Scan statistic of Lévy noises and marked empirical processes

Zakhar Kabluchko, Evgeny Spodarev

July 2, 2008

Institut für Mathematische Stochastik, Universität Göttingen,
Maschmühlenweg 8-10, D-37073 Göttingen, Germany
and
Institut für Stochastik, Universität Ulm, Helmholtzstr. 18, D-89069 Ulm,
Germany

Abstract

Let n points be chosen independently and uniformly in the unit cube $[0,1]^d$, and suppose that each point is supplied with a mark, the marks being i.i.d. random variables independent from the location of the points. To each cube R contained in $[0,1]^d$ we associate its score $\mathcal{X}_n(R)$ defined as the sum of marks of all points contained in R. The scan statistic is defined as the maximum of $\mathcal{X}_n(R)$, taken over all cubes R contained in $[0,1]^d$. We show that if the marks are non-lattice random variables with finite exponential moments, having negative mean and assuming positive values with non-zero probability, then after appropriate normalization the distribution of the scan statistic converges as $n \to \infty$ to the Gumbel distribution. We prove also a corresponding result for the scan statistic of a Lévy noise with negative mean. The more elementary cases of zero and positive mean are also considered.

Keywords: scan statistic, independently scattered Lévy measure, marked empirical process, extremes, Pickands' double sum method.
AMS 2000 Subject Classification: 60G70

1 Introduction

Let $\{U_i, i = 1, \dots, n\}$ be n points chosen independently and uniformly from the d-dimensional unit cube $[0,1]^d$. Suppose that to each point U_i a mark

X_i is attached, the marks being i.i.d. real-valued random variables independent from the location of points U_i. The collection $\{(U_i, X_i), i = 1, \dots, n\}$ is called *marked empirical process*. A natural problem is how to detect inhomogeneities, e.g. clustering of unusually big marks, in the marked empirical process. To this end, one may consider the *scan statistic*, whose definition we now recall (see Glaz et al. (2001), Glaz and Balakrishnan (1999)). For a set $R \subset [0,1]^d$ define its score $\mathcal{X}_n(R)$ as the sum of marks of all points contained in R, that is

$$\mathcal{X}_n(R) = \sum_{i \in \{1,\dots,n\}: U_i \in R} X_i. \tag{1}$$

Then the scan statistic is defined as $\sup_{R \in \mathcal{R}(1)} \mathcal{X}_n(R)$, where $\mathcal{R}(1)$ is some collection of subsets ("windows") of $[0,1]^d$. Since no a priori assumptions about the size of clusters are made, it is natural to require $\mathcal{R}(1)$ to contain windows of all sizes. For example, one can take $\mathcal{R}(1)$ to be the collection of all cubes contained in $[0,1]^d$ (a cube is a translate of the set $[0,x]^d$ for some $x > 0$).

The main question is then how the scan statistic is distributed as $n \to \infty$. To state our main result we make the following assumptions about the distribution of random marks.

(X1) The logarithmic moment generating function $\psi(\theta) = \log \mathbb{E} e^{\theta X_1}$ of X_1 exists as long as $\theta \in [0, \theta_0)$ for some $\theta_0 \in (0, \infty]$ which is supposed to be maximal with this property.

(X2) The function ψ has a zero $\theta^* \in (0, \theta_0)$.

(X3) The distribution of X_1 is non-lattice.

Note that the second condition implies that $\mathbb{E}X_1 = \psi'(0) < 0$. A further corollary is that $\mathbb{P}[X_1 > 0] \neq 0$. Conversely, if 1) $\mathbb{E}X_1 < 0$, 2) $\mathbb{P}[X_1 > 0] \neq 0$, and 3) condition (X1) is satisfied for $\theta_0 = \infty$, then condition (X2) is fulfilled automatically.

Recall that $\mathcal{R}(1)$ is the collection of all cubes contained in $[0,1]^d$. For a constant H^* to be specified later, set

$$u_n(\tau) = \frac{1}{\theta^*}(d \log n + (d-1) \log \log n + H^* + \tau), \quad \tau \in \mathbb{R}. \tag{2}$$

Our main result reads as follows.

Theorem 1.1. *Let $\{(U_i, X_i), i = 1, \dots, n\}$ be a marked empirical process satisfying the above conditions X1-X3. Then*

$$\lim_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(1)} \mathcal{X}_n(R) \le u_n(\tau)] = \exp\{-e^{-\tau}\}.$$

The scan statistic of Theorem 1.1 may be interpreted as a multiscale test statistic in the following sense. Suppose we are given a set of points $\{U_i, i = 1, \dots, n\}$ in $[0,1]^d$, the point U_i being marked by a number Y_i. Let F_0 and F_1 be two distribution functions, such that the density $p = dF_1/dF_0$ exists. Consider the following hypotheses (here, $R \in \mathcal{R}(1)$):

$$
\begin{aligned}
H_0 &: U_i \text{ are i.i.d. uniform in } [0,1]^d \text{ and } Y_i \text{ are i.i.d. with } Y_i \sim F_0;\\
H_R &: U_i \text{ are i.i.d. uniform in } [0,1]^d,\\
&\qquad \text{whereas } Y_i \text{ are independent with } Y_i \sim \begin{cases} F_1 \text{ if } U_i \in R,\\ F_0 \text{ if } U_i \notin R;\end{cases}\\
H_1 &: \cup_{R \in \mathcal{R}(1)} H_R.
\end{aligned}
$$

Of course, we always suppose that the families $\{U_i, i = 1, \dots, n\}$ and $\{Y_i, i = 1, \dots, n\}$ are mutually independent. It is easy to see that the log-likelihood statistic for testing H_0 against H_R is given by $\mathcal{X}_n(R)$ defined in formula (1) with $X_i = \log p(Y_i)$. Thus, the scan statistic considered in Theorem 1.1 may be interpreted as a generalized likelihood ratio statistic for testing H_0 against H_1.

In the one-dimensional case, the distribution of scan statistic applied to an i.i.d. sequence with negative mean was studied starting with Iglehart (1972), who showed an analogue of Theorem 1.1 in dimension 1 with marked empirical process replaced by an i.i.d. sequence of random variables. This result was extended from i.i.d. to Markov-dependent sequences in Karlin and Dembo (1992), a version for Lévy processes was obtained in Doney and Maller (2005). The scan statistic considered in Iglehart (1972) appears in a variety of settings. For example, it may be interpreted as the statistic used in the CUSUM stopping procedure in change-point analysis, as a maximal waiting time among the first n customers in a GI/G/1 queue or, in bioinformatics, as a maximal segmental score when comparing two random sequences.

The papers cited above use fluctuation theory of random walks and Lévy processes. Fluctuation theory, giving very elegant solutions in dimension $d = 1$, does not allow an extension to the case $d \geq 2$. To prove Theorem 1.1 we use the method of double sums introduced by Pickands in Pickands (1969b), Pickands (1969a), see also Chapter 12 in Leadbetter et al. (1983). For the development of the method, see Piterbarg (1996). Although Pickands' method was developed originally to study extremes of gaussian processes, it can be applied in the non-gaussian case as well, see for example Piterbarg and Kozlov (2003).

A question closely related to that considered in Theorem 1.1 is about the distribution of $\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R)$, where $\mathcal{Z}$ is an independently scattered homogeneous random Lévy measure on $\mathbb{R}^d$ with negative mean, and $\mathcal{R}(n)$ is the collection of all cubes contained in $[0,n]^d$. The analogue of Theorem 1.1 in this situation, Theorem 2.1, will be stated in Section 2. In fact, it will be more convenient for us to prove Theorem 2.1 first and then to deduce Theorem 1.1 from it using a close relation between empirical process and Poisson processes.

It will be seen in the proof of Theorem 2.1 that the main contribution to the extremes of the scan statistic is made by cubes of some "optimal" volume $v_n \approx c^* \log n$ for some constant c^*, as well as by cubes having a volume differing from the optimal one by a quantity of order $\sqrt{v_n}$. Thus, the situation we encounter is close to that of Hüsler and Piterbarg (2004) who considered scan statistic applied to a fractional Brownian noise with negative mean. Using a change of variables, Hüsler and Piterbarg (2004) reduced their problem to studying extremes of a gaussian field with non-constant variance, the points of maximal variance corresponding to the intervals of "optimal" size. In our case, random fields under consideration are non-gaussian, which makes non-applicable many results from the extreme-value theory of gaussian processes and causes some technical difficulties.

In Theorem 1.1 and Theorem 2.1 below, the distribution of scan statistic applied to noises with *negative* mean is considered. One may ask, what happens if the mean is zero or positive. Compared to the negative mean case, these two cases, which will be treated in Section 3, are much simpler. See Iglehart (1972) for the similar problem in the case of i.i.d. sequences and Zeevi and Glynn (2000) for the case of fractional Brownian noise.

Finally, let us note that although we are considering only scan statistic with *variable* window size, the same method, with considerable simplifications, can be used to obtain Erdös-Rényi-type laws in distribution for the scan statistic taken over all windows of *fixed* volume $c \log n$ (resp. $c \log n/n^d$ in the case of marked empirical process). The corresponding result in the case of one-dimensional i.i.d. sequence was proved in J. Komlós and G. Tusnády (1975), Piterbarg and Kozlov (2003). In the case of d-dimensional compound Poisson process, this can be deduced from Chan (2007), where large deviations estimates are proved for the scan statistic taken over a set of windows with fixed shape and size (the windows need not be cubes). Using such a statistic in applications requires a preknowledge about the size of clusters to be discovered.

The organization of the paper is as follows. In Section 2 we state Theorem 2.1, an analogue of Theorem 1.1 for scan statistic applied to a Lévy noise. Limiting distribution of scan statistic in the case of zero or positive mean is considered in Section 3. The proof of Theorem 2.1 will be carried out in Section 4. Finally, in Section 5 we deduce Theorem 1.1 from Theorem 2.1.

2 Scan statistic of a Lévy noise with negative mean

Let $\{\xi(t), t \geq 0\}$ be a Lévy process. An independently scattered homogeneous Lévy measure on $\mathbb{R}^d$ (*Lévy noise* for short) is a stochastic process $\{\mathcal{Z}(R), R \in \mathcal{B}(\mathbb{R}^d)\}$, indexed by the collection $\mathcal{B}(\mathbb{R}^d)$ of Borel sets in $\mathbb{R}^d$, such that the following conditions are satisfied.

(Z1) $\mathcal{Z}(R)$ has the same distribution as $\xi(|R|)$, where $|R|$ is the Lebesgue measure of a Borel set R.

(Z2) $\mathcal{Z}(R_1), \ldots, \mathcal{Z}(R_n)$ are independent whenever $R_1, \ldots, R_n$ are disjoint Borel subsets of $\mathbb{R}^d$.

