Geometric properties of heavy-tailed random fields

Lettisia George
Supervisor: Andriy Olenko
La Trobe University

February 2014

1 Introduction

This project aims to investigate the asymptotic distribution of the volume of the excursion set for heavy-tailed random fields. It is based on and extends the work [2] which gives asymptotic results for the distribution of the volume of the excursion sets of Fisher-Snecdor and Student random fields under short and long range dependence conditions. My tasks included learning the relevant theory and running simulations of random fields to verify the results in [2] and in particular to examine the effect of increasing the excursion set level. A bonus challenge was to write my report in LaTeX.

A random field is a set of random variables defined at points in space [1]. The value of the random variable at each point has a particular distribution. From the distribution we can find statistical properties such as the mean and variance. For any random variable the distribution tells us the probability that the variable takes a certain value. In this project we use Gaussian, Student and Fisher-Snecdor random fields in which each value has the Gaussian, Student or Fisher-Snecdor distribution respectively. In addition for any set of points in a Gaussian random field the values are multivariate normally distributed. The distribution of values at a point is known as the marginal distribution. Student or Fisher-Snecdor random fields have heavy-tailed marginal distributions [4].

In this paper we use η to denote a Gaussian field, $F_{m,n}(x)$ for a Fisher field with m and n degrees of freedom and $T_n(x)$ for a Student field with n degrees of freedom. Student and Fisher fields can be constructed from Gaussian fields. For this reason Student and Fisher fields are known as Gaussian-related fields and some of the properties of Gaussian fields can be modified for use with these fields [1].

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia
Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

Consider the vector random field

$$\eta(x) = [\eta_1(x), ..., \eta_m(x), \eta_{m+1}(x), ..., \eta_{m+n}(x)]'$$

which consists of $n+m$ independent copies of a measurable mean-square continuous homogeneous isotropic zero-mean and unit variance Gaussian random field $\eta_1(x)$, $x \in \mathbb{R}^d$.

Definition 1 *The Fisher random field $F_{m,n}(x)$, $x \in \mathbb{R}^d$, is defined by*

$$F_{m,n}(x) := \frac{\frac{1}{m}\left(\eta_1^2(x) + \cdots + \eta_m^2(x)\right)}{\frac{1}{n}\left(\eta_{m+1}^2(x) + \cdots + \eta_{m+n}^2(x)\right)}, \quad x \in \mathbb{R}^d.$$

Definition 2 *The Student random field $T_n(x)$, $x \in \mathbb{R}^d$, is defined by*

$$T_n(x) := \frac{\eta_1(x)}{\sqrt{\frac{1}{n}\left(\eta_2^2(x) + \cdots + \eta_{n+1}^2(x)\right)}}, \quad x \in \mathbb{R}^d.$$

Note that $[T_n(x)]^2 = F_{1,n}(x),\ x \in \mathbb{R}^d$.

In order to further specify a random field we need to know the covariance between points. To simplify analysis we would like to consider covariance as a function of a single variable. So we consider isotropic random fields meaning that the covariance depends only on the distance between the points. Weak stationarity of a random field implies both homogeneity (translation invariance) and isotropy (rotation invariance). Weak stationarity means that statistics such as mean and variance are the same everywhere in the field.

The shape of the covariance function determines the dependence characteristics of the random field. We look at two classes of covariance in this paper, short range dependence and long range dependence. For short range dependence the value at a point is dependent on the points close to it whereas for long range dependence, points at large distances influence each other. Mathematically speaking the covariance of a random field with short range dependence is integrable while the covariance of a random field with long range dependence is not integrable.

In this paper we want to know about the volume of excursion set within a bounded observation window. The volume of the excursion set is also known as the sojourn measure. The excursion set is the set of points whose value is above a certain level a. Minkowski functionals are geometric properties including area, volume, perimeter, surface area, mean curvature and Euler characteristic. For an n-dimensional space there are n+1 Minkowski functionals. In this project we work with the first Minkowski functional which gives us length, area or volume depending on the number of spatial dimensions.

Consider a Jordan-measurable convex bounded set $\Delta \subset \mathbb{R}^d$, such that $|\Delta| > 0$ and Δ contains the origin in its interior. Δ is an initial bounded observation window. Let $\Delta(r), r > 0$, be the homothetic image of the set Δ, with the centre of homothety in the origin and the coefficient $r > 0$, that is $|\Delta(r)| = r^d |\Delta|$. So $\Delta(r)$ is our original window blown up by a factor of r.

