 = Var\,\xi(1) > 0$.
  Lévy noise $\mathcal{Z} = \{\mathcal{Z}(B), B \in \mathcal{B}(\mathbb{R}^d)\}$ is an independently scattered stationary random measure on $\mathbb{R}^d$ driven by ξ, i.e. $\mathcal{Z}(B) \stackrel{\mathrm{d}}{=} \xi(|B|)$ for Borel sets $B \in \mathcal{B}(\mathbb{R}^d)$ where $|\cdot|$ is the volume in $\mathbb{R}^d$.
- Multiscale scan statistic:

$$T_n = \sup_{W_o \in \mathcal{W}_n} \mathcal{Z}(W_o), \quad n \in \mathbb{N}$$

for $\mathcal{W}_n = \{$all cubes within $W_n = [0, n]^d\}$.

LT for the scan statistic of Lévy noise

Theorem (Kabluchko, S. (2009))

- *If* $\mu > 0$ *then* $(T_n - \mu n^d)/(\sigma n^{d/2}) \xrightarrow{d} Y \sim N(0,1)$
- *If* $\mu = 0$ *then* $T_n/(\sigma n^{d/2}) \xrightarrow{d} \sup_{W_o \in \mathcal{W}_1} Z(W_o)$, *where* $Z = \{Z(B),\ B \in \mathcal{B}(\mathbb{R}^d)\}$ *is the standard Gaussian white noise on* $[0,1]^d$.
- *If the distribution of* $\xi(1)$ *is non–lattice,* $\varphi(s) = \log \mathbb{E}\, e^{s\xi(1)}$ *exists for* $s \in [0, s_0)$ *with the maximal* $s_0 \in (0, \infty]$ *and* $\exists\, s^* \in (0, s_0)$: $\varphi(s^*) = 0$ $(\mu < 0)$ *then*
$$s^* T_n - d \log n - (d-1) \log\log n - c \xrightarrow{d} Y,$$
where Y *is standard Gumbel distributed r. v. and* c *is a constant.*

LT for the scan statistic of Lévy noise

Ideas of the proof: $\mu > 0$

- Show that $T_n \sim \mathcal{Z}([0,n]^d)$ as $n \to \infty$ using
- the invariance principle for multidimensionally indexed random fields (Bickel and Wichura, 1971). It holds

  $$\Xi(n\cdot)/(\sigma n^{d/2}) \to Z(\cdot), \qquad n \to \infty$$

  weakly in Skorohod space $D[0,1]^d$, where $\Xi = \{\Xi(x),\ x \in \mathbb{R}^d\}$ is the Lévy sheet defined by $\Xi(x) = \mathcal{Z}([o,x])$, $x \in [0,1]^d$ and $Z = \{Z(x),\ x \in \mathbb{R}^d\}$ is the Brownian sheet on $[0,1]^d$ with continuous paths.
- Use a classical CLT for iid random variables ($\mathcal{Z}$ is a noise!).

LT for the scan statistic of Lévy noise

Ideas of the proof: $\mu = 0$

- Use the above invariance principle once again. Apply the continuous sup-functional to get $T_n/(\sigma n^{d/2}) \xrightarrow{d} \sup_{W_o \in \mathcal{W}_1} Z(W_o)$.

LT for the scan statistic of Lévy noise

Ideas of the proof: $\mu < 0$

- Use Pickands' method of double sums
- Only those cubes of volume $v_n = c^* \log n$ and $v_n \pm A\sqrt{v_n}$ contribute substantially to the scan statistic T_n where c^* and A are some positive constants and A is large enough.
- Use the large deviation result for Lévy processes by V. Petrov (1965).

LT for the scan statistic of the compound Poisson process

Corollary (Kabluchko, S. (2009))

The same result holds for the case if $\Phi = \{(X_i, M_i),\ i \in \mathbb{N}\}$ *is an independently marked stationary Poisson point process in* $\mathbb{R}^d$ *with unit intensity and iid marks* $\{M_i\}$, $\mathbb{E}\, M_1 = \mu$, $\sigma^2 = Var\, M_1 \in (0, \infty)$.