The Lévy noise may be reconstructed from the corresponding Lévy sheet $\{\Xi(x_1, \ldots, x_d), (x_1, \ldots, x_d) \in [0, \infty)^d\}$ defined by

$$\Xi(x_1, \ldots, x_d) = \mathcal{Z}([0, x_1] \times \ldots \times [0, x_d]).$$

We always suppose that the sample paths of Ξ belong to the Skorokhod space in d dimensions as defined e.g. in Bickel and Wichura (1971).
Concerning the underlying Lévy process ξ, we suppose that the following three conditions are satisfied.

(L1) The logarithmic moment generating function $\varphi(\theta) = \log \mathbb{E} e^{\theta \xi(1)}$ of $\xi(1)$ exists as long as $\theta \in [0, \theta_0)$ for some $\theta_0 \in (0, \infty]$ which is supposed to be maximal with this property.

(L2) The function φ has a zero $\theta^* \in (0, \theta_0)$.

(L3) The distribution of $\xi(1)$ is non-lattice.

Let $\mathcal{R}(n)$ be the collection of all cubes contained in $[0, n]^d$.

Theorem 2.1. *Let $\{\mathcal{Z}(R), R \in \mathcal{B}(\mathbb{R}^d)\}$ be a Lévy noise on $\mathbb{R}^d$ as defined above such that conditions L1-L3 are satisfied. Define $u_n(\tau)$ as in* (2). *Then*

$$\lim_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n(\tau)] = \exp\{-e^{-\tau}\}.$$

For $d = 1$ this theorem was proved in Doney and Maller (2005) by a method which uses fluctuation theory of Lévy processes and thus cannot be extended to higher dimensions.

3 Results in the case of zero and positive mean

In the preceding sections we considered the limiting distribution of scan statistic applied to a Lévy noise (resp. marked empirical process) assuming, essentially, that the mean is negative. The case of negative mean is the most difficult one. Here, we state the corresponding results in the case of zero and positive mean, restricting ourselves for simplicity to Lévy noises. First we treat the case of zero mean. Let $\{\mathcal{W}(R), R \in \mathcal{B}([0,1]^d)\}$ be a standard gaussian white noise on $[0,1]^d$; we suppose that the corresponding Brownian sheet has continuous sample paths. Recall that $\mathcal{R}(n)$ is the collection of all cubes contained in $[0,n]^d$. For a Borel set R let $|R|$ be its Lebesgue measure.

Theorem 3.1. *Let $\{\mathcal{Z}(R), R \in \mathcal{B}(\mathbb{R}^d)\}$ be a Lévy noise such that*

$$\mathbb{E}\mathcal{Z}(R) = 0, \quad \mathrm{Var}\mathcal{Z}(R) = \sigma^2|R|$$

for each Borel set $R \subset \mathbb{R}^d$. Then the distribution of $\sigma^{-1}n^{-d/2}\sup_{R\in\mathcal{R}(n)}\mathcal{Z}(R)$ converges as $n \to \infty$ to the distribution of $\sup_{R\in\mathcal{R}(1)}\mathcal{W}(R)$.

Proof. The Lévy sheet $\Xi_{\mathcal{Z}}$ corresponding to the noise $\mathcal{Z}$ (resp. the Brownian sheet $\Xi_{\mathcal{W}}$ corresponding to the noise $\mathcal{W}$) are defined by

$$\begin{aligned}\Xi_{\mathcal{Z}}(x_1,\ldots,x_d) &= \mathcal{Z}([0,x_1]\times\ldots\times[0,x_d]),\\ \Xi_W(x_1,\ldots,x_d) &= \mathcal{W}([0,x_1]\times\ldots\times[0,x_d]).\end{aligned}$$

By the invariance principle for multidimensionally indexed random fields, see e.g. Bickel and Wichura (1971), we have

$$\sigma^{-1}n^{-d/2}\Xi_{\mathcal{Z}}(n\cdot) \Rightarrow \Xi_{\mathcal{W}}(\cdot) \quad \text{as } n \to \infty, \tag{3}$$

where "$\Rightarrow$" denotes the weak convergence in the Skorohod space $D([0,1]^d)$. We define a continuous functional $F : D([0,1]^d) \to \mathbb{R}$ by

$$F(\Xi) = \sup_{R\in\mathcal{R}(1)} \Xi(R), \quad \Xi \in D([0,1]^d),$$

where $\Xi(R)$ is defined in a straightforward way (so that e.g. $\Xi_{\mathcal{Z}}(R) = \mathcal{Z}(R)$ and $\Xi_{\mathcal{W}}(R) = \mathcal{W}(R)$). It follows from (3) that the random variable $F(\sigma^{-1}n^{-d/2}\Xi_{\mathcal{Z}}(n\cdot)) = \sigma^{-1}n^{-d/2}\sup_{R\in\mathcal{R}(n)}\mathcal{Z}(R)$ converges in distribution to $F(\Xi_{\mathcal{W}}(\cdot)) = \sup_{R\in\mathcal{R}(1)}\mathcal{W}(R)$ as $n \to \infty$. This proves the theorem. □

Theorem 3.2. *Let $\{\mathcal{Z}(R), R \in \mathcal{B}(\mathbb{R}^d)\}$ be a Lévy noise such that for some $\mu > 0$ and $\sigma^2 > 0$*

$$\mathbb{E}\mathcal{Z}(R) = \mu|R|, \quad \mathrm{Var}\mathcal{Z}(R) = \sigma^2|R|.$$

Then the distribution of $\sigma^{-1}n^{-d/2}\left(\sup_{R\in\mathcal{R}(n)}\mathcal{Z}(R) - n^d\mu\right)$ converges as $n \to \infty$ to the standard normal distribution.

Proof. The idea is to show that $\sup_{R\in\mathcal{R}(n)} \mathcal{Z}(R)$ behaves essentially like $\mathcal{Z}([0,n]^d)$ and then to apply the central limit theorem. We show that for every $a>0$

$$\mathbb{P}[\sup_{R\in\mathcal{R}(n)} \mathcal{Z}(R) - \mathcal{Z}([0,n]^d) \geq an^{d/2}] \to 0, \quad n\to\infty. \tag{4}$$

Denoting the left-hand side by P_n and taking $\varepsilon>0$ small we have $P_n \leq P_n' + P_n''$, where

$$\begin{aligned} P_n' &= \mathbb{P}[\sup_{R\in\mathcal{R}'(n)} \mathcal{Z}(R) - \mathcal{Z}([0,n]^d) \geq 0], \\ P_n'' &= \mathbb{P}[\sup_{R\in\mathcal{R}''(n)} \mathcal{Z}(R) - \mathcal{Z}([0,n]^d) \geq an^{d/2}], \end{aligned}$$

and $\mathcal{R}'(n) = \{R\in\mathcal{R}(n) : |R| < (1-\varepsilon)n^d\}$, $\mathcal{R}''(n) = \{R\in\mathcal{R}(n) : |R| \geq (1-\varepsilon)n^d\}$.
Define the centered noise $\{\mathcal{Z}_0(R), R\in\mathcal{B}(\mathbb{R}^d)\}$ by $\mathcal{Z}_0(R) = \mathcal{Z}(R) - \mu|R|$. Then

$$P_n' \leq \mathbb{P}[\sup_{R\in\mathcal{R}'(n)} \mathcal{Z}_0(R) - \mathcal{Z}_0([0,n]^d) \geq \mu\varepsilon n^d].$$

By the multidimensional invariance principle of Bickel and Wichura (1971), applied to $\mathcal{Z}_0$, $\lim_{n\to\infty} P_n' = 0$. Further,

$$P_n'' \leq \mathbb{P}[\sup_{R\in\mathcal{R}''(n)} \mathcal{Z}_0(R) - \mathcal{Z}_0([0,n]^d) \geq an^{d/2}].$$

Again using the multidimensional invariance principle, we see that this converges to

$$c(\varepsilon) = \mathbb{P}[\sup_{R\in\mathcal{R}(1), |R|\geq 1-\varepsilon} \mathcal{W}(R) - \mathcal{W}([0,1]^d) \geq a].$$

It is easy to see that $\lim_{\varepsilon\to 0} c(\varepsilon) = 0$. It follows that

$$\limsup_{n\to\infty} P_n \leq \limsup_{n\to\infty} P_n' + \limsup_{n\to\infty} P_n'' \leq c(\varepsilon).$$

Letting $\varepsilon\to 0$ we obtain $\lim_{n\to\infty} P_n = 0$, which proves (4). Now, the statement of the theorem follows from the central limit theorem applied to $\mathcal{Z}([0,n]^d)$ in combination with (4). □

4 Proof of Theorem 2.1

This section is devoted to the proof of Theorem 2.1. In the first two subsections we introduce some notation and prove technical lemmas which will be often used in the sequel. In what follows, $C>0$ (resp. $\delta>0$) denotes a large (resp. small) constant whose value may change from line to line.

4.1 Notation and preliminaries

Normalizing constants. Let $\tau \in \mathbb{R}$ be fixed once for all. For the constants H^* and α^* to be specified later define

$$u_n = \frac{1}{\theta^*}(d \log n + (d-1)\log\log n + H^* + \tau); \quad (5)$$

$$v_n = u_n/\alpha^*, \quad l_n = v_n^{1/d}. \quad (6)$$

The space of cubes. A d-dimensional cube (denoted usually by R) is a set of the form $\times_{i=1}^{d}[x_i - x/2, x_i + x/2]$, where $(x_1, \ldots, x_d) \in \mathbb{R}^d$ are the coordinates of the center and $x > 0$ is the side length. The space of all cubes, denoted by $\mathcal{R}$, will be identified with $\mathbb{R}^d \times (0, \infty)$, a cube R being identified with the tuple $(x_1, \ldots, x_d; x)$. We denote by $|R| = x^d$ the volume of the cube R.
The underlying Lévy process. Let $\{\xi(t), t \geq 0\}$ be a Lévy process satisfying conditions L1-L3 of Section 2. The function φ is real analytic, convex function on $[0, \theta_0)$. The zero θ^* is necessary unique by convexity of φ. It follows from condition L2 that $\mathbb{E}[\xi(1)] < 0$. Further, condition L2 implies that $\mathbb{P}[\xi(1) > 0] \neq 0$ and it follows that $\xi(1)$, being infinitely divisible, can attain arbitrarily large values. Using this, it is not difficult to show that $\lim_{\theta \to \theta_0} \varphi'(\theta) = \infty$. Note also that φ' is monotone increasing and $\varphi'(0) = \mathbb{E}[\xi(1)] < 0$. For each $\alpha \in (\mathbb{E}\xi(1), \infty)$ let $\theta(\alpha)$ be the unique solution of $\varphi'(\theta(\alpha)) = \alpha$ and let $\sigma(\alpha) = \sqrt{\varphi''(\theta(\alpha))}$. Define the Cramér-Chernoff information function $I : [\mathbb{E}\xi(1), \infty) \to [0, \infty)$ by

$$I(\alpha) = \sup_{\theta \geq 0}(\alpha\theta - \varphi(\theta)) = \alpha\theta(\alpha) - \varphi(\theta(\alpha)). \quad (7)$$

Define

$$\alpha^* = \varphi'(\theta^*), \quad \sigma^* = \sqrt{\varphi''(\theta^*)}. \quad (8)$$

Note that $\alpha^* > 0$, $\theta(\alpha^*) = \theta^*$ and $\sigma(\alpha^*) = \sigma^*$.