Definition 3 *The first Minkowski functional is defined as*

$$M_r\{S\} := |\{x \in \Delta(r) : S(x) > a(r)\}| = \int_{\Delta(r)} \chi(S(x) > a(r))\, dx,$$

where $|\cdot|$ denotes the Lebesgue measure (length, area, volume) $\chi(\cdot)$ is an indicator function and $a(r)$ is a continuous nondecreasing function.

In the simplest case $a(r) = a$ is a constant. The functional $M_r\{S\}$ has an interpretation of the volume of the excursion set of the random field $S(x)$ above the constant level a, or the moving level $a(r)$. The aim of this paper is to investigate the distribution of the volume of the excursion set as the level $a(r)$ increases with r.

The paper [2] provides the asymptotic distributions for the volume of excursion sets of Fisher-Snecdor and Student random fields. One of the aims of this project was to work with these distributions and examine how they behave as the level $a(r)$ is increased.

In the following theorems $I_\mu(p, q)$ is the incomplete beta function defined in [2]. There are a number of constants that will not be defined here, (refer to [2]). I also will not detail any of the required assumptions about the fields again (refer to [2]). Most important to observe about these asymptotic distributions is whether they are normal or not and that each is obtained for different normalisations the Minkowski functional.

The asymptotic distribution of the excursion set for a Student field with short range dependence is normal.

The fist two theorems apply to the constant level $a(r) \equiv a$.

Theorem 1 *If the covariance matrix of the Student random field $T_n(x)$, $x \in \mathbb{R}^d$, satisfies the conditions as in [2] then*

$$r^{-d/2} M_r\{T_n\} - |\Delta|\, r^{d/2} \left(\frac{1}{2} - \frac{1}{2} \left(1 - I_{\frac{n}{n+a^2}} \left(\frac{n}{2}, \frac{1}{2} \right) \right) \cdot \operatorname{sgn}(a) \right) \xrightarrow{\mathcal{D}} \tilde{Y}_\Delta,$$

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia
Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

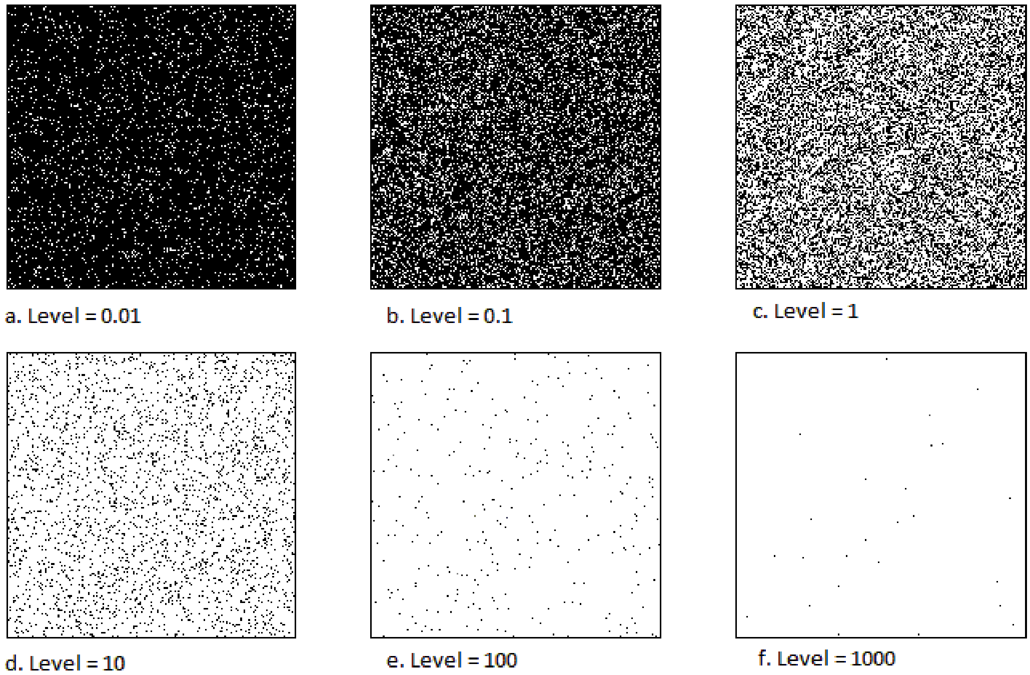

Figure 1: Excursion sets of a short range dependent Fisher field

where $r \to \infty$, $|\Delta|^{-1/2}\, \tilde{Y}_\Delta \sim N(0, \sigma_T^2)$,

$$\sigma_T^2 := \int_{\mathbb{R}^d} \mathbf{E}\left[\chi\left(T_n(0) > a\right)\chi\left(T_n(x) > a\right)\right] dx.$$

The asymptotic distribution of the excursion set for a Fisher field with short range dependence is normal.

