Thesis for the degree of Doctor of Philosophy

Noise sensitivity and FK-type representations for Gaussian and stable processes

Malin Palö Forsström

CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

Division of Analysis and Probability
Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
Gothenburg, Sweden, 2019.

Noise sensitivity and FK-type representations for Gaussian and stable processes

Malin Palö Forsström

ISBN 978-91-7597-855-0

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4536
ISSN 0346-718X

Department of Mathematical Sciences
Chalmers University of Techology and University of Gothenburg
SE 412 96 Gothenburg
Sweden
Phone: +46 (0)31 772 10 00
Author email: palo@chalmers.se

Typeset with LaTeX
Printed by Chalmers Reproservice, Sweden, 2019.

Noise sensitivity and FK-type representations for Gaussian and stable processes

Malin Palö Forsström

Department of Mathematical Sciences,
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis contains four papers on probability theory.

Paper A concerns the question of whether the exclusion sensitivity and exclusion stability of a sequence of Boolean functions are monotone with respect to adding edges to the underlying sequence of graphs.

In paper B, we use the tools developed in Paper A to give an elementary proof of the behaviour of the mixing time of a random interchange process on a complete graph.

In Paper C we discuss the relationship between the noise sensitivity, noise stability and volatility of sequences of Boolean functions. In particular, we show that the set of volatile such sequences is dense in the set of all sequences of Boolean functions. Moreover, we construct a noise stable and volatile sequence of Boolean functions which is not $o(1)$-close to any non-volatile sequence of Boolean functions.

Finally, in Paper D, we investigate which threshold Gaussian and threshold stable random vectors have color representations. We discuss this from many different perspectives, and results include formulae for the dimension of the kernel of the associated linear operator, geometric conditions on the Gaussian vectors whose threshold have color representations and explicit examples of stable vectors with phase transitions at any stability index for the corresponding threshold process to have a color representation for large h.

Key words: Noise sensitivity, noise stability, volatility, Bernoulli random vector, color process, color representation, mixing time, exclusion process, interchange process, threshold Gaussian vector, threshold stable vector, multivariate stable distribution.

LIST OF APPENDED PAPERS

A **Malin Palö Forsström**.
Monotonicity properties of exclusion sensitivity.
Electronic Journal of Probability, Vol. 21, paper no. 45, 2016.

B **Malin Palö Forsström** and Johan Jonasson.
The spectrum and convergence rates of exclusion and interchange processes of the complete graph.
Journal of Theoretical Probability, Vol. 30, Issue 2, pp. 639–654, 2017.

C **Malin Palö Forsström**.
Denseness of volatile and nonvolatile sequences of functions.
Stochastic Processes and their Applications, Vol. 128, Issue 11, pp. 3880–3896, 2018.

D **Malin Palö Forsström** and Jeffrey E. Steif.
Fortuin-Kastelyn representations for threshold Gaussian and stable vectors.
Submitted in January 2019.

Papers not included in this thesis

E **Malin Palö Forsström**.
Exact Haussdorff measures of Cantor sets.
Real Analysis Exchange, Vol. 39, No. 2, pp. 367-384, 2014.

F **Malin Palö Forsström**.
Noise sensitivity and noise stability for Markov chains.
Preprint, included in licenciate thesis.

G **Malin Palö Forsström**.
A noise sensitivity theorem for Schreier graphs.
Preprint, included in licenciate thesis.

My contribution to the appended papers

B I significantly contributed to the development of ideas and theory, and also did some of the writing.

D I significantly contributed to the development of ideas and theory, as well as to the writing.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Jeffrey E. Steif for sharing his love of mathematics, for constantly reminding me that there are always new questions and for always encouraging me to stop to think things through carefully. In particular, thank you for all the discussions about mathematics we had during our last project. Next, I would like to thank my co-supervisor Johan Jonasson for supervising me during one year of my doctural studies. Also, thank you Olle Häggström for agreeing to be my co-supervisor, and for giving several excellent courses during my years at the department.

One of the most enjoyable things for me during the last six years has been meeting and getting to know my fellow Ph.D. students. In particular, I would like to thank Claes – for discussing your research with me and for letting me use your remarkable stapler, Magnus and Mariana – for board games, rides home and ice cream, Anna J., Anna H., Erik, Fanny, Ivar, Mikael, Olle, Tobias and Viktor – for all lunches and great company, Anders M. – for sharing your great enthusiasm for intriguing problems, Peter, Dawan, Hossein and Oscar – for the book club and for making sure there were always coffee breaks, and Lotta, Marie, and Fia – for being so welcoming and for always helping me more than you had to.

There are also a number of people outside the department that has made, and continues to make, a significant impact on me. In particular, thank you, Harry, for enabling me to continue learning mathematics, and for making me perceive mathematics as something exciting, interesting and completely natural to think about. Thank you, Sara, for reminding me that my view of mathematics was perhaps not utterly normal while always making me fully convinced that was not at all important. Also, thank you for playing Bartok with me for hours even when both I and my violin have been slightly out of tune and in less good shape than we should be. I am also grateful to Ann and Jan for always being extremely welcoming, and providing an environment where being different was not perceived as something negative, but rather precisely what made someone interesting.

Finally, I would like to thank my family. In particular, thank you Étienne for everything you are to me, and Hedvig for always reminding me of what is most important.