Lemma 4.1. *The function $J : (0, \infty) \to (0, \infty)$ defined by $J(\alpha) = I(\alpha)/\alpha$ has a unique minimum at $\alpha = \alpha^*$. Furthermore*

$$J(\alpha^*) = \theta^*; \quad J'(\alpha^*) = 0; \quad J''(\alpha^*) = \frac{1}{\alpha^* \sigma^{*2}}. \quad (9)$$

Proof. Substituting $\alpha = \alpha^*$ into (7) gives $I(\alpha^*) = \alpha^*\theta^* - \varphi(\theta^*) = \alpha^*\theta^*$, which proves that $J(\alpha^*) = \theta^*$. Differentiating (7) at $\alpha = \alpha^*$ we obtain $I'(\alpha^*) = \theta(\alpha^*) = \theta^*$ and $I''(\alpha^*) = \theta'(\alpha^*) = 1/\varphi''(\theta(\alpha^*)) = 1/\sigma^{*2}$. Substituting this into

$$J'(\alpha) = \alpha^{-2}(\alpha I'(\alpha) - I(\alpha)), \quad J''(\alpha) = \alpha^{-3}(\alpha^2 I''(\alpha) - 2\alpha I'(\alpha) + 2I(\alpha))$$

we obtain (9). In order to show that $\alpha = \alpha^*$ is the *unique* minimum of J note that it follows from the above equation that $J'(\alpha) = \alpha^{-2}\varphi(\theta(\alpha))$ and that $\alpha = \alpha^*$ is the unique solution of $\varphi(\theta(\alpha)) = 0$, $\alpha > 0$. □

Large deviations. We need the following precise large deviations theorem due to Petrov (1965).

Theorem 4.1. *Let $\{\xi(t), t \geq 0\}$ be a Lévy process satisfying conditions L1-L3 of Section 2. Let $\alpha \in (\mathbb{E}\xi(1), \infty)$. We have as $v \to \infty$*

$$\mathbb{P}[\xi(v)/v > \alpha] \sim \frac{1}{\sqrt{2\pi}\theta(\alpha)\sigma(\alpha)} \frac{1}{\sqrt{v}} \exp\{-vI(\alpha)\}.$$

Moreover, the above holds uniformly in α as long as α stays bounded from $\mathbb{E}\xi(1)$ and $+\infty$.

The next lemma is a simple consequence of Markov's inequality and will be often used in the sequel.

Lemma 4.2. *For every $u, v > 0$*

$$\mathbb{P}[\xi(v) > u] \leq \exp(-uJ(u/v)).$$

Proof. By Markov's inequality we have, for each $t > 0$,

$$\mathbb{P}[\xi(v) > u] \leq e^{-ut}\mathbb{E}e^{t\xi(v)} = \exp(-ut + v\varphi(t)) = \exp(-v(ut/v - \varphi(t))).$$

Since the above is true for every $t > 0$, we obtain

$$\mathbb{P}[\xi(v) > u] \leq \exp(-vI(u/v)) = \exp(-uJ(u/v)),$$

which finishes the proof. □

Corollary 4.1. *For every $u, v > 0$*

$$\mathbb{P}[\xi(v) > u] \leq \exp(-\theta^* u).$$

Proof. Use the Lemma 4.2 and recall that $J(u/v) \geq \theta^*$ by Lemma 4.1. □

4.2 Modulus of continuity estimate

Let $\{\mathcal{Z}(R), R \in \mathcal{B}(\mathbb{R}^d)\}$ be a Lévy noise such that the underlying Lévy process ξ satisfies assumptions L1-L3 of Section 2 and let Ξ be the corresponding Lévy sheet. The next lemma gives a large deviations estimate for the supremum of Ξ over $[0, c]^d$, $c > 0$.

Lemma 4.3. *For every $\theta < \theta_0$, $c > 0$ there is $C = C(\theta, c)$ such that*

$$\mathbb{P}[\sup_{(x_1,\dots,x_d)\in[0,c]^d} \Xi(x_1, \dots, x_d) > u] < Ce^{-\theta u} \quad \forall u > 0. \tag{10}$$

Proof. For simplicity we assume that $c = 1$. For $d = 1$ the lemma was proved in Willekens (1987). We use induction over d combined with the method of Willekens (1987). Suppose that the statement of the lemma was proved in dimensions $1, \ldots, d-1$. Let

$$\tau_1 = \inf\{x_1 \geq 0 : \exists x_2, \ldots, x_d \in [0,1]^{d-1} : \Xi(x_1, \ldots, x_d) > u\}.$$

The left-hand side of (10) is the probability of the event $A = \{\tau_1 \leq 1\}$. We have $A = A_1 \cup A_2$ where

$$\begin{aligned} A_1 &= \{\tau_1 \leq 1 \cap \sup_{(x_2,\ldots,x_d)\in[0,1]^{d-1}} \Xi(1, x_2, \ldots, x_d) > u - 1\}, \\ A_2 &= \{\tau_1 \leq 1 \cap \sup_{(x_2,\ldots,x_d)\in[0,1]^{d-1}} \Xi(1, x_2, \ldots, x_d) < u - 1\}. \end{aligned}$$

Now, by the induction hypothesis,

$$\mathbb{P}[A_1] \leq \mathbb{P}[\sup_{(x_2,\ldots,x_d)\in[0,1]^{d-1}} \Xi(1, x_2, \ldots, x_d) > u - 1] \leq Ce^{-\theta u}.$$

We estimate

$$\begin{aligned} \mathbb{P}[A_2] &\leq \mathbb{P}[A \cap \inf_{(x_2,\ldots,x_d)\in[0,1]^{d-1}} (\Xi(1, x_2, \ldots, x_d) - \Xi(\tau_1, x_2, \ldots, x_d)) < -1] \\ &\leq \mathbb{P}[A \cap \inf_{(x_1,x_2,\ldots,x_d)\in[0,1]^{d}} (\Xi(\tau_1 + x_1, x_2, \ldots, x_d) - \Xi(\tau_1, x_2, \ldots, x_d)) < -1] \\ &= \mathbb{P}[A]\mathbb{P}[\inf_{(x_1,x_2,\ldots,x_d)\in[0,1]^{d}} (\Xi(\tau_1 + x_1, x_2, \ldots, x_d) - \Xi(\tau_1, x_2, \ldots, x_d)) < -1] \\ &= \mathbb{P}[A]\mathbb{P}[\inf_{(x_1,x_2,\ldots,x_d)\in[0,1]^{d}} \Xi(x_1, x_2, \ldots, x_d) < -1] \\ &= p\mathbb{P}[A] \end{aligned}$$

for some $p < 1$. We obtain $\mathbb{P}[A] = \mathbb{P}[A_1] + \mathbb{P}[A_2] \leq Ce^{-\theta u} + p\mathbb{P}[A]$ for some $p < 1$, from which the statement of the lemma follows. □

In the sequel, we shall often use the following technical lemma, which estimates the modulus of continuity of the random field $\{\mathcal{Z}(R), R \in \mathcal{R}\}$.

Lemma 4.4. *Let $c > 0$ be a fixed constant. Let $x > c^d$ and let $q < cx^{1-d}$. Define a set of cubes $\mathcal{B} = [-q/2, q/2]^d \times [x, x+q]$. Let $R_0 = [-(x-q)/2, (x-q)/2]^d$ be the intersection of all cubes from $\mathcal{B}$ and define the random variable M by*

$$M = \sup_{R\in\mathcal{B}} \mathcal{Z}(R) - \mathcal{Z}(R_0).$$

Then, for every $\theta < \theta_0$, there is a constant $C = C(c, \theta)$ such that uniformly in x and q

$$\mathbb{P}[M > t] \leq Ce^{-\theta t} \quad \forall t > 0.$$

Proof. We show that $\mathbb{E}e^{\theta M} < C(c, \theta)$, the lemma follows then from Markov's inequality. For $h = (h_1, \dots, h_d)$, $h_i > 0$, and $\varepsilon_1, \dots, \varepsilon_d \in \{-1, 0, 1\}$ define a rectangle

$$R(\varepsilon_1, \dots, \varepsilon_d; h) = I(\varepsilon_1; h_1) \times \dots \times I(\varepsilon_d; h_d),$$

where

$$I(\varepsilon_i; h_i) = \begin{cases} [-(x-q)/2, (x-q)/2], & \text{if } \varepsilon_i = 0, \\ [(x-q)/2, (x-q)/2 + h_i], & \text{if } \varepsilon_i = 1, \\ [-(x-q)/2 - h, (x-q)/2], & \text{if } \varepsilon_i = -1. \end{cases}$$

Note that $R(0, \dots, 0; h) = R_0$ (in particular, the left-hand side does not depend on h). Let

$$M(\varepsilon_1, \dots, \varepsilon_d) = \sup_{h \in [0, 3q/2]^d} \mathcal{Z}(R(\varepsilon_1, \dots, \varepsilon_d; h)).$$

The random variables $M(\varepsilon_1, \dots, \varepsilon_d)$ are independent and

$$M \leq \sum_{(\varepsilon_1, \dots, \varepsilon_d) \in \{-1,0,1\}^d \setminus (0, \dots, 0)} M(\varepsilon_1, \dots, \varepsilon_d). \tag{11}$$

Furthermore, if r is the number of $+1$ and -1 among ε_i and if $r \neq 0$, then the random variable $M(\varepsilon_1, \dots, \varepsilon_d)$ has the same distribution as the supremum of an r-dimensionally indexed Lévy sheet on $[0, \frac{3}{2}q(x-q)^{(d-r)/r}]^r$. Since $\frac{3}{2}q(x-q)^{(d-r)/r} \leq \frac{3}{2}qx^{d-1} < \frac{3}{2}c$ by the assumption of the lemma, we have, by Lemma 4.3,

$$\mathbb{E}e^{\theta M(\varepsilon_1, \dots, \varepsilon_d)} < C(c, \theta).$$

To finish the proof of the lemma use (11). □

4.3 Cubes of nearly optimal size

Idea of the proof. Now we are ready to start the proof of Theorem 2.1. We are interested in the high-crossing probability $\mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) > u_n]$. Intuitively, too small or too large cubes have asymptotically no chance to contribute to the above probability (for large cubes this is due to the assumption that the mean of the Lévy noise is negative). We shall see later that, asymptotically, the probability $\mathbb{P}[\mathcal{Z}(R) > u_n]$ achieves its maximum if the volume of the cube R is equal to v_n (equivalently, if its side length is equal to l_n). Furthermore, we shall see that cubes of volume differing from the optimal volume v_n by a quantity of order more than $\sqrt{v_n}$ have asymptotically no chance to contribute to the extremes of the field $\mathcal{Z}$.