Proof.

Use the Lévy noise $\mathcal{Z}(B) = \sum_{i: X_i \in B} M_i$ for bounded Borel sets in $\mathbb{R}^d$. $\square$

LT for the scan statistic of marked empirical processes

Let $\Phi = \{(X_i, M_i), i = 1, \dots, n\}$ be an independently marked empirical process in $W = [0,1]^d$ with iid marks $\{M_i\}$. Let $\mathbb{E}\, M_1 = \mu$, $\sigma^2 = Var\, M_1 \in (0, \infty)$. The multiscale scan statistic is here

$$T_n = \sup_{W_o \in \mathcal{W}} \sum_{i: X_i \in W_o} M_i$$

with $\mathcal{W} = \{\text{all cubes } W_1 \subset W\}$.

LT for the scan statistic of marked empirical processes

Corollary (Kabluchko, S. (2009))

- *If* $\mu > 0$ *then* $(T_n - \mu n^d)/(\sigma n^{d/2}) \xrightarrow{d} Y \sim N(0,1)$
- *If* $\mu = 0$ *then* $T_n/(\sigma n^{d/2}) \xrightarrow{d} \sup_{W_o \in \mathcal{W}_1} Z(W_o)$, *where* $Z = \{Z(B),\ B \in \mathcal{B}(\mathbb{R}^d)\}$ *is the standard Gaussian white noise on* $[0,1]^d$.
- *If the distribution of* M_1 *is non–lattice,* $\varphi(s) = \log \mathbb{E}\, e^{sM_1}$ *exists for* $s \in [0, s_0)$ *with the maximal* $s_0 \in (0, \infty]$ *and* $\exists\, s^* \in (0, s_0)$*:* $\varphi(s^*) = 0$ $(\mu < 0)$ *then*
$$s^* T_n - \log n - (d-1) \log\log n - c \xrightarrow{d} Y,$$
where Y *is standard Gumbel distributed r. v. and* c *is a constant.*

LT for the scan statistic of marked empirical processes

Proof.

Approximate the distribution of the marked empirical process in $[0,1]^d$ by the distribution of the stationary compound Poisson process Φ with unit intensity in the cube $[0,t_n]^d$, $t_n = \inf\{t > 0 : \Phi([0,t]^d) = n+1\}$. Then use the first corollary. $\square$

Outlook

- Erdös-Rényi-type laws for the scan statistics with fixed cube size $c \log n$ for Lévy noise ($c \log n / n$ for marked empirical processes, resp.)
- LT for other types of scan statistics e.g. those based on other empirical quantiles of scans or on their Lorenz curve.

References

- H. P. Chan, Maxima of moving sums in a Poisson random field, Adv. Appl. Probab. 41 (2009), No. 3, 647-663.
- J. Cohen, Extreme value distribution for the M/G/1 and the G/M/1 queueing systems. Ann. Inst. Henri Poincaré, Nouv. Ser., Sect. B (1968) 4, 83–98.
- R. Doney, R. Maller, Cramér's estimate for a reflected Lévy process. Ann. Appl. Probab. 15 (2005), 1445–1450.
- J. Glaz, S. Wallenstein, J. Naus, Scan Statistics, Springer, 2001.
- J. Glaz, N. Balakrishnan, Scan Statistics and Applications, Birkhäuser, 1999.
- D. L. Iglehart, Extreme values in the GI/G/1 queue. Ann. Math. Statist. 43 (1972), 627–635.
- Z. Kabluchko, E. Spodarev, Scan statistic of Lévy noises and marked empirical processes, Adv. Appl. Probab. 41 (2009), No. 1, 13-37.
- S. Karlin, A. Dembo, Limit distributions of maximal segmental score among Markov-dependent partial sums. Adv. Appl. Probab. 24 (1992), 113–140.