Malin Palö Forsström
Gothenburg, February 2019

CONTENTS

INTRODUCTION

The questions and answers discussed in the papers appended to this thesis have occupied a large proportion of my thoughts during the last five years. The purpose of this first part of the thesis is to describe the language and the setting in which these questions were asked. To this end, in Chapter 2 we give an introduction to Markov chains on finite state spaces. In Chapter 3 we use the definitions from Chapter 2 to introduce concepts such as noise stability, noise sensitivity, and volatility which will be central in several of the appended papers. Next, in Chapter 4, we define and briefly discuss what a color representation of a binary random vector is, and give several examples. After this, in Chapter 5 we introduce the family of so-called stable distributions in both a univariate and multivariate setting, and compare some of their properties to those of Gaussian distributions. Finally, in Chapter 6 we give a summary of the appended papers.

In this chapter, we will introduce Markov chains on finite state spaces in both discrete and continuous time. Markov chains are central to this thesis since they constitute the main component in the definitions of mixing time, noise sensitivity, noise stability, and volatility, which are central in the appended papers. Common for all these definitions is that the Markov chains are used to model noise in the underlying model. One reason that Markov chains are a reasonable way to model noise is that they are *memoryless*, in the sense that if you know the current state of the process, then further knowledge of earlier behavior of the process gives no additional information about the probabilities of future events. This behavior is natural to assume for noise since noise is often thought of as randomly affecting the process by chance or accident when it is at some state, as opposed to by an intelligent agent that observes a system over time.

The first examples of Markov chains in discrete time, were introduced by Andrei Andreevich Markov in 1906. In 1931, Kolmogorov introduced Markov chains in continuous time as we know them today ([12, 18]).

2.1 Discrete time Markov chains

In this section, we define and discuss discrete time Markov chains. Two good references on this topic are [2] and [12].

We now give a definition. To this end, let Ω be a finite set. A sequence of Ω-valued random variables $X_0, X_1, X_2, \dots,$ is said to be a *Markov chain* with *state space* Ω if for all $n \geq 1$ and $y, x_1, x_2, \dots, x_n \in \Omega$ we have that

$$\begin{aligned} &P(X_{n+1} = y \mid X_1 = x_1, \dots, X_n = x_n) \\ &\quad = P(X_{n+1} = y \mid X_n = x_n). \end{aligned}$$

Further, we will always assume that $(X_i)_{i \geq 0}$ is *time homogenous*, meaning that for all $x, y \in \Omega$ and $n \geq 0$, we have that

$$P(X_{n+1} = y \mid X_n = x) = P(X_1 = y \mid X_0 = x).$$

To simplify notation, for $x, y \in \Omega$ we write

$$P(x, y) \coloneqq P(X_1 = y \mid X_0 = x)$$

and call this the *transition probability* from x to y. The matrix $P = (P(x, y))_{x,y \in \Omega}$ is called the *transition matrix* of X. One consequence of this definition is that for

any $m \geq 0$, we have that

$$P(X_{m+1} = x_{m+1} \mid X_m = x_m) = e_{x_m}^T P(\cdot, x_{m+1})$$

and hence, by induction, it follows that for any $n \geq m \geq 0$ we have that

$$P(X_n = x_n \mid X_m = x_m) = e_{x_m}^T P^{n-m}(\cdot, x_n).$$

Also, it follows that if we let π_0 be a vector of length $|\Omega|$ with $\pi_0(x) = P(X_0 = x)$ for all $x \in \Omega$, then for any $n \geq 0$ we have that

$$P(X_n = x_n) = \pi_0^T P^n(\cdot, x_n)$$

and hence the vector $\pi_0^T P^n$ describes the distribution at time n given that the Markov chain has distribution π_0 at time zero.

If there is a vector π such that

$$\forall x, y \in \Omega \colon \pi(x)P(x, y) = \pi(y)$$

then $(\pi(x))_{x \in \Omega}$ is said to be a *stationary distribution* of X.

We often think of the index set $\{0, 1, 2, \ldots\}$ as describing time, and say that the Markov chain jumps from X_0 to X_1 at time one, from X_1 to X_2 at time two etc.

Three important properties which a discrete time Markov chain can have is irreducibility, aperiodicity and reversibility. A Markov chain X is said to be *irreducible* if

$$\forall x, y \in \Omega,\ \exists n \colon P(X_n = y \mid X_0 = x) > 0$$

aperiodic if

$$\forall x \in \Omega \colon \ \gcd\{n \geq 1 \colon P(X^n = x \mid X_0 = x) > 0\} = 1,$$

and *reversible* if

$$\forall x, y \in \Omega \colon \pi(x)P(X_1 = y \mid X_0 = x) = \pi(y)P(X_1 = x \mid X_0 = y).$$

Heuristically, aperiodicity means that the Markov chain is not periodic, irreducibility means that for any current state of the Markov chain, any other state can be attained at some future time, and reversibility means that for any $x_1, x_2 \in \Omega$, the events $(X_1, X_2) = (x_1, x_2)$ and $(X_1, X_2) = (x_2, x_1)$ are equally likely. One can show that any Markov chain has at least one stationary distribution, which is unique if X is irreducible (see e.g. Proposition 1.14 on page 12 and Corollary 1.18 on page 14 of [12]).