In this section we are dealing with cubes of nearly optimal size, that is with cubes whose volume differs from v_n by a quantity of order $\sqrt{v_n}$. To be more precise, we fix a very large $A > 0$ and define

$$l_n^- = (v_n - A\sqrt{v_n})^{1/d}, \;\; l_n^+ = (v_n + A\sqrt{v_n})^{1/d}. \tag{12}$$

The main result of this subsection is Lemma 4.11 below, in which the limit as $n \to \infty$ of $\mathbb{P}[\max_{R \in \mathcal{R}_A(n)} \mathcal{Z}(R) \le u_n]$ is calculated, where $\mathcal{R}_A(n)$ is the set of cubes from $\mathcal{R}(n)$ whose side length is in the interval $[l_n^-, l_n^+]$.
Cubes of nearly optimal size. First we evaluate the high crossing probability $\mathbb{P}[\mathcal{Z}(R) > u_n]$ for cubes R having the optimal volume v_n.

Lemma 4.5. *We have as $n \to \infty$*

$$\mathbb{P}[\xi(v_n) > u_n] \sim \frac{\sqrt{\alpha^*}}{\sqrt{2\pi}\theta^* \sigma^*} \frac{1}{\sqrt{u_n}} e^{-\theta^* u_n}.$$

Proof. This follows from Petrov's Theorem 4.1. □

Now we consider cubes with volume differing from the optimal one by a quantity of order $\sqrt{v_n}$. Comparably to cubes of optimal volume, the high crossing probability changes by a constant factor.

Lemma 4.6. *We have as $n \to \infty$*

$$\mathbb{P}[\xi(v_n + s\sqrt{v_n}) > u_n + t] \sim e^{-\theta^* t} e^{-\frac{(\alpha^* s)^2}{2\sigma^{*2}}} \mathbb{P}[\xi(v_n) > u_n].$$

The above holds uniformly in s, t as long as $s = O(1)$ and $t = o(\sqrt{u_n})$.

Proof. Let $\alpha_n = (u_n + t)/(v_n + s\sqrt{v_n})$. Note that $\lim_{n\to\infty} \alpha_n = \alpha^*$. We obtain by Petrov's Theorem 4.1

$$\mathbb{P}[\xi(v_n + s\sqrt{v_n}) > (u_n + t)] \sim \frac{1}{\sqrt{2\pi}\theta^* \sigma^*} \frac{1}{\sqrt{v_n}} \exp\left\{-(u_n + t) J(\alpha_n)\right\}.$$

Now, an easy calculation shows that $\alpha_n = \alpha^*(1 - s/\sqrt{v_n} + o(1/\sqrt{v_n}))$. Using Lemma 4.1 we obtain

$$J(\alpha_n) = \theta^* + \frac{1}{2} \frac{\alpha^* s^2}{\sigma^{*2} v_n} + o\left(\frac{1}{v_n}\right).$$

It follows that

$$\mathbb{P}[\xi(v_n + s\sqrt{v_n}) > (u_n + t)] \sim \left(\frac{1}{\sqrt{2\pi}\theta^* \sigma^*} \frac{1}{\sqrt{v_n}} e^{-\theta^* u_n}\right) e^{-\theta^* t} e^{-\frac{(\alpha^* s)^2}{2\sigma^{*2}}}.$$

The statement of the lemma follows by noting that the first factor on the right-hand side is asymptotically equivalent to $\mathbb{P}[\xi(v_n) > u_n]$ by Lemma 4.5. □

Let $q_n = l_n^{1-d} = v_n^{(1-d)/d}$. Note that if $d = 1$ then $q_n = 1$, whereas otherwise $\lim_{n\to\infty} q_n = 0$. In the next lemma, we consider a high-crossing probability over a set of size of order q_n in the space of cubes.

Lemma 4.7. *For $x > 0$ and a fixed $m \in \mathbb{N}$ define a set of cubes*

$$\mathcal{B}_x^m(n) = [-mq_n/2, mq_n/2]^d \times [x, x + mq_n].$$

Then, for some constant $H_m > 0$, the following asymptotic equality holds as $n \to \infty$ uniformly in x as long as $x \in [l_n^-, l_n^+]$

$$\mathbb{P}\left[\max_{R \in \mathcal{B}_x^m(n)} \mathcal{Z}(R) > u_n\right] \sim H_m \mathbb{P}[\xi(x^d) > u_n]. \tag{13}$$

Proof. Let $R_0 = [-(x - mq_n)/2, (x - mq_n)/2]^d$ be the intersection of all cubes from $\mathcal{B}_x^m(n)$. Note that $|R_0| = |x - mq_n|^d = x^d + O(1)$ as $n \to \infty$. Applying Lemma 4.6 twice, we obtain that

$$\mathbb{P}[\mathcal{Z}(R_0) > u_n - t] \sim e^{\theta^* t} \mathbb{P}[\xi(x^d) > u_n]. \tag{14}$$

Let

$$M_n = \max_{R \in \mathcal{B}_x^m(n)} \mathcal{Z}(R) - \mathcal{Z}(R_0)$$

Then it is easy to see that $\mathcal{Z}(R_0)$ and M_n are independent and that M_n converges in distribution as $n \to \infty$ to the random variable

$$M_\infty = \max_{(l_1, \ldots, l_d; l) \in [-1/2, 1/2]^d \times [0,1]} \sum_{i=1}^d (\xi_i(l_i + \frac{l+m}{2}) - \xi_i(l_i - \frac{l+m}{2})),$$

where $\xi_1(\cdot), \ldots, \xi_d(\cdot)$ are independent copies of the Lévy process $\{\xi(t), t \in \mathbb{R}\}$. Denote the probability on the left-hand side of (13) by P_n. Then

$$P_n = \int_{-\infty}^{\infty} \mathbb{P}[\mathcal{Z}(R_0) > u_n - t] d\mathbb{P}[M_n = t].$$

Using (14), we obtain, at least formally, that as $n \to \infty$

$$P_n \sim \left(\int_{-\infty}^{\infty} e^{\theta^* t} d\mathbb{P}[M_\infty = t]\right) \mathbb{P}[\xi(x^d) > u_n],$$

which proves the lemma with $H_m = \mathbb{E}e^{\theta^* M_\infty}$. In the rest of the proof we justify this step. Take $T > 0$ large. We have

$$\begin{aligned} \frac{P_n}{\mathbb{P}[\xi(x^d) > u_n]} &= \int_{-\infty}^{\infty} \frac{\mathbb{P}[\mathcal{Z}(R_0) > u_n - t]}{\mathbb{P}[\xi(x^d) > u_n]} d\mathbb{P}[M_n = t] \\ &= \int_{-T}^{+T} + \int_{T}^{u_n^{1/3}} + \int_{u_n^{1/3}}^{\infty} + \int_{-\infty}^{-T} \\ &= I + II + III + IV. \end{aligned}$$

Since by Lemma 4.6 the convergence in (14) is uniform for $t \in [-T, T]$ and since M_n converges in distribution to M_∞, we obtain

$$\lim_{n\to\infty} I = \int_{-T}^{T} e^{\theta^* t} d\mathbb{P}[M_\infty = t].$$

The convergence in (14) remains uniform for $t = o(\sqrt{u_n})$. Using the fact that by Lemma 4.4, applied to M_n, $\mathbb{E}e^{\theta M_n} < C(\theta, m)$ for every $\theta < \theta_0$, we obtain

$$II \le \int_{T}^{u_n^{1/3}} Ce^{\theta^* t} d\mathbb{P}[M_n = t] < Ce^{-\delta T}$$

for some $\delta > 0$. To estimate the third term note that $\mathbb{P}[\mathcal{Z}(R_0) > u_n - t] \le e^{-\theta^*(u_n - t)}$ by Corollary 4.1 and $\mathbb{P}[\xi(x^d) > u_n] \ge cu_n^{-1/2} e^{-\theta^* u_n}$ by Lemma 4.6. Thus

$$III \le \int_{u_n^{1/3}}^{\infty} \frac{e^{-\theta^*(u_n - t)}}{cu_n^{-1/2} e^{-\theta^* u_n}} d\mathbb{P}[M_n = t] \le Cu_n^{1/2} \int_{u_n^{1/3}}^{\infty} e^{\theta^* t} d\mathbb{P}[M_n = t].$$

The right-hand side of the above inequality converges to 0 as $n \to \infty$ since $\mathbb{E}e^{\theta M_n} < C$ for every $\theta < \theta_0$ by Lemma 4.4. Thus, $\lim_{n\to\infty} III = 0$.
We estimate the last term:

$$IV \le \mathbb{P}[\mathcal{Z}(R_0) > u_n + T]/\mathbb{P}[\xi(x^d) > u_n].$$

It follows from Lemma 4.6 applied twice that the right-hand side of the above inequality converges to $e^{-\theta^* T}$ as $n \to \infty$ and hence $\limsup_{n\to\infty} IV \le e^{-\theta^* T}$. The statement of the lemma follows from above by letting first $n \to \infty$ and then $T \to \infty$. □

Let s_n be a sequence satisfying $s_n = O(l_n)$, $s_n > 1$. In the next lemma, we evaluate the high crossing probability of the scan statistic taken over the set of all cubes of nearly optimal volume with centers contained in $[0, s_n]^d$.

Lemma 4.8. *Let $\mathcal{L}_A(n) = [0, s_n]^d \times [l_n^-, l_n^+]$. Then we have as $n \to \infty$*

$$\mathbb{P}\left[\max_{R\in\mathcal{L}_A(n)} \mathcal{Z}(R) > u_n\right] \sim H \left(\int_{-A}^{A} e^{-\frac{(\alpha^* t)^2}{2\sigma^{*2}}} dt\right) s_n^d v_n^{d-(1/2)} \mathbb{P}[\xi(v_n) > u_n]. \quad (15)$$

Here, $H \in (0, \infty)$ is a constant defined by

$$H = \frac{1}{d} \lim_{m\to\infty} \frac{H_m}{m^{d+1}}. \quad (16)$$

Proof. Define

$$\mathcal{L}_A^m(n) = mq_n \mathbb{Z}^{d+1} \cap \mathcal{L}_A(n).$$

For $R = (x_1, \ldots, x_d; x) \in \mathcal{L}_A^m(n)$ let

$$\mathcal{B}_R^m(n) = [x_1 - \frac{q_n}{2}, x_1 + (m - \frac{1}{2})q_n] \times \ldots \times [x_d - \frac{q_n}{2}, x_d + (m - \frac{1}{2})q_n] \times [x, x + mq_n]$$

and define $B_R^m(n) = B_{x_1,\ldots,x_d;x}^m(n)$ to be the random event $\{\max_{Q \in \mathcal{B}_R^m(n)} \mathcal{Z}(Q) > u_n\}$. Denote the probability on the left-hand side of (15) by P_n. Then

$$P_n \le S_1^m(n), \tag{17}$$

where

$$S_1^m(n) = \sum_{R \in \mathcal{L}_A^m(n)} \mathbb{P}[B_R^m(n)] = \left(\frac{s_n}{mq_n}\right)^d \sum_{x \in mq_n\mathbb{Z} \cap [l_n^-, l_n^+]} \mathbb{P}[B_{0,\ldots,0;x}^m(n)].$$