Theorem 2 *If the covariance matrix of the Fisher random field $F_{m,n}(x)$, $x \in \mathbb{R}^d$, satisfies the conditions as in [2], then*

$$r^{-d/2} M_r \left\{F_{m,n}\right\} - |\Delta|\, r^{d/2}\left(1 - I_{\frac{ma}{n+ma}}\left(\frac{m}{2}, \frac{n}{2}\right)\right) \overset{\mathcal{D}}{\to} Y_\Delta,\ r \to \infty,$$

where $|\Delta|^{-1/2}\, Y_\Delta \sim N(0, \sigma_F^2(a))$,

$$\sigma_F^2(a) := \int_{\mathbb{R}^d} \mathbf{E}\left[\chi\left(F_{m,n}(0) > a\right)\chi\left(F_{m,n}(x) > a\right)\right] dx.$$

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

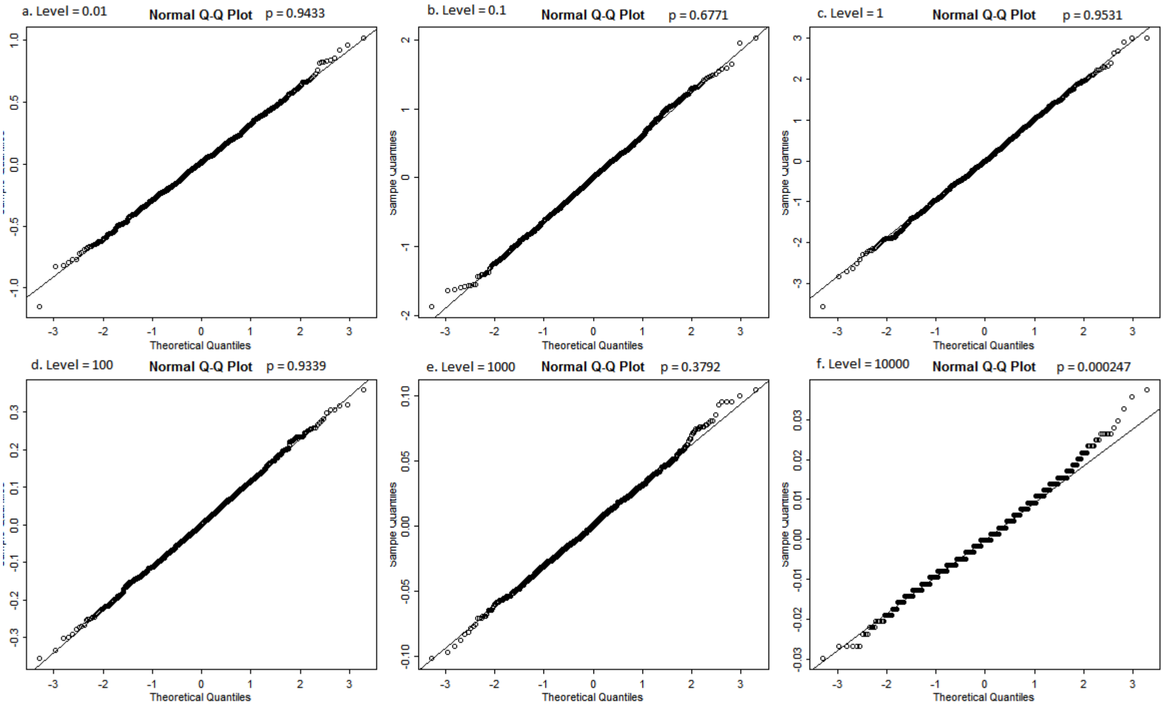

Figure 2: Q-Q plots for short range dependent Fisher field simulations. P values are from the Shapiro-Wilk test for normality

The third and fourth theorems apply to the long range dependent case with varying levels.

The asymptotic distribution of the excursion set for a Student field with long range dependence is normal.

Theorem 3 *Let* $\eta(x) = [\eta_1(x), \ldots, \eta_{n+1}(x)]'$, $x \in \mathbb{R}^d$, *satisfy Assumption 1 for* $\alpha \in (0, d)$, *and Assumption 2 hold for the spectral density of each component* $\eta_j(\cdot)$. *(see [2] for details of these assumptions) If* $a(r) = o(r^{\gamma/2n})$, $\gamma \in (0, \min(\alpha, d-\alpha))$, $r \to \infty$, *then the random variable*

$$\sqrt{2\pi}\left(1 + a(r)^2/n\right)^{n/2} \frac{M_r\{T_n\} - |\Delta|\, r^d I_{\frac{n}{n+a^2(r)}}\left(\frac{n}{2}, \frac{1}{2}\right)}{r^{d-\alpha/2} L^{1/2}(r) \sqrt{c_2(d,\alpha)\, c_3(1,d,\alpha)}}$$

is asymptotically $\mathcal{N}(0,1)$.