2.2 Continuous time Markov chains

A continuous time Markov chain differs from a discrete time Markov chain in that the time between two consecutive jumps is not fixed, but instead a random variable whose law is an exponential distribution. The following definition makes this precise. For more background and theory on the contents in this section, we refer the reader to [13].

Definition 2.2.1. Let Ω be a finite set. A right-continuous Ω-valued stochastic process $(X_t)_{t\geq 0}$ is a *continuous time Markov chain* if

(i) $(X_t)_{t\geq 0}$ changes its value at most finitely many times in any bounded interval, and

(ii) for all $x_1, x_2, \ldots, x_{n-1}, y \in \Omega$ and $0 \leq t_1 < t_2 < \ldots < t_n$,

$$\begin{aligned} &P(X_{t_n} = y \mid X_{t_1} = x_1, \ldots, X_{t_{n-1}} = x_{n-1}) \\ &\quad = P(X_{t_n} = y \mid X_{t_{n-1}} = x_{n-1}). \end{aligned}$$

The continuous time Markov chains we consider will always be *homogenous* in time, in the sense that we will always assume that for all $0 \leq t_1 < t_2$ and $x, y \in \Omega$

$$P(X_{t_2} = y \mid X_{t_1} = x) = P(X_{t_2 - t_1} = y \mid X_0 = x).$$

This motivates the definition of P_t; for $x, y \in \Omega$ and $t \geq 0$, we define $P_t(x, y) := P(X_t = y \mid X_0 = x)$ is continuous in t. Since Ω is finite, the limits

$$\begin{cases} q_{xy} := \lim_{t\to 0} P_t(x,y)/t \\ q_{xx} := \lim_{t\to 0}(P_t(x,x) - 1)/t \end{cases}$$

exist, and for $x, y \in \Omega$ and $t > 0$, we have that

$$P(X_{\inf\{s>t\colon X_s \neq x\}} = y \mid X_t = x) = \frac{q_{xy}}{\sum_{z\in\Omega\setminus\{x\}} q_{xz}}.$$

Moreover, given the assumptions above, then for all $x \in \Omega$ and $t \geq 0$ we have that

$$\inf\{s > t\colon X_s \neq x \mid X_t = x\} - t \sim \exp\left(-\sum_{y\in\Omega\setminus\{x\}} q_{xy}\right).$$

In other words, if the Markov chain is at $x \in \Omega$ at time t, then the time until it jumps to another state follows an exponential distribution with rate $\sum_{y\in\Omega\setminus\{x\}} q_{xy}$. For this reason, q_{xy} is called the *transition rate* from x to y, and the matrix $Q :=$

$(q_{xy})_{x,y\in\Omega}$, called the *generator* of $(X_t)_{t\geq 0}$, is a continuous analogue to the transition matrix of a discrete time Markov chain. The ratios $q_{xy}/\sum_{z\in\Omega\setminus\{x\}} q_{xz}$ are called the *jump probabilities* of the Markov chain.

If π is a probability distribution on Ω which is such that for all $x \in \Omega$ and all $t > 0$,

$$\pi(x) = \sum_{y\in\Omega} \pi(y)P_t(y,x)$$

then we say that π is a *stationary distribution* of $(X_t)_{t\geq 0}$.

A continuous time Markov chain X is said to be *reversible* if for all $x, y \in \Omega$ we have that

$$\pi(x)q_{xy} = \pi(y)q_{yx}$$

and *irreducible* if for all $t > 0$ and $x, y \in \Omega$ we have that

$$P(X_t = y \mid X_0 = x) > 0.$$

As in the discrete case, if $(X_t)_{t\geq 0}$ is irreducible then there will always exist a unique stationary distribution. If not otherwise stated, we will choose X_0 according to this distribution.

A rich source of examples of continuous time Markov chains is so called *continuous time random walks*.

Definition 2.2.2. Let G be a graph with vertex set $V(G)$ and edge set $E(G)$. We say that a continuous time Markov chain with state space $S = V(G)$ is a *continuous time random walk* on G if there is $\alpha > 0$ such that for any two distinct vertices $x, y \in V(G)$, we have

$$q_{xy} = \begin{cases} 0 & \text{if } (x,y) \notin E(G) \\ \alpha & \text{if } (x,y) \in E(G). \end{cases}$$

The constant α is called the *rate* of the random walk.

The following example of a continuous time random walk will be central in several of the appended papers.

Example 2.2.3. Let $n \geq 1$ and let G_n be the graph with $V(G_n) = \{-1,1\}^n$ which is such that for distinct $x, y \in \Omega$ we have $(x,y) \in E(G_n)$ if and only if $\sum_{i=1}^n |x(i) - y(i)| = 2$. Then G_n is said to be an *n-dimensional hypercube*. The continuous time random walk $(X_t)_{t\geq 0}$ with rate one on G_n can also be described as follows. Let $X_0 \in \{-1,1\}^n$. Then for any $\varepsilon > 0$, consider the binary string obtained by rerandomizing each bit $X_0(i)$, $i \in [n]$, using a uniform distribution on $\{-1,1\}$, with probability $1 - e^{-\varepsilon}$. Then the distribution of this binary string is exactly $\mathcal{L}(X_\varepsilon)$.