Applying Lemma 4.7 and then Lemma 4.6

$$\mathbb{P}[B_{0,\ldots,0;x}^m(n)] \sim H_m \mathbb{P}[\xi(v_n) > u_n] \exp\left\{-\frac{\alpha^{*2}}{2\sigma^{*2}} \frac{(x^d - v_n)^2}{v_n}\right\}.$$

If the values of x are in $mq_n\mathbb{Z} \cap [l_n^-, l_n^+]$, then the values of $(x^d - v_n)/\sqrt{v_n}$ form a lattice in the interval $[-A, A]$, whose (variable) mesh size is $\approx md/\sqrt{v_n}$. Thus, approximating the Riemann sum by an integral, we obtain

$$\sum_{x \in mq_n\mathbb{Z} \cap [l_n^-, l_n^+]} \mathbb{P}[B_{0,\ldots,0;x}^m(n)] \sim H_m \mathbb{P}[\xi(v_n) > u_n] \left(\frac{\sqrt{v_n}}{md} \int_{-A}^{A} e^{-\frac{(\alpha^* t)^2}{2\sigma^{*2}}} dt\right).$$

This shows that

$$S_1^m(n) \quad \sim \quad \frac{1}{dm^{d+1}} H_m \left(\int_{-A}^{A} e^{-\frac{(\alpha^* t)^2}{2\sigma^{*2}}} dt\right) s_n^d v_n^{d-(1/2)} \mathbb{P}[\xi(v_n) > u_n]. \tag{18}$$

Since the above is true for every m, we obtain by letting $m \to \infty$ that the left-hand side of (15) is asymptotically not greater than the right-hand side of (15). In order to prove the converse we use the Bonferroni inequality

$$P_n \ge S_1^m(n) - S_2^m(n), \tag{19}$$

where $S_1^m(n)$ is as above and

$$S_2^m(n) = \sum_{R_1, R_2 \in \mathcal{L}_A^m(n),\ R_1 \ne R_2} \mathbb{P}[B_{R_1}^m(n) \cap B_{R_2}^m(n)].$$

The sum $S_1^m(n)$ was already treated above. The proof will be finished in Lemma 4.10, where it will be shown that $S_2^m(n)$ can be asymptotically ignored as $n \to \infty$ and $m \to \infty$.

For a cube $R = (x_1, \ldots, x_d; x)$ let

$$\mathcal{B}_R(n) = [x_1 - q_n/2, x_1 + q_n/2] \times \ldots \times [x_d - q_n/2, x_d + q_n/2] \times [x, x + q_n]$$

and let $B_R(n)$ be the random event $\{\sup_{R \in \mathcal{B}_R(n)} \mathcal{Z}(R) > u_n\}$. Let $R_1, R_2 \in \mathcal{L}_A(n)$ be two cubes. Denote $\Delta(R_1, R_2) = |R_1 \triangle R_2|$, the volume of the symmetric difference of R_1 and R_2.

Lemma 4.9. *For arbitrary cubes $R_1, R_2 \in \mathcal{L}_A(n)$ we have*

$$\mathbb{P}[B_{R_1}(n) \cap B_{R_2}(n)] \leq Ce^{-\delta\Delta(R_1, R_2)} \mathbb{P}[\xi(v_n) > u_n].$$

Proof. For a cube $R = (x_1, \ldots, x_d; x)$, $h = (h_1, \ldots, h_d)$, $h_i > 0$, and $(\varepsilon_1, \ldots, \varepsilon_d) \in \{-1, 0, 1\}^d$ let

$$R(\varepsilon_1, \ldots, \varepsilon_d; h) = I_1(\varepsilon_1; h_1) \times \ldots \times I_d(\varepsilon_d; h_d),$$

where

$$I_j(\varepsilon_j; h_j) = \begin{cases} [x_j - (x - q_n)/2, x_j + (x - q_n)/2], & \text{if } \varepsilon_j = 0, \\ [x_j + (x - q_n)/2, x_j + (x - q_n)/2 + h_j], & \text{if } \varepsilon_j = 1, \\ [x_j - (x - q_n)/2 - h_j, x_j - (x - q_n)/2], & \text{if } \varepsilon_j = -1. \end{cases}$$

Let $R_1(\varepsilon_1, \ldots, \varepsilon_d; h)$ and $R_2(\varepsilon_1, \ldots, \varepsilon_d; h)$ be defined analogously with R replaced by R_1 resp. R_2. Let $h(n) = (\frac{3}{2}q_n, \ldots, \frac{3}{2}q_n)$ and

$$\bar{R}_1 = \cup_{(\varepsilon_1, \ldots, \varepsilon_d) \in \{-1,0,1\}^d} R_1(\varepsilon_1, \ldots, \varepsilon_d; h(n)).$$

Define

$$\begin{aligned} R_2'(\varepsilon_1, \ldots, \varepsilon_d; h) &= R_2(\varepsilon_1, \ldots, \varepsilon_d; h) \cap \bar{R}_1, \\ R_2''(\varepsilon_1, \ldots, \varepsilon_d; h) &= R_2(\varepsilon_1, \ldots, \varepsilon_d; h) \backslash \bar{R}_1. \end{aligned}$$

Note that $R_0' := R_2'(0, \ldots, 0; h)$ and $R_0'' := R_2''(0, \ldots, 0; h)$ do not depend on h. Let

$$\begin{aligned} M' &= \sum_{\{\varepsilon_1, \ldots, \varepsilon_d\} \in \{-1,0,1\}^d \backslash \{0, \ldots, 0\}} \sup_{h \in [0, 3q_n/2]^d} \mathcal{Z}(R_2'(\varepsilon_1, \ldots, \varepsilon_d; h)), \\ M'' &= \sum_{\{\varepsilon_1, \ldots, \varepsilon_d\} \in \{-1,0,1\}^d \backslash \{0, \ldots, 0\}} \sup_{h \in [0, 3q_n/2]^d} \mathcal{Z}(R_2''(\varepsilon_1, \ldots, \varepsilon_d; h)). \end{aligned}$$

Finally, fix some small $a > 0$ and let

$$\begin{aligned} B' &= \{\mathcal{Z}(R_0') + M' > u_n + a\Delta\}, \\ B'' &= \{\mathcal{Z}(R_0'') + M'' > -a\Delta\}. \end{aligned}$$

We have trivially

$$\begin{aligned}\mathbb{P}[B_{R_1}(n)\cap B_{R_2}(n)] &= \mathbb{P}[B_{R_1}(n)\cap B_{R_2}(n)\cap B'] + \mathbb{P}[(B_{R_1}(n)\cap B_{R_2}(n))\backslash B'] \\ &\leq \mathbb{P}[B'] + \mathbb{P}[B_{R_1}(n)\cap B''] \\ &= \mathbb{P}[B'] + \mathbb{P}[B_{R_1}(n)]\mathbb{P}[B''].\end{aligned}$$

By Lemma 4.7 with $m = 1$ we have $\mathbb{P}[B_{R_1}(n)] \leq C\mathbb{P}[\xi(v_n) > u_n]$. Thus, in order to prove the lemma we need to show the following two inequalities

$$\mathbb{P}[B'] \leq Ce^{-\delta\Delta(R_1,R_2)}\mathbb{P}[\xi(v_n) > u_n], \tag{20}$$

$$\mathbb{P}[B''] \leq Ce^{-\delta\Delta(R_1,R_2)}. \tag{21}$$

We prove (20). Take $\varepsilon > 0$ sufficiently small. We write $\Delta := \Delta(R_1, R_2)$

$$\begin{aligned}\mathbb{P}[B'] &= \int_0^\infty \mathbb{P}[\mathcal{Z}(R_0') \geq u_n + a\Delta - t]d\mathbb{P}[M' = t] = \int_0^{(1-\varepsilon)u_n} + \int_{(1-\varepsilon)u_n}^\infty \\ &= I + II.\end{aligned}$$

To estimate I suppose first that $|R_0'| > \varepsilon' v_n$, where ε' is much smaller than ε. As in Lemma 4.4, we have $\mathbb{E}e^{\theta M'} < C(\theta)$ for every $\theta < \theta_0$. Then by Petrov's theorem

$$\begin{aligned}I &= \int_0^{(1-\varepsilon)u_n} \mathbb{P}[\mathcal{Z}(R_0') \geq u_n + a\Delta - t]d\mathbb{P}[M' = t] \\ &\leq \int_0^{(1-\varepsilon)u_n} Cu_n^{-1/2}e^{-\theta^*(u_n+a\Delta-t)}d\mathbb{P}[M' = t] \\ &\leq Cu_n^{-1/2}e^{-\theta^* u_n}e^{-\delta\Delta}\int_0^\infty e^{\theta^* t}d\mathbb{P}[M' = t] \\ &\leq Ce^{-\delta\Delta}\mathbb{P}[\xi(v_n) > u_n].\end{aligned}$$

Now suppose that $|R_0'| < \varepsilon' v_n$. Then, by Lemma 4.2, and if ε' is small enough,

$$I \leq \mathbb{P}[\mathcal{Z}(R_0') > \varepsilon u_n] < e^{-(1+\delta)\theta^* u_n}.$$

To estimate II note that by Lemma 4.4 and if ε is sufficiently small

$$II \leq \mathbb{P}[M' > (1-\varepsilon)u_n] \leq e^{-(1+\delta)\theta^* u_n}.$$

This proves (20). We prove (21). By symmetry we may assume that $|R_1\backslash R_2| \leq |R_2\backslash R_1|$ and hence $|R_0''| \geq \Delta/2 - O(1)$. By Markov inequality, for $t > 0$ small,

$$\begin{aligned}\mathbb{P}[B''] &= \mathbb{P}[\mathcal{Z}(R_0'') + M'' > -a\Delta] \leq e^{ta\Delta}\mathbb{E}e^{t\mathcal{Z}(R_0'')}\mathbb{E}e^{tM''} \leq Ce^{ta\Delta}\mathbb{E}e^{t\mathcal{Z}(R_0'')} \\ &= Ce^{ta\Delta + |R_0''|\varphi(t)} \leq Ce^{\Delta(2ta+\varphi(t))/2}.\end{aligned}$$

Now, since $a > 0$ is small enough and $\varphi'(0) < 0$, we may choose $t > 0$ so small that $2ta + \varphi(t) < 0$. This proves (21). □

Lemma 4.10. *With the notation of Lemma 4.8 and its proof*

$$\lim_{m\to\infty}\limsup_{n\to\infty} S_2^m(n)/(s_n^d v_n^{d-(1/2)}\mathbb{P}[\xi(v_n)>u_n])=0.$$