The asymptotic distribution of the excursion set for a Fisher field with long range dependence is a linear combination of the Rosenblatt type.

Theorem 4 *Let $\eta(x) = [\eta_1(x), \ldots, \eta_{n+m}(x)]'$, $x \in \mathbb{R}^d$, satisfy Assumption 1 for $\alpha \in (0, d/2)$, and Assumption 2 hold for the spectral density of each component $\eta_j(\cdot)$. (see [2] for details of these assumptions) If $a(r) = o(r^{\gamma/n})$, $\gamma \in (0, \min(\alpha, d-\alpha))$, $r \to \infty$, then the distribution of the random variable*

$$\frac{M_r\{F_{m,n}\} - |\Delta|\, r^d \left(1 - I_{\frac{ma(r)}{n+ma(r)}}\left(\frac{m}{2}, \frac{n}{2}\right)\right)}{c_4(a(r), n, m)\, r^{d-\alpha} L(r)}$$

converges to the distribution of the random variable

$$\frac{X_{2,1} + \ldots + X_{2,m}}{m} - \frac{X_{2,m+1} + \ldots + X_{2,m+n}}{n},$$

where $X_{2,j}$, $j = 1, \ldots, m+n$, are defined in [2].

2 Simulations

2.1 Process

We simulated Fisher-Snecdor and Student random fields with short range and long range dependence. The simulation used R and in particular the *RandomFields* package [3]. The aim of the simulations was to demonstrate the limit behaviour of the sojourn measure and verify the results from [2] and secondly to show the changes in limiting behaviour at high levels.

The simulation process is as follows:

- Generate three Gaussian random fields with gauss or Cauchy covariance model
- Combine them to give a Fisher or Student random field
- Measure the area of the excursion set at various levels
- Normalise the area according to the theorems from [2]
- Repeat 1000 times to simulate the distribution.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

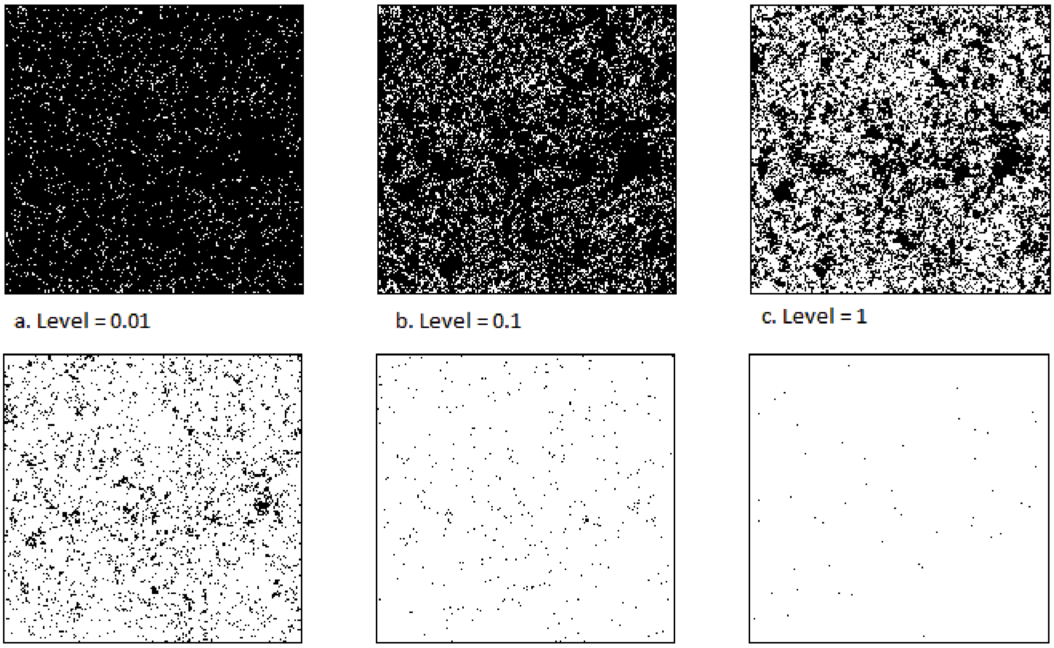

Figure 3: Excursion sets of a long range dependent Fisher field

2.2 Results

Figure 1 shows some excursion sets with various levels of a Fisher random field with a Gaussian covariance model. Excursion sets are shown in black throughout. The Gaussian covariance model has short range dependence so there are only small clusters.