We end this section by using the definition of a continuous time random walk to define what we mean by an exclusion process.

Definition 2.2.4. Let G be a connected graph with vertex set $V(G)$ and edge set $E(G)$. Define another graph G' with vertex set $V(G') = \{-1, 1\}^{V(G)}$ and edge set $E(G')$ such that the following hold. For $x, y \in V(G')$, let $(x, y) \in E(G')$ if and only if there are distinct $v_1, v_2 \in V(G)$ such that

1. $x(v) = y(v)$ for all $v \in V(G) \backslash \{v_1, v_2\}$,
2. $(v_1, v_2) \in E(G)$, and
3. $x(v_1) = y(v_2)$ and $y(v_1) = x(v_2)$.

We say that a continuous time Markov chain X is an *exclusion process* on G if it is a continuous time random walk on G'. In other words, and exclusion process on G is a continuous time Markov chain on $\{0, 1\}^{V(G)}$ which at the event times of a Poisson process picks an edge in the graph at random and interchanges the $\{0, 1\}$-valued labels at its endpoints. The rate of the continuous time random walk, or equivalently of the Poisson process, is also said to be the *rate* of the resulting exclusion process.

2.3 Measuring speed of convergence – mixing times and relaxation times

In this section, we will define the mixing time and the relaxation time of a Markov chain. An excellent introduction to this topic is given in Chapters 4 and 12 in [12]. As a motivation for this topic, we first mention the following theorem, which says that an irreducible continuous time Markov chain will converge to its stationary distribution if it is run for long enough.

Theorem 2.3.1 (The Convergence Theorem)**.** *Suppose that X is an irreducible continuous time Markov chain on a finite state space Ω with stationary distribution π. Then there are constants $C > 0$ and $\alpha \in (0, 1)$ such that for each $t > 0$,*

$$\max_{x \in \Omega} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| \leq C\alpha^t.$$

The convergence theorem above motivates the following definition. The *total variation distance* between two probability distribution μ and ν on a finite space Ω is defined by

$$\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

One can with some work (see e.g. Proposition 4.2 in [12]) show that

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

One motivation for introducing the mixing time is to try to be more precise about the speed of convergence in Theorem 2.3.1. To this end, we define a distance function $d(t)$ by

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}.$$

Using this function, we define

$$t_{mix}(\varepsilon) := \min\{t \colon d(t) \leq \varepsilon\}$$

and finally the mixing time of a Markov chain $(X_t)_{t \geq 0}$ by

$$t_{mix} := t_{mix}(1/4).$$

The constant $1/4$ above is completely arbitrary, and could be replaced by any number in $(0, 1/2)$. If fact, one can show that for any $\varepsilon \in (0, 1/2)$,

$$t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil t_{mix}.$$

We now give a few examples.

Example 2.3.2 (Random walk on a hypercube)**.** The mixing time of a continuous time rate one random walk on a n-dimensional hypercube (see Example 2.2.3) is of order $\log n$ (see Theorem 18.3 in [12]).

Example 2.3.3 (Random walk on an n-cycle)**.** Let G_n be the graph with $V(G_n) = \mathbb{Z}_n$ which is such that for distinct $x, y \in \mathbb{Z}_n$ we have that $(x, y) \in E(G_n)$ if and only if $x - y = \pm 1$ in $\mathbb{Z}_n$. Let $X^{(n)}$ be a continuous time rate one random walk on G_n. Then the mixing time of $X^{(n)}$ is of order n^2 (see e.g. Subsubsection 5.3.1 in [12]).

Another measure of how long we need to run a Markov chain until it is well mixed is the so-called relaxation time of a Markov chain. While the mixing time is the time required on average for the worst possible (binary) function and the worst possible starting position, the relaxation time is instead the time required for the worst possible function if we choose a starting point according to the stationary distribution of the Markov chain. We now do this formally. To this end, let $(X_t)_{t \geq 0}$ be a reversible and irreducible continuous time Markov chain on Ω and let Q be its

generator. The *spectral gap* of X is defined as the second smallest eigenvector of $-Q$.

$$\lambda_2 := \inf_{\substack{f\colon \Omega\to\mathbb{R},\, f\not\equiv 0,\\ \mathbb{E}[f(X_0)]=0}} \frac{\langle f, -Qf\rangle}{\langle f, f\rangle}. \tag{2.1}$$

The quotient in (2.1) is called the *Rayleigh quotient* of $-Q$. As

$$\begin{aligned}\langle f, -Qf\rangle &= \sum_{x\in\Omega}\pi(x)f(x)\sum_{y\in\Omega}-q_{xy}f(y)\\ &= \sum_{x\in\Omega}\pi(x)f(x)\sum_{y\in\Omega\setminus\{x\}}q_{xy}(f(x)-f(y))\\ &= \sum_{x\in\Omega}\pi(x)\sum_{y\in\Omega\setminus\{x\}}q_{xy}f(x)(f(x)-f(y))\\ &= \frac{1}{2}\sum_{x\in\Omega}\pi(x)\sum_{y\in\Omega\setminus\{x\}}q_{xy}(f(x)-f(y))^2\end{aligned} \tag{2.2}$$