Proof. For each $R\in\mathcal{L}_A(n)$ the set $\mathcal{B}_R^m(n)$ may be written as a union of m^{d+1} sets of the form $\mathcal{B}_Q(n)$, where $Q\in q_n\mathbb{Z}^{d+1}\cap\mathcal{L}_A(n)$. For two cubes $Q_1,Q_2\in q_n\mathbb{Z}^{d+1}\cap\mathcal{L}_A(n)$ we write $Q_1\sim^m Q_2$ if there is $R\in\mathcal{L}_A^m(n)$ such that both $\mathcal{B}_{Q_1}(n)$ and $\mathcal{B}_{Q_2}(n)$ are subsets of $\mathcal{B}_R^m(n)$. It is not difficult to see that

$$S_2^m(n)\le\sum_{\substack{Q_1,Q_2\in q_n\mathbb{Z}^{d+1}\cap\mathcal{L}_A^n\\ Q_1\nsim^m Q_2}}\mathbb{P}[B_{Q_1}(n)\cap B_{Q_2}(n)].$$

Applying Lemma 4.9 we obtain

$$S_2^m(n)\le C\mathbb{P}[\xi(v_n)>u_n]\sum_{\substack{Q_1,Q_2\in q_n\mathbb{Z}^{d+1}\cap\mathcal{L}_A^n\\ Q_1\nsim^m Q_2}}e^{-\delta\Delta(Q_1,Q_2)}.$$

If $Q_1=q_nw_1$, $Q_2=q_nw_2$ for $w_1,w_2\in\mathbb{Z}^{d+1}$ then $\Delta(Q_1,Q_2)>c\|w_1-w_2\|$ for some $c>0$, where $\|\cdot\|$ is any norm on $\mathbb{R}^{d+1}$. The lattice $\mathbb{Z}^{d+1}$ can be decomposed into a disjoint union of discrete cubes of size-length m, the cubes having the form $w+K_m$ for $w\in m\mathbb{Z}^{d+1}$, where $K_m=\{0,\dots,m-1\}^{d+1}$. For $w_1,w_2\in\mathbb{Z}^{d+1}$ we write $w_1\sim^m w_2$ if w_1 and w_2 are contained in the same cube of the form described above. It is clear that $Q_1\sim^m Q_2$ iff $w_1\sim^m w_2$. It follows

$$S_2^m(n)\le C\mathbb{P}[\xi(v_n)>u_n]\sum_{\substack{w_1,w_2\in\mathbb{Z}^{d+1}\cap q_n^{-1}([0,s_n]^d\times[l_n^+,l_n^-])\\ w_1\nsim^m w_2}}e^{-\delta\|w_1-w_2\|}.$$

The set $\mathbb{Z}^{d+1}\cap q_n^{-1}([0,s_n]^d\times[l_n^+,l_n^-])$ contains $O(1)s_n^d v_n^{d-(1/2)}$ points. If (for c sufficiently large) $\|w_1-w_2\|>cm$ then $w_1\nsim^m w_2$. Using this, we obtain

$$S_2^m(n)\le C\mathbb{P}[\xi(v_n)>u_n]s_n^d v_n^{d-(1/2)}(I+II),$$

where

$$I=\sum_{w\in\mathbb{Z}^{d+1},\|w\|\ge cm}e^{-\delta\|w\|},\quad II=\frac{1}{m^{d+1}}\sum_{\substack{w_1\in K_m,w_2\in\mathbb{Z}^{d+1}\\ \|w_1-w_2\|\le cm}}e^{-\delta\|w_1-w_2\|}.$$

Both I and II do not depend on n and a straightforward calculation shows that $\lim_{m\to\infty}I=\lim_{m\to\infty}II=0$. This finishes the proof of Lemma 4.10. □

Now we can finish the proof of Lemma 4.8. Equation (15), and in particular the existence of the limit in (16), follows from the Bonferroni inequalities (17), (19) as well as from the above asymptotic equalities for $S_1^m(n)$, $S_2^m(n)$. It remains only to show that $H > 0$.
We have

$$P_n \geq \sum_{R \in \mathcal{L}_A^m(n)} \mathbb{P}[B_R(n)] - \sum_{\substack{R_1, R_2 \in \mathcal{L}_A^m(n) \\ R_1 \neq R_2}} \mathbb{P}[B_{R_1}(n) \cap B_{R_2}(n)] = I - II.$$

By the above I is asymptotically greater than $cm^{-(d+1)} s_n^d v_n^{d-(1/2)} \mathbb{P}[\xi(v_n) > u_n]$ for some $c > 0$, whereas

$$II \leq C\mathbb{P}[\xi(v_n) > u_n] s_n^d v_n^{d-(1/2)} m^{-(d+1)} \sum_{w \in m\mathbb{Z}^{d+1}} e^{-\delta \|w\|}.$$

It follows, that if m is sufficiently large, then $P_n \geq I - II \geq cs_n^d v_n^{d-(1/2)} \mathbb{P}[\xi(v_n) > u_n]$ for some $c > 0$. It follows that the constant H in (15) is positive. $\square$

Remark 4.1. *If the constant H^* in* (5) *is defined by $H^* = \log H + (d-1)\log d - d \log(\alpha^* \theta^*)$, then the statement of Lemma 4.8 may be written as*

$$\mathbb{P}\left[\max_{R \in \mathcal{L}_A(n)} \mathcal{Z}(R) > u_n\right] \sim D_A \frac{e^{-\tau}}{(n/s_n)^d},$$

where $D_A = \alpha^* (\sqrt{2\pi}\sigma^*)^{-1} \int_{-A}^{A} e^{-\frac{(\alpha^* t)^2}{2\sigma^{*2}}} dt$.

Finally, we are ready to prove the main result of this subsection.

Lemma 4.11. *Let $\mathcal{R}_A(n) = [0, n]^d \times [l_n^-, l_n^+]$. We have*

$$\lim_{n \to \infty} \mathbb{P}[\max_{R \in \mathcal{R}_A(n)} \mathcal{Z}(R) \leq u_n] = \exp\left\{-D_A e^{-\tau}\right\}.$$

Proof. The set $\mathcal{R}_A(n)$ can be decomposed in $(n/s_n)^d$ translates of the set $\mathcal{L}_A(n)$, which was considered in Lemma 4.8 and Remark 4.1. The lemma follows then from the Poisson limit theorem, the only problem to overcome is that the events under consideration are dependent.
Let $s_n = l_n$. For $(m_1, \ldots, m_d) \in \mathbb{Z}^d \cap [0, n/s_n]^d$ let

$$\mathcal{L}_A(n; m_1, \ldots, m_d) = [m_1 s_n, (m_1+1)s_n] \times \ldots \times [m_d s_n, (m_d+1)s_n] \times [l_n^-, l_n^+].$$

and define the random event $L_A(n; m_1, \ldots, m_d) = \{\sup_{R \in \mathcal{L}_A(n; m_1, \ldots, m_d)} \mathcal{Z}(R) > u_n\}$. By Lemma 4.8 and Remark 4.1

$$\mathbb{P}[L_A(n; m_1, \ldots, m_d)] \sim D_A \frac{e^{-\tau}}{(n/s_n)^d}. \tag{22}$$

Now we want to apply the Poisson limit theorem to the events

$$\{L_A(n; m_1, \ldots, m_d), (m_1, \ldots, m_d) \in \mathbb{Z}^d \cap [0, n/s_n]^d\}.$$

Note that the events are finite-range dependent. More precisely, the events $L_A(n; m'_1, \ldots, m'_d)$ and $L_A(n; m''_1, \ldots, m''_d)$ are independent if $|m'_i - m''_i| > 1$ for at least one $i = 1, \ldots, d$. In order to justify the use of the Poisson limit theorem we have to show that

$$\mathbb{P}[L_A(n; m'_1, \ldots, m''_d) \cap L_A(n; m''_1, \ldots, m''_d)] = o((n/s_n)^{-d}) \tag{23}$$

as $n \to \infty$, where $m''_i = m'_i + \varepsilon_i$, $\varepsilon_i \in \{-1, 0, 1\}$ are not all 0, see e.g. Theorem 1 in Arratia et al. (1989). To this end, we use Lemma 4.8 again, this time for $s'_n = 3l_n$. We obtain

$$\mathbb{P}[\bigcup_{(\varepsilon_1, \ldots, \varepsilon_d) \in \{-1,0,1\}^d} L_A(n, m'_1 + \varepsilon_1, \ldots, m'_d + \varepsilon_d)] \sim 3^d D_A \frac{e^{-\tau}}{(n/s_n)^d}.$$

On the other hand, by (22),

$$\sum_{(\varepsilon_1, \ldots, \varepsilon_d) \in \{-1,0,1\}^d} \mathbb{P}[L_A(n, m'_1 + \varepsilon_1, \ldots, m'_d + \varepsilon_d)] \sim 3^d D_A \frac{e^{-\tau}}{(n/s_n)^d}.$$

Then (23) follows by Bonferroni inequality. □

4.4 Cubes of non-optimal size

In this subsection we are dealing with cubes whose volume differs significantly from the optimal volume v_n. More precisely, we consider cubes with volume outside the interval $[v_n - A\sqrt{v_n}, v_n + A\sqrt{v_n}]$. We show that, if $A \to \infty$ and $n \to \infty$, these cubes do not contribute to the extremal behavior of the random field $\{\mathcal{Z}(R), R \in \mathcal{R}(n)\}$. Let $\varepsilon > 0$ be sufficiently small.

Lemma 4.12. *There is $\delta > 0$ such that the following inequality holds uniformly s and t as long as $|s| \le \varepsilon\sqrt{v_n}$ and $t = o(\sqrt{u_n})$*

$$\mathbb{P}[\xi(v_n + s\sqrt{v_n}) > u_n + t] \le Ce^{-\theta^* t} e^{-\delta s^2} \mathbb{P}[\xi(v_n) > u_n]. \tag{24}$$

Proof. Let $\alpha_n = (u_n + t)/(v_n + s\sqrt{v_n})$. Note that $\alpha_n \in [(1-\eta)\alpha^*, (1+\eta)\alpha^*]$ where $\eta = \eta(\varepsilon)$ is small if ε is small. By Petrov's theorem

$$\mathbb{P}[\xi(v_n + s\sqrt{v_n}) > u_n + t] \le C \frac{1}{\sqrt{v_n}} \exp\{-(u_n + t) J(\alpha_n)\}. \tag{25}$$

An easy calculation shows that for some $c > 0$

$$(\alpha_n - \alpha^*)^2 \ge cs^2/v_n.$$

Applying Lemma 4.1 we obtain that for some $\delta > 0$

$$J(\alpha_n) \geq \theta^* + \delta \frac{s^2}{v_n}.$$

Substituting this into (25), we obtain the statement of the lemma. $\square$

Let $A > 0$ be large. Recall that l_n^- and l_n^+ were defined in (12).