For a short range dependent Fisher field we expect to see a normal distribution of the sojourn measure (theorem 2). The normal quantile-quantile plots in Figure 2 and the Shapiro-Wilk tests confirm that the distribution is normal at low levels and non-normal at high levels. The level of the graph (c) in Figure 2 is 1 which is in the middle for this field. If we increase the level to 10000 as in graph (f) the distribution is no longer normal (Shapiro-Wilk test gives p = 0.000247.) This demonstrates that at sufficiently high levels the distribution from [2] does not apply.

In contrast, the Fisher random field in Figure 3 has a Cauchy covariance model which has long range dependence so there is clustering over large distances.

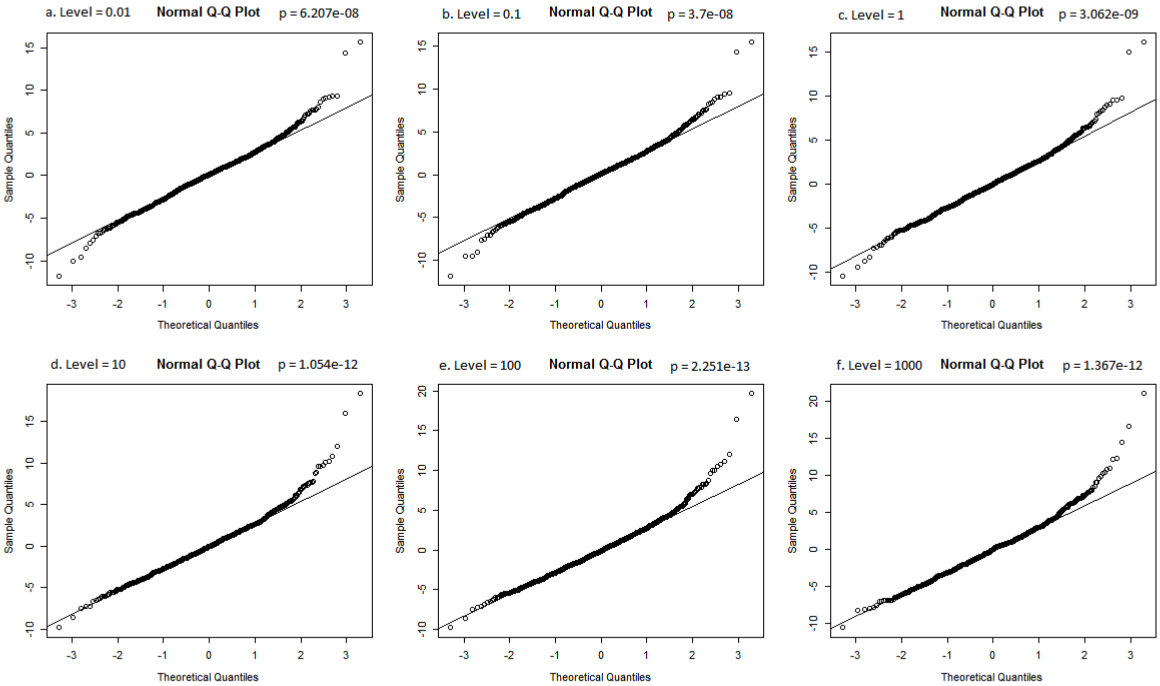

Figure 4: Q-Q plots for long range dependent Fisher field simulations. P values are from the Shapiro-Wilk test for normality

In the case of a Fisher field with long range dependence the limiting distribution is non-normal (see theorem 4). The Q-Q plots in Figure 4 are clearly non-normal and in addition it appears that the distribution is different between high and low levels. A Kolmogorov-Smirnov test confirms this with p=0.09710 when comparing the data with level 1 to those at level 100, for example. This is because the limit distribution no longer applies when the level is too high.

Figures 5 and 7 show excursion sets of Student random fields with short and long range dependence, respectively. Notice they look quite similar to the Fisher equivalents. One of the important differences between Student and Fisher fields is that Student fields take both positive and negative values while Fisher fields take only positive values. This is useful in many applications. The values in my simulations of Student fields tended to be lower than my Fisher simulations so the levels of interest are also lower.