we can directly deduce that $\lambda_2 \geq 0$. Next, if $f\colon \Omega\to\mathbb{R}$ satisfies $\mathbb{E}[f(X_0)] = 0$ and $f \not\equiv 0$, then there must be at least two distinct $x, y \in \Omega$ such that $f(x) \neq f(y)$. Since (X_t) is irreducible, there must be a sequence of states $x_1, x_2, \ldots, x_n$ such that $q_{x,x_1} > 0$, $q_{x_1,x_2} > 0$, $q_{x_2,x_3} > 0$, $\ldots$, $q_{x_{n-1},x_n} > 0$ and $q_{x_n,y} > 0$, and hence it follows from (2.2) that in fact $\lambda_2 > 0$. This implies in particular that λ_2^{-1} is well defined. Using the eigenvalues and eigenvectors of $-Q$ and the fact that $P_t = e^{Qt}$ one can show that

$$\mathrm{Cov}(f(X_0), f(X_t)) \leq e^{-\lambda_2 t}\,\mathrm{Var}(f(X_0)).$$

This suggests that if one want to study the decorrelations of a function f of a continuous time Markov chain, it would be natural to pick $t = O(1/\lambda_2)$. For this reason, we define the *relaxation time* t_{rel} of X by $t_{rel} := \lambda_2^{-1}$.

One can show that (see e.g. Theorems 12.3 and 12.4 in [12]) for any reversible and irreducible Markov chain, we have that

$$-\log 2\varepsilon\cdot(t_{rel}-1) \leq t_{mix}(\varepsilon) \leq -\log\Big(\varepsilon\min_x \pi(x)\Big)t_{rel}. \tag{2.3}$$

This implies that the relaxation time can also be used to get upper and lower bounds for the mixing time of a Markov chain, which is one of the tools we use in Paper B to determine the mixing time of the so called interchange process.

We now give an example.

Example 2.3.4 (Random walk on a hypercube)**.** Let G_n be the n-dimensional hypercube (see Example 2.2.3) and let X be a continuous time rate one random walk on G_n. Then the relaxation time of X is equal to $1/2$ (see e.g. Example 12.15

in [12]). Interestingly, this is of a strictly smaller order than the mixing time, which is of order $\log n$ (see Example 2.3.2). Comparing this with (2.3) and using that the unique stationary distribution π of X is given by $\pi(x) = 2^{-n}$ for all $x \in V(G_n)$, we see that in this case, the leftmost inequality in (2.3) gives the wrong order for the mixing time while the rightmost inequality in (2.3) gives the correct order for the mixing time.

As humans, we ask questions all the time, such as "Do a majority of all people like classical music?", "Are there any birds visible from my bedroom window?" or "Will I freeze today if I do not put my hat on?". Common for all these questions is that the answer we get when we ask the question might be different a week, day, hour or even second later. Also, the degree to which an answer we got in the past can help us predict the answer at a future time decreases over time. Naturally, the speed at which such facts change depends on both the questions we ask and on how fast the object about which we ask the question evolve. In this chapter, we will describe some different ways in which this type of questions can be asked, and have been asked, within a mathematical framework. To this end, we will look at sequences of Boolean functions defined on state spaces of Markov chains , and use the Markov chains as a source of noise to their input. The goal of this chapter is to introduce ways to describe how sensitive different sequences of functions, describing answers to questions, are to such noise.

3.1 Noise sensitivity and noise stability

The noise sensitivity and noise stability of a sequence were first defined in [4] as measures of how sensitive sequences of Boolean functions $f_n\colon \{-1,1\}^n \to \{-1,1\}$ were to rerandomizing the sign of a small proportion of the bits in their domains. If we let $(X_t^{(n)})_{t\geq 0}$ be a continuous time random walk on the n-dimensional hypercube with rate one, then running this Markov chain for a short time will have exactly this effect.

Definition 3.1.1. For each $n \geq 1$, let $(X_t^{(n)})_{t\geq 0}$ be a continuous time random walk on the n-dimensional hypercube with rate one. A sequence $(f_n)_{n\geq 1}$ of Boolean functions, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, is said to be *noise sensitive* (wrt. $((X_t^{(n)})_{t\geq 0})_{n\geq 1}$) if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{Cov}(f_n(X_0), f_n(X_\varepsilon)) = 0.$$

The noise sensitivity of a sequence of Boolean functions, as defined above, was first introduced by Benjamini, Kalai and Schramm in [4]. In the same paper, the authors also introduced the concept of noise stability, which captures a possible opposite behaviour.

Definition 3.1.2. For each $n \geq 1$, let $(X_t^{(n)})_{t\geq 0}$ be a continuous time random walk on the n-dimensional hypercube with rate one. A sequence $(f_n)_{n\geq 1}$ of Boolean

functions, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, is said to be *noise stable* (wrt. $((X_t^{(n)})_{t\geq 0})_{n\geq 1}$) if

$$\lim_{\varepsilon\to 0}\sup_n P(f_n(X_\varepsilon) \neq f_n(X_0)) = 0.$$

We now give a few examples.