Lemma 4.13. *Define a set of cubes* $\mathcal{R}_1(n) = [0, n]^d \times [(1-\varepsilon)l_n, l_n^-]$ *and let*

$$P_n(A) = \mathbb{P}[\max_{R \in \mathcal{R}_1(n)} \mathcal{Z}(R) > u_n].$$

Then, for $c(A) = \limsup_{n\to\infty} P_n(A)$, *we have* $\lim_{A\to\infty} c(A) = 0$.

Proof. Recall that $q_n = l_n^{1-d}$. Define the set $\mathcal{B}_x(n) = [-q_n/2, q_n/2]^d \times [x, x + q_n]$. Then

$$P_n(A) \leq (n/q_n)^d \sum_{x \in q_n\mathbb{Z} \cap [(1-\varepsilon)l_n, l_n^-]} \mathbb{P}[\max_{R \in \mathcal{B}_x(n)} \mathcal{Z}(R) > u_n]. \tag{26}$$

Given $x \in [(1-\varepsilon)l_n, l_n^-]$ define a cube

$$R_0 = [-1/2(x - q_n), 1/2(x - q_n)]^d.$$

and $M = \max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) - \mathcal{Z}(R_0)$. Note that M and $\mathcal{Z}(R_0)$ are independent and $|R_0| = x^d + O(1)$. We have

$$\begin{aligned}
\mathbb{P}[\max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) > u_n] &\leq \mathbb{P}[\mathcal{Z}(R_0) > u_n] + \int_0^\infty \mathbb{P}[\mathcal{Z}(R_0) > u_n - t] d\mathbb{P}[M = t] \\
&= \mathbb{P}[\mathcal{Z}(R_0) > u_n] + \int_0^{u_n^{1/3}} + \int_{u_n^{1/3}}^\infty \\
&= I + II + III.
\end{aligned}$$

Let $s = s_n(x)$ be chosen such that $x^d = v_n - s\sqrt{v_n}$. By Lemma 4.12

$$I \leq Ce^{-\delta s^2} \mathbb{P}[\xi(v_n) > u_n].$$

We estimate the second term using first Lemma 4.12 and then Lemma 4.4:

$$\begin{aligned}
II &\leq Ce^{-\delta s^2} \mathbb{P}[\xi(v_n) > u_n] \int_0^{u_n^{1/3}} e^{\theta^* t} d\mathbb{P}[M = t] \\
&\leq Ce^{-\delta s^2} \mathbb{P}[\xi(v_n) > u_n].
\end{aligned}$$

Using Corollary 4.1 and then Lemma 4.4 the third term may be estimated by

$$\begin{aligned} III &\leq \int_{u_n^{1/3}}^{\infty} e^{-\theta^*(u_n-t)} d\mathbb{P}[M=t] = e^{-\theta^* u_n} \int_{u_n^{1/3}}^{\infty} e^{\theta^* t} d\mathbb{P}[M=t] \\ &\leq Ce^{-\theta^* u_n} e^{-\delta u_n^{1/3}}. \end{aligned}$$

Bringing all three estimates together and recalling Lemma 4.5 we obtain

$$\mathbb{P}[\max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) > u_n] \leq Cu_n^{-1/2} e^{-\theta^* u_n} e^{-\delta s^2} + Ce^{-\theta^* u_n} e^{-\delta u_n^{1/3}}.$$

It follows from (26) that $P_n(A) \leq I' + II'$, where

$$\begin{aligned} I' &= C(n/q_n)^d u_n^{-1/2} e^{-\theta^* u_n} \sum_{x\in q_n\mathbb{Z}\cap[(1-\varepsilon)l_n, l_n^-]} e^{-\delta s_n(x)^2}, \\ II' &= C(n/q_n)^d e^{-\theta^* u_n} e^{-\delta u_n^{1/3}} \sum_{x\in q_n\mathbb{Z}\cap[(1-\varepsilon)l_n, l_n^-]} 1. \end{aligned}$$

It is easy to see that $\lim_{n\to\infty} II' = 0$. We estimate I'. If $x \in q_n\mathbb{Z} \cap [(1-\varepsilon)l_n, l_n^-]$, then the possible values of $s_n(x)$ form a lattice in $[A,\infty)$ with mesh size $O(u_n^{-1/2})$. Thus, estimating the Riemann sum by an integral, we obtain

$$I' \leq C(n/q_n)^d e^{-\theta^* u_n} \int_A^{\infty} e^{-\delta s^2} ds < C \int_A^{\infty} e^{-\delta s^2} ds.$$

The statement of the lemma follows. □

Lemma 4.14. *Let* $\mathcal{R}_2(n) = [0,n]^d \times [0,(1-\varepsilon)l_n]$ *and define*

$$P_n = \mathbb{P}[\max_{R\in\mathcal{R}_2(n)} \mathcal{Z}(R) > u_n].$$

Then we have $\lim_{n\to\infty} P_n = 0$.

Proof. The proof starts similar to the proof of the previous lemma. For $x \in [0,(1-\varepsilon)l_n]$ define a set of cubes $\mathcal{B}_x(n) = [-q_n/2, q_n/2]^d \times [x, x+q_n]$. Then

$$P_n \leq (n/q_n)^d \sum_{x\in q_n\mathbb{Z}\cap[0,(1-\varepsilon)l_n]} \mathbb{P}[\max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) > u_n]. \tag{27}$$

Define, as in the proof of the previous lemma,

$$R_0 = [-1/2(x-q_n), 1/2(x-q_n)]^d.$$

and $M = \max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) - \mathcal{Z}(R_0)$ (if $x < q_n$, we set $R_0 = \emptyset$). Note that M and $\mathcal{Z}(R_0)$ are independent. We have

$$\begin{aligned}
\mathbb{P}[\max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) > u_n] &\leq \mathbb{P}[\mathcal{Z}(R_0) > u_n] + \int_0^\infty \mathbb{P}[\mathcal{Z}(R_0) > u_n - t] d\mathbb{P}[M = t] \\
&= \mathbb{P}[\mathcal{Z}(R_0) > u_n] + \int_0^{\varepsilon u_n/2} + \int_{\varepsilon u_n/2}^\infty \\
&= I + II + III.
\end{aligned}$$

To estimate the first term we use Lemma 4.2 and the fact that $|R_0| < (1 - \varepsilon)^d v_n$ (and thus $u_n/|R_0| > (1+\delta)\alpha^*$)

$$I \leq e^{-u_n J(u_n/|R_0|)} \leq e^{-(1+\delta)\theta^* u_n}.$$

The second term is estimated analogously, using Corollary 4.1 and Lemma 4.4

$$\begin{aligned}
II &\leq \int_0^{\varepsilon u_n/2} e^{-(u_n - t)J((u_n-t)/|R_0|)} d\mathbb{P}[M = t] \\
&\leq \int_0^{\varepsilon u_n/2} e^{-(1+\delta)\theta^*(u_n - t)} d\mathbb{P}[M = t] \\
&= e^{-(1+\delta)\theta^* u_n} \int_0^{\varepsilon u_n/2} e^{(1+\delta)\theta^* t} d\mathbb{P}[M = t] \leq C e^{-(1+\delta)\theta^* u_n}.
\end{aligned}$$

To estimate the third term we use again Corollary 4.1 and Lemma 4.4

$$\begin{aligned}
III &\leq \int_{\varepsilon u_n/2}^\infty e^{-\theta^*(u_n - t)} d\mathbb{P}[M = t] = e^{-\theta^* u_n} \int_{\varepsilon u_n/2}^\infty e^{\theta^* t} d\mathbb{P}[M = t] \\
&\leq C e^{-(1+\delta)\theta^* u_n}.
\end{aligned}$$

Bringing the three estimates together, we obtain

$$\mathbb{P}[\max_{R\in\mathcal{B}_x(n)} \mathcal{Z}(R) > u_n] \leq C e^{-(1+\delta)\theta^* u_n}.$$

It follows from (27) that

$$P_n \leq C(n/q_n)^d e^{-(1+\delta)\theta^* u_n} \sum_{x\in q_n\mathbb{Z}\cap[0,(1-\varepsilon)l_n]} 1,$$

which converges to 0 as $n \to \infty$. This finishes the proof. □

Lemma 4.15. *Define* $\mathcal{R}_1^+(n) = [0, n]^d \times [l_n^+, (1+\varepsilon)l_n]$ *and let*

$$P_n(A) = \mathbb{P}[\max_{R\in\mathcal{R}_1^+(n)} \mathcal{Z}(R) > u_n].$$

Then, for $c(A) = \limsup_{n\to\infty} P_n(A)$*, we have* $\lim_{A\to\infty} c(A) = 0$.

Proof. Analogous to the proof of Lemma 4.13. □

Lemma 4.16. *Let* $\mathcal{R}_2^+(n) = [0,n]^d \times [(1+\varepsilon)l_n, n]$ *and define*

$$P_n = \mathbb{P}[\max_{R\in\mathcal{R}_2^+(n)} \mathcal{Z}(R) > u_n].$$

Then we have $\lim_{n\to\infty} P_n = 0$.

Proof. Define $q_l = l^{1-d}$ and, for $x \in [l, l+1]$, define a set of cubes $\mathcal{B}_x = [-q_l/2, q_l/2]^d \times [x, x+q_l]$. Let M and R_0 be defined in previous lemmas. Then $P_n \le P_n' + P_n''$, where

$$\begin{aligned} P_n' &= \sum_{l=[(1+\varepsilon)l_n]}^{[l_n^2]} (nq_l^{-1})^d \sum_{x\in q_l\mathbb{Z}\cap[l,l+1]} \mathbb{P}[\max_{R\in\mathcal{B}_x} \mathcal{Z}(R) > u_n], \\ P_n'' &= \sum_{l=[l_n^2]}^{n} (nq_l^{-1})^d \sum_{x\in q_l\mathbb{Z}\cap[l,l+1]} \mathbb{P}[\max_{R\in\mathcal{B}_x} \mathcal{Z}(R) > u_n]. \end{aligned}$$

If $l \in [(1+\varepsilon)l_n, l_n^2]$, then we use the estimate

$$\begin{aligned} \mathbb{P}[\max_{R\in\mathcal{B}_x} \mathcal{Z}(R) > u_n] &\le \mathbb{P}[\mathcal{Z}(R_0) > u_n] + \int_0^\infty \mathbb{P}[\mathcal{Z}(R_0) > u_n - t] d\mathbb{P}[M=t] \\ &= I + II. \end{aligned}$$

The first term may be estimated using Lemma 4.2 and the fact that $|R_0| \ge (1+\varepsilon)v_n$

$$I \le e^{-(1+\delta)\theta^* u_n}.$$

To estimate the second term use additionally Lemma 4.4

$$\begin{aligned} II &\le \int_0^\infty e^{-(1+\delta)\theta^*(u_n-t)} d\mathbb{P}[M=t] \le e^{-(1+\delta)\theta^* u_n} \int_0^\infty e^{(1+\delta)\theta^* t} d\mathbb{P}[M=t] \\ &< Ce^{-(1+\delta)\theta^* u_n}. \end{aligned}$$