For a Student field with short range dependence the theorem predicts a normal distribution however the simulation showed normal distributions only for levels less

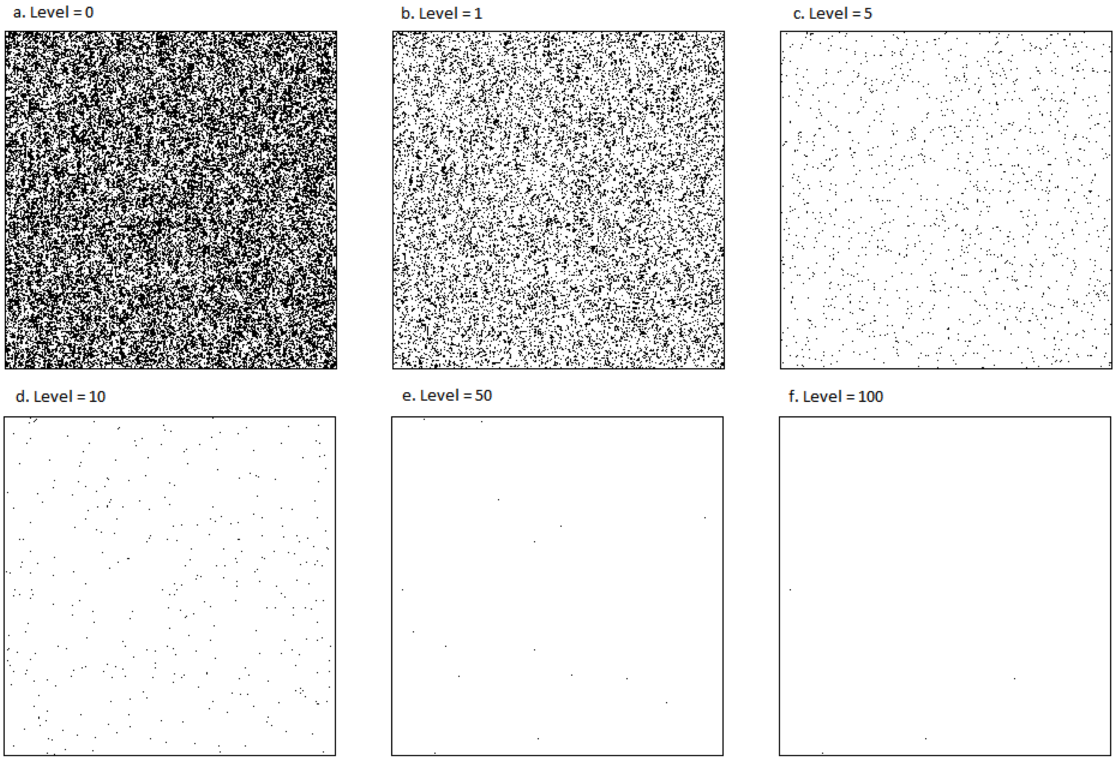

Figure 5: Excursion sets for short range dependent Student field

than 10 and non-normal distributions for levels as low as 50 (see Figure 6). A possible explanation for this is that the restriction on the growth of $a(r)$ is stricter in the case of a Student field ($a(r) = o(r^{\gamma/2n})$) compared to a long range dependent Fisher field ($a(r) = o(r^{\gamma/n})$).

Theorem 3 predicts the long range dependent Student field to give normal distributions at low levels. Figure 8 shows that even at level 0 the result is barely normal (p = 0.03) and all other levels tested the distribution is non-normal. A possible explanation is that the window size was too small too show limiting behaviour. I increased the window size of my simulation in an attempt to correct this problem but the results were similar. More time and computing power would allow me to check this option. The condition that $a(r) = o(r^{\gamma/2n})$ is stricter in the Student case than in the Fisher case which requires $a(r) = o(r^{\gamma/n})$. This may also account for the lack of normality. Further investigation is required.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