Example 3.1.3 (The Dictator function)**.** For $n \geq 1$, define $\textsc{Dict}_n\colon \{-1,1\}^n \to \{-1,1\}$ by $\textsc{Dict}_n(x) := x(1)$. In other words, the function $\textsc{Dict}_n$ returns the value of the first bit of its input. Since we for any $n \geq 1$ and $\varepsilon > 0$ have

$$P(\textsc{Dict}_n(X_\varepsilon) \neq \textsc{Dict}_n(X_0)) = P(X_\varepsilon(1) \neq X_0(1)) = \frac{1-e^{-\varepsilon}}{2}$$

it follows that the sequence $(\textsc{Dict}_n)_{n\geq 1}$ is noise stable.

Example 3.1.4 (The Parity function)**.** For $n \geq 1$, define $\textsc{Parity}_n\colon \{-1,1\}^n \to \{-1,1\}$ by $\textsc{Parity}_n(x) := \prod_{i=1}^n x(i)$. In other words, $\textsc{Parity}_n$ returns the parity of its input. Since we for any $n \geq 1$ and $\varepsilon > 0$ have

$$\mathrm{Cov}(\textsc{Parity}_n(X_\varepsilon), \textsc{Parity}_n(X_0)) = e^{-\varepsilon n}$$

it follows that the sequence $(\textsc{Parity}_n)_{n\geq 1}$ is noise sensitive.

Example 3.1.5 (The Majority function)**.** For $n \geq 1$, define $\textsc{Maj}_n\colon \{-1,1\}^n \to \{-1,1\}$ by $\textsc{Maj}_n(x) := I(\sum_{i=1}^n x(i) \geq 0)$. In other words, $\textsc{Maj}_n$ returns the value held by the majority of the bits in its input. One can show (see e.g. Theorem 2.45 in [15]) that the sequence $(\textsc{Maj}_n)_{n\geq 1}$ is noise stable.

Example 3.1.6. There are several ways to construct sequences $(f_n)_{n\geq 1}$ of Boolean functions $f_n\colon \{-1,1\}^n \to \{-1,1\}$ that are neither noise stable nor noise sensitive:

1. Let $(f_n)_{n\geq 1}$, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, be a sequence of functions such that

   $$\lim_{n\to\infty} P(f_n(X_0) = 1) \in \{0,1\}.$$

   Then $(f_n)_{n\geq 1}$ is both noise stable and noise sensitive.

2. For $x \in \{-1,1\}^n$, define

   $$f_n := \begin{cases} \textsc{Maj}_n & \text{if } n \text{ is even} \\ \textsc{Parity}_n & \text{if } n \text{ is odd.} \end{cases}$$

   Since $(\textsc{Maj}_n)_{n\geq 1}$ is noise stable (and not noise sensitive) and $(\textsc{Parity}_n)_{n\geq 1}$ is noise sensitive (and not noise stable), it follows that $(f_n)_{n\geq 1}$ can be neither noise stable nor noise sensitive.

3. For $x \in \{-1,1\}^n$, define

$$f_n(x) := \begin{cases} \textsc{Maj}_n & \text{if } x(1) = 1 \\ \textsc{Parity}_n & \text{else.} \end{cases}$$

Then it is easy to verify that $(f_n)_{n\geq 1}$ is neither noise stable nor noise sensitive.

In [4], the authors exclusively studied noise corresponding to a continuous time random walk on a n-dimensional hypercube. However, since then, similar definitions for several other state spaces and sources of noise have been considered. One such generalization was given in [5], where the authors considered noise being provided by an exclusion process, and gave the following definitions.

Definition 3.1.7 (Definition 1.1 in [5])**.** Let $(G_n)_{n\geq 1}$ be a sequence of graphs, and let $(\alpha_n)_{n\geq 1}$ be a sequence of positive numbers satisfying $\alpha_n \leq 1/\max_{v\in V(G_n)} \deg(v)$ for all n. For each $n \geq 1$, let $(X_t^{(n)})_{t\geq 0}$ be a continuous time exclusion process with rate α_n on G_n, and let $X_0^{(n)}$ be chosen according to the uniform measure on $\{-1,1\}^{G_n}$. A sequence $(f_n)_{n\geq 1}$ of Boolean functions, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, is said to be *exclusion sensitive* (wrt. $(G_n)_{n\geq 1}$ and $(\alpha_n)_{n\geq 1}$) if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{Cov}(f_n(X_0), f_n(X_\varepsilon)) = 0.$$

Definition 3.1.8 (Definition 1.2 in [5])**.** Let $(G_n)_{n\geq 1}$ be a sequence of graphs, and let $(\alpha_n)_{n\geq 1}$ be a sequence of positive numbers satisfying $\alpha_n \leq 1/\max_{v\in V(G_n)} \deg(v)$ for all n. For each $n \geq 1$, let $(X_t^{(n)})_{t\geq 0}$ be a continuous time exclusion process with rate α_n on G_n, and let $X_0^{(n)}$ be chosen according to the uniform measure on $\{-1,1\}^{G_n}$. A sequence $(f_n)_{n\geq 1}$ of Boolean functions, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, is said to be *exclusion stable* (wrt. $(X_t^{(n)})$) if

$$\lim_{\varepsilon\to 0} \sup_n P(f_n(X_\varepsilon) \neq f_n(X_0)) = 0.$$

We now give an example.