Using this, we obtain $P_n' \le Cn^d l_n^C e^{-(1+\delta)\theta^* u_n}$, which converges to 0 as $n\to\infty$. Now suppose that $l \in [l_n^2, n]$. Then (the constant b is large)

$$\mathbb{P}[\max_{R\in\mathcal{B}_x} \mathcal{Z}(R) > u_n] \le \mathbb{P}[\mathcal{Z}(R_0) > -bu_n] + \mathbb{P}[M > bu_n] = I + II.$$

Since $|R_0| > cl_n^{2d} = cv_n^2$, $c > 0$, the first term may be estimated using e.g. Petrov's theorem

$$\mathbb{P}[\mathcal{Z}(R_0) > -bu_n] \le Ce^{-cv_n^2 I(bu_n/v_n^2)} < Ce^{-\delta v_n^2} < Ce^{-\delta u_n^2} < Cn^{-D}$$

for any given D. To estimate the second term we use Lemma 4.4

$$\mathbb{P}[M > bu_n] < Ce^{-b\theta^* u_n} < 1/n^D$$

for any given D if b is sufficiently large. Bringing everything together we obtain $P_n'' \le Cn^{d^2+1} l_n^C n^{-D}$, which converges to 0 for D large. Thus, $\lim_{n\to\infty} P_n'' = 0$. This finishes the proof of the lemma. □

4.5 Proof of Theorem 2.1

Now we are able to finish the proof of Theorem 2.1. For every $A > 0$ we have

$$\mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n] \leq \mathbb{P}[\sup_{R \in \mathcal{R}_A(n)} \mathcal{Z}(R) \leq u_n].$$

Letting $n \to \infty$ and applying Lemma 4.11 to the right-hand side we obtain

$$\limsup_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n] \leq \exp\left\{-D_A e^{-\tau}\right\}.$$

Now, letting $A \to \infty$ and using the fact that $\lim_{A \to \infty} D_A = 1$, we obtain

$$\limsup_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n] \leq \exp\left\{-e^{-\tau}\right\}. \tag{28}$$

On the other hand, we have

$$\mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n] \geq \mathbb{P}[\sup_{R \in \mathcal{R}_A(n)} \mathcal{Z}(R) \leq u_n] - \mathbb{P}[\sup_{R \in \mathcal{R}(n) \backslash \mathcal{R}_A(n)} \mathcal{Z}(R) > u_n].$$

Again, by Lemma 4.11 the first term on the right-hand side converges to $\exp\{-D_A e^{-\tau}\}$ as $n \to \infty$. By Lemmas 4.13, 4.14, 4.15, 4.16

$$\lim_{A \to \infty} \limsup_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(n) \backslash \mathcal{R}_A(n)} \mathcal{Z}(R) > u_n] = 0.$$

Thus, letting first $n \to \infty$ and then $A \to \infty$, we obtain

$$\liminf_{n \to \infty} \mathbb{P}[\sup_{R \in \mathcal{R}(n)} \mathcal{Z}(R) \leq u_n] \geq \exp\left\{-e^{-\tau}\right\}. \tag{29}$$

The statement of the theorem follows from (28) and (29). □

5 Proof of Theorem 1.1

In this section we deduce Theorem 1.1 from Theorem 2.1 using a relation between marked empirical process and compound Poisson process stated below.

Proof. First, let $\{(U_i, X_i), i = 1, \ldots, n\}$ be a marked empirical process as in Section 1. On the other hand, let $\{V_i, i \in \mathbb{N}\}$ be a Poisson point process on $\mathbb{R}^d$ with unit intensity. To each point V_i we attach a mark Y_i. We suppose that Y_i are i.i.d. with the same distribution as the marks X_i used in the construction of the marked empirical process and that Y_i's do not depend on V_i's. For a Borel set R let

$$\mathcal{Z}(R) = \sum_{i \in \mathbb{N}: V_i \in R} Y_i.$$

The compound Poisson process $\mathcal{Z}$ is an example of Lévy noise. Now we are going to show that $\mathcal{Z}$ satisfies conditions L1-L3 provided that X_i satisfy conditions X1-X3. First note that if φ is the logarithmic moment generating function of X_1 then the logarithmic moment generating function of $\mathcal{Z}([0,1]^d)$ is $\psi(t) = e^{\varphi(t)} - 1$. Thus, if φ is finite on $[0, \theta_0)$ and has a zero at θ^*, then the same holds for ψ. Finally, it is clear that if X_1 is non-lattice then $\mathcal{Z}([0,1]^d)$ is also non-lattice. This shows that conditions L1-L3 are satisfied.

We denote by $N_t = \#\{i \in \mathbb{N} : V_i \in [0,t]^d\}$ the number of points of the compound Poisson process, contained in the cube $[0,t]^d$. For $n \in \mathbb{N}$ let $T_n = \inf\{t > 0 : N_t = n+1\}$. Then we have the equality in distribution

$$\{\mathcal{X}_n(R), R \in \mathcal{R}(1)\} \sim \{\mathcal{Z}(T_n R), R \in \mathcal{R}(1)\}. \tag{30}$$

To see this consider the right-hand side of (30) conditioned on $\{T_n \in [t, t+dt]\}$, where dt is infinitesimal. The condition may be reformulated as $\{N_t = n; \exists i : V_i \in [0, t+dt]^d \backslash [0,t]^d\}$, and, under this condition, the square $[0,t]^d$ contains n points of the Poisson point process $\{V_i, i \in \mathbb{N}\}$, which have the same distribution as n points chosen independently and uniformly in $[0,t]^d$. Thus,

$$\{\mathcal{Z}(T_n R), R \in \mathcal{R}(1)\} \mid \{T_n \in [t, t+dt]\} \sim \{\mathcal{X}_n(R), R \in \mathcal{R}(1)\}.$$

Since this is true for every t, we obtain (30).

Define $T_n^+ = n + n^{2/3}$, $T_n^- = n - n^{2/3}$. Then $\mathbb{P}[T_n > T_n^+] = \mathbb{P}[N_{n+n^{2/3}} < n+1]$, which converges to 0 as $n \to \infty$ by the central limit theorem. Thus, $\lim_{n\to\infty} \mathbb{P}[T_n > T_n^+] = 0$ and analogously $\lim_{n\to\infty} \mathbb{P}[T_n < T_n^-] = 0$. We have

$$\begin{aligned}
\mathbb{P}[\sup_{R \in \mathcal{R}(1)} \mathcal{X}_n(R) \le u_n] &= \mathbb{P}[\sup_{R \in \mathcal{R}(T_n)} \mathcal{Z}(R) \le u_n] \\
&\le \mathbb{P}[\sup_{R \in \mathcal{R}(T_n)} \mathcal{Z}(R) \le u_n \cap T_n \ge T_n^-] + \mathbb{P}[T_n < T_n^-] \\
&\le \mathbb{P}[\sup_{R \in \mathcal{R}(T_n^-)} \mathcal{Z}(R) \le u_n] + \mathbb{P}[T_n < T_n^-].
\end{aligned}$$

Now the first term converges to $\exp\{-e^{-\tau}\}$ by Theorem 2.1 (note that $u_n = u_{T_n^-} + o(1)$) and the second term was shown to converge to 0. This shows that

$$\limsup_{n\to\infty} \mathbb{P}[\sup_{R \in \mathcal{R}(1)} \mathcal{X}_n(R) \le u_n] \le \exp\{-e^{-\tau}\}.$$

On the other hand

$$\begin{aligned}
\mathbb{P}[\sup_{R \in \mathcal{R}(1)} \mathcal{X}_n(R) \le u_n] &= \mathbb{P}[\sup_{R \in \mathcal{R}(T_n)} \mathcal{Z}(R) \le u_n] \\
&\ge \mathbb{P}[\sup_{R \in \mathcal{R}(T_n^+)} \mathcal{Z}(R) \le u_n \cap T_n \le T_n^+] \\
&\ge \mathbb{P}[\sup_{R \in \mathcal{R}(T_n^+)} \mathcal{Z}(R) \le u_n] - \mathbb{P}[T_n > T_n^-].
\end{aligned}$$

As above, the first term converges to $\exp\{-e^{-\tau}\}$ by Theorem 2.1, whereas the second converges to 0. This shows that

$$\liminf_{n\to\infty} \mathbb{P}[\sup_{R\in\mathcal{R}(1)} \mathcal{X}_n(R) \le u_n] \ge \exp\{-e^{-\tau}\}.$$

The proof is finished. □

Acknowledgements. The authors are grateful to M. Schlather for numerous useful remarks.

References

R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: The Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

P.J. Bickel and M.J. Wichura. Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Stat.*, 42:1656–1670, 1971.

H.P. Chan. Maxima of moving sums in a Poisson random field. *Available at http://arxiv.org/abs/0708.2764*, 2007.

R.A. Doney and R.A. Maller. Cramér's estimate for a reflected Lévy process. *Ann. Appl. Probab.*, 15(2):1445–1450, 2005.

J. Glaz and N. Balakrishnan. *Scan statistics and applications.* Statistics for Industry and Technology. Boston: Birkhäuser. xxi, 324 p., 1999.

J. Glaz, J. Naus, and S. Wallenstein. *Scan statistics.* Springer Series in Statistics. New York, NY: Springer. xv, 370 p., 2001.

J. Hüsler and V. Piterbarg. Limit theorem for maximum of the storage process with fractional Brownian motion as input. *Stochastic Processes Appl.*, 114(2):231–250, 2004.

D. L. Iglehart. Extreme values in the GI/G/1 queue. *Ann. Math. Stat.*, 43: 627–635, 1972.

J. Komlós and G. Tusnády. On sequences of "pure heads". *Ann. Probab.*, 3: 608–617, 1975.

S. Karlin and A. Dembo. Limit distributions of maximal segmental score among Markov-dependent partial sums. *Adv. Appl. Probab.*, 24(1):113–140, 1992.

M.R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. New York - Heidelberg - Berlin: Springer - Verlag., 336 p., 1983.

V.V. Petrov. On the probabilities of large deviations for sums of independent random variables. *Theor. Probab. Appl.*, 10:287–298, 1965.

J. Pickands. Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Am. Math. Soc.*, 145:75–86, 1969a.

J. Pickands. Upcrossing probabilities for stationary Gaussian processes. *Trans. Am. Math. Soc.*, 145:51–73, 1969b.

V.I. Piterbarg. *Asymptotic methods in the theory of Gaussian processes and fields.* . Translations of Mathematical Monographs. 148. Providence, RI: AMS., 206 p. , 1996.

V.I. Piterbarg and A.M. Kozlov. On large jumps of a Cramer random walk. *Theory Probab. Appl.*, 47(4):719–729, 2003.

E. Willekens. On the supremum of an infinitely divisible process. *Stochastic Processes Appl.*, 26:173–175, 1987.

A. J. Zeevi and P. W. Glynn. On the maximum workload of a queue fed by fractional Brownian motion. *Ann. Appl. Probab.*, 10(4):1084–1099, 2000.