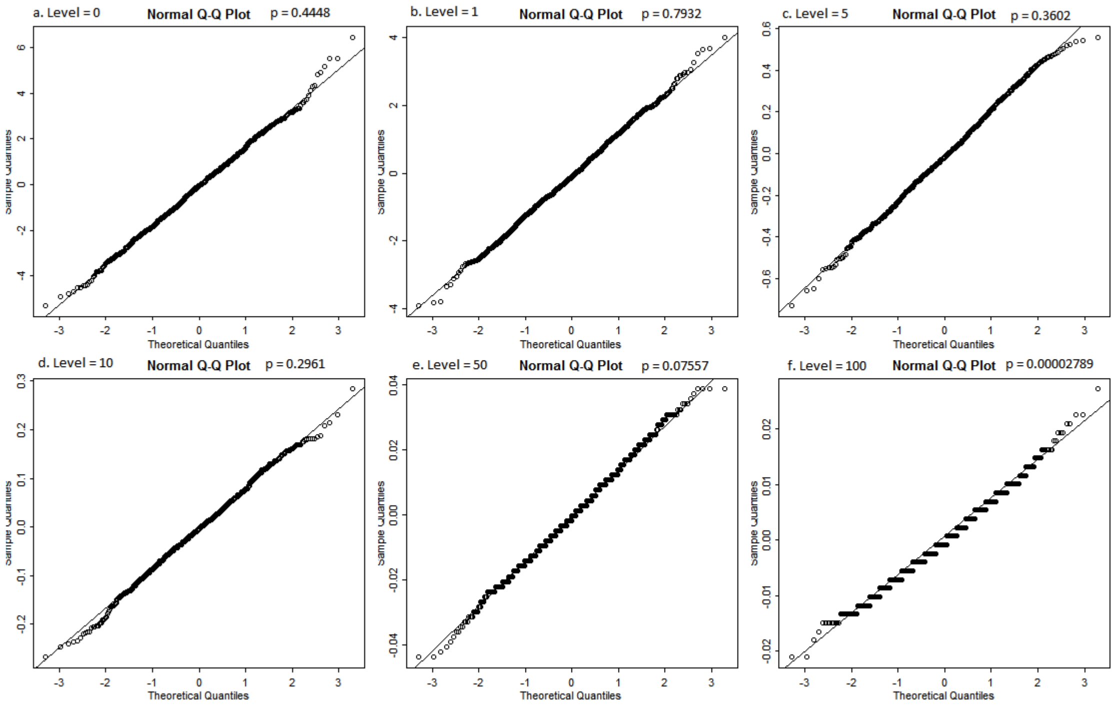

Figure 6: Q-Q plots for short range dependent Student field simulations. P values are from the Shapiro-Wilk test for normality

3 Conclusion

This project investigated the asymptotic distribution of the volume of the excursion set for heavy-tailed random fields. I learned about the theory of random fields. As preparation for my honours year I have gained a solid basis of knowledge of the subject material. I feel well prepared for further research on the topic.

I simulated the distribution of the excursion set at various levels. My results confirm the predictions from [2] at low levels but demonstrate that the distributions cannot be used at higher levels. Further work will include describing mathematically the conditions on the level at which the distributions cannot be applied. Conditions on the window size required for the limiting distribution to apply can also be found.

It would also be interesting to extend the work to other types of random fields for example χ^2 fields can be constructed from Gaussian fields. Another possibility is to