Example 3.1.9 (The Parity function)**.** Let $\textsc{Parity}_n\colon \{-1,1\}^n \to \{-1,1\}$ be defined as in Example 3.1.4. Since $\textsc{Parity}_n(X_t^{(n)})$ depends on $X^{(n)}$ only through $\sum_{i=1}^n X^{(n)}(i)$, which is not affected by an exclusion process on any graph, it follows that the sequence $(\textsc{Parity}_n)_{n\geq 1}$ is not exclusion sensitive. In fact, it is exclusion stable. We can however quite easily define a similar function which is exclusion sensitive. To this end, define $\textsc{Parity'}_n\colon \{-1,1\}^n \to \{-1,1\}$ by

$$\textsc{Parity}'_n(x) := \textsc{Parity}_{\lfloor n/2\rfloor}(x_1, x_2, \ldots, x_{\lfloor n/2\rfloor}).$$

Then it is easy to verify that $(\textsc{Parity}'_n)$ is exclusion sensitive with respect to e.g. a sequence of complete graphs.

In [5], the authors show that any sequence $(f_n)_{n\geq 1}$ of Boolean functions which is exclusion sensitive with respect to some sequence of graphs is also noise sensitive, and conversely, that any sequence of Boolean functions which is noise stable is also exclusion stable with respect to any sequence of graphs. However, [5] does not include any results on how the properties of being exclusion sensitive with respect to different sequences of graphs relate. To obtain results in this direction was the primary goal of Paper A.

3.2 Volatility

Suppose that you are planning for a full day outside. Then it is probably more relevant to ask "*How likely is it that it starts raining at some point during the day?*" than "*How likely is it that it rains at lunch, given that it does not rain now?*" This observation motivates the definition of volatility.

The volatility of a sequence of Boolean functions was first introduced in [10], as a measure of the stability of a sequence of Boolean functions complementing the definitions of noise sensitivity and noise stability. One motivation for introducing a new definition was that both noise sensitive and noise stability, although giving information about $f_n(X_t)$ at two distinct times $t = 0$ and $t = \varepsilon$, gives no information about whether f_n changed its value at any intermediate time $t \in (0, \varepsilon)$.

Definition 3.2.1 (Definition 1.3 in [10])**.** For each $n \geq 1$, let $(X_t^{(n)})$ be a continuous time random walk on the n-dimensional hypercube with rate one. A sequence $(f_n)_{n\geq 1}$ of Boolean functions, $f_n\colon \{-1,1\}^n \to \{-1,1\}$, is said to be *volatile* if for all $\delta > 0$,

$$\lim_{n\to\infty} P(\tau_{\partial f_n} > \delta) = 0,$$

where $\tau_{\partial f_n}$ is the hitting time of the set $\{x\colon f_n(x) \neq f_n(X_0)\}$.

In words, a sequence of functions is volatile if for any small time interval, the probability that $f_n(X_t^{(n)})$ has changed its value at least once in that time interval tends to one as n tends to infinity. In [10] the authors show that any noise sensitive sequence of Boolean functions is also volatile. However, the converse is not true, and in Paper C we give an example which shows that there are sequences of Boolean functions which are both noise stable and volatile. Such examples can quite easily by constructing sequences of functions which, in the limit, changes their value infinitely often, but only for such short time periods that if you look at the value of

the function at any fixed set of times, it is extremely unlikely that you will see that the function ever changed its value at all.

In this chapter, we define and discuss the concept of color representations which is the main topic of Paper D. To this end, consider the following construction. Let S be an index set, let $\mathcal{P}(S)$ be the set of all partitions of S and let μ be a probability measure on $\mathcal{P}(S)$. Then $Y \sim \mu$ is a random partition of S. Given such a random partition Y, we obtain a $\{0,1\}^S$-valued process X by for each partition element of Y, independently for different partition elements, assigning all elements in the partition element the value one with probability p_1 and value zero with probability $p_0 := 1 - p_1$. We write $\Phi_{p_1}(\nu)$ to denote the law of X.

Definition 4.0.1 (Section 1.2 in [19])**.** Let ν be a probability measure on $\{0,1\}^S$ with marginal distributions (p_0, p_1). If there is a probability measure μ on $\mathcal{P}(S)$ such that $\Phi_{p_1}(\mu) = \nu$ we say that ν is a *color process* and that μ is a *color representation* of ν.

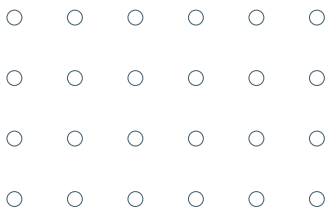
Even though color representations as described above were first defined in [19], the concept had arisen earlier. One of the most prominent examples of this is the coupling between the Ising model with $p = 1/2$ and the Fortuin-Kastelyn random cluster model (see Figure 4.2(d)). It is well known that the random cluster model is, in fact, a color representation of the Ising model (see, e.g. [8] and [19]). This coupling has been used extensively to better understand the Ising model. Another example is the so-called *divide and color model* defined in [9] (see Figure 4.2(b)). This random model is constructed by first performing independent bond percolation on some graph G, and then coloring the resulting connected components independently according to some probability measure on a set of colors C. If there are exactly two colors, then the resulting model has a color representation by definition, namely the measure of the connected components in bond percolation. In [9], Häggström studied the properties of the resulting measure on $C^{V(G)}$.

Given Definition 4.0.1, one might ask the following questions.

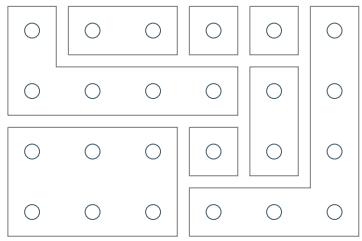
(i) Are there any natural conditions on ν that guarantee the existence or non-existence of a color representation?

(ii) If ν has a color representation, what does that tell us about ν?

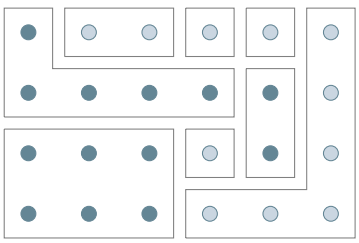
In [19], the authors primarily studied the second of these questions, while the focus of Paper D, and hence the discussion below, is the first. Note that by definition, there clearly are probability distributions which have color representations. The following example however, shows that this does not always hold, hence motivating the question of existance.



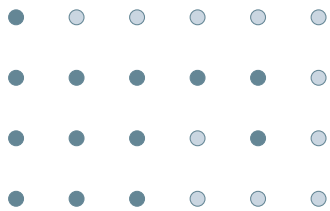
(a) A possible index set S.



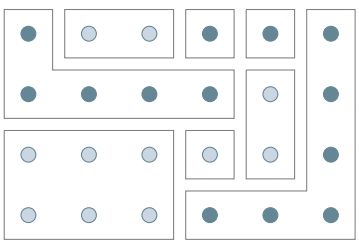
(b) A partition σ of S (i.e. an element $\sigma \in \mathcal{P}(S)$).



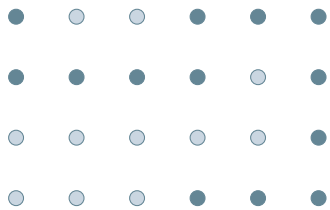
(c) A coloring of S which is compatible with the partition σ.



(d) The color process obtained from the coloring in (c).

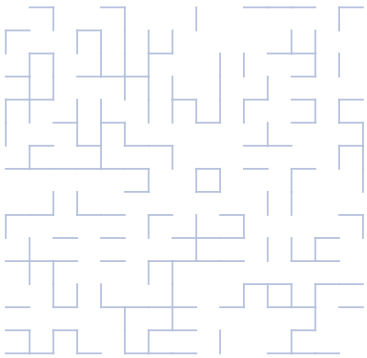


(e) A coloring of S which is compatible with the partition σ.

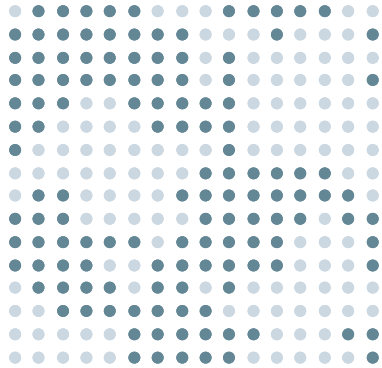


(f) The color process obtained from the coloring in (e).

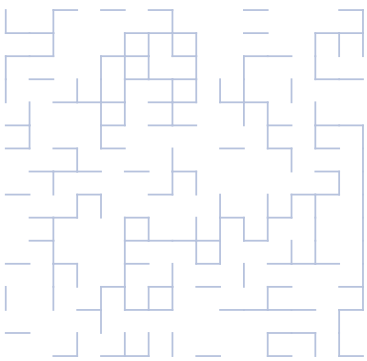
Figure 4.1: In Figures (a)-(f) above we give an example of a possible set S, a partition σ of S, two possible colorings of S which are compatible with σ, as well as the correspong color processes.



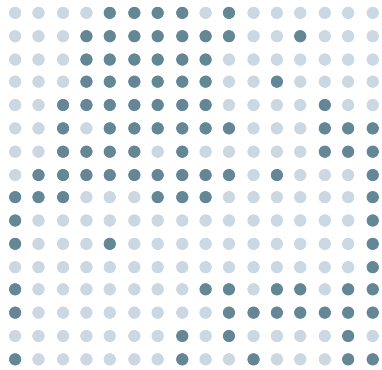
(a) A simulation of bond percolation on a subset of $\mathbb{Z}^2$ using edge probability $p = 1 - e^{-1/4}$. Bond perculation induces a partition of the same subset by letting two elements of $\mathbb{Z}^2$ be in the same partition elements exactly if there is a set of edges connecting them.



(b) An example of a divide and color model, using the partition in (a) and color distribution $(1/2, 1/2)$.



(c) An example of the Fortuin-Kastelyn random cluster model, using parameters $p = 1 - e^{-1/4}$ and $q = 2$. As in (a), the corresponding partition of $\mathbb{Z}^2$ is exactly the connected components of this graph.



(d) An example of the Ising model, with no external field and $\beta = 1/4$, on a subset of $\mathbb{Z}^2$, obtained as a color process using the partition in (c) using $p_1 = 1/2$.

Figure 4.2: In Figures (a)-(d) above we show simulations of bond percolation, the color and divide model, the Fortuin-Kastelyn random cluster mode and the Ising model.

Example 4