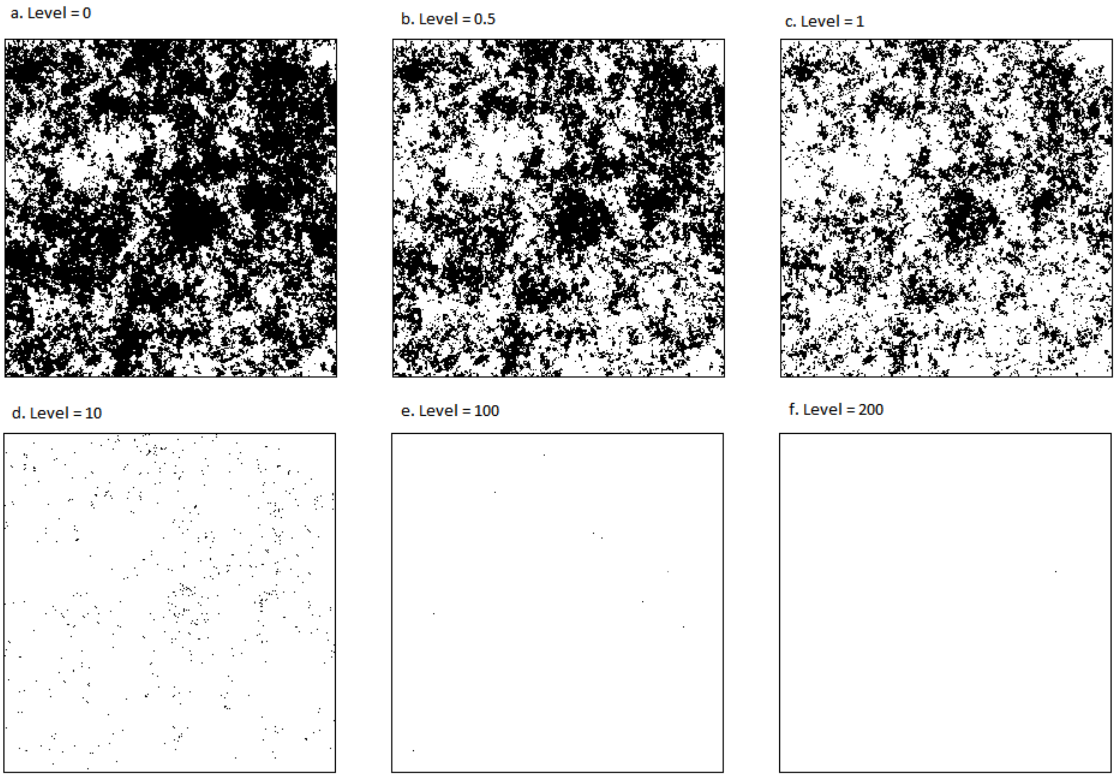

Figure 7: Excursion sets for long range dependent Student field

find asymptotic results for other Minkowski functionals.

Thank you to Andriy Olenko for being such a patient supervisor. I would also like to thank AMSI and CSIRO for their generous support and for hosting the Big Day In. The opportunity to speak in front of my peers is not common as an undergraduate.

References

[1] Adler, R. J., Taylor, J. E. and Worsley, K.J., (2009) *Applications of Random Fields and Geometry*, http://webee.technion.ac.il/people/adler/hrf.pdf

[2] N. Leonenko, A.Olenko. *Sojourn measures of Student and Fisher-Snedecor random fields*, will appear in Bernoulli.

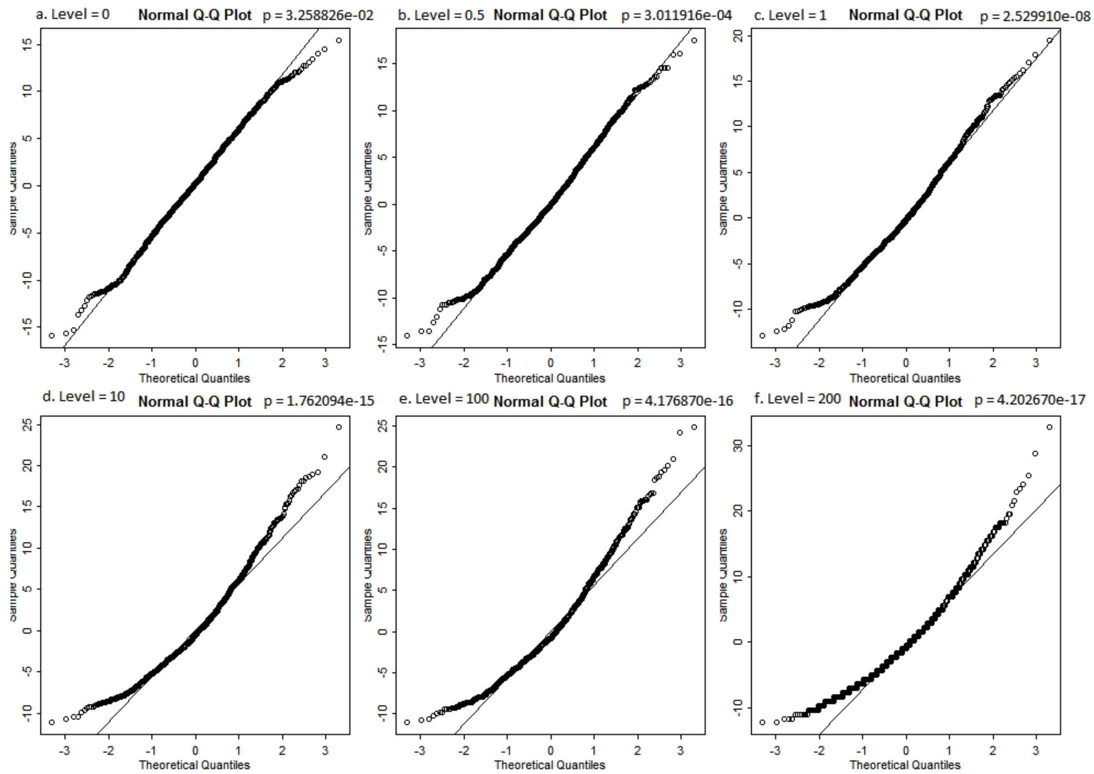

Figure 8: Q-Q plots for long range dependent Student field simulations. P values are from the Shapiro-Wilk test for normality

[3] Schlather, M., (1999), /em An introduction to positive definite functions and to unconditional simulation of random fields. Technical report ST 99-10, Dept. of Maths and Statistics, Lancaster University.

[4] Worsley, K. J. (1994) /em Local maxima and the expected Euler characteristic of excursion sets of χ^2 , F and t Fields. Adv. In Appl. Probab., 26 (1), 13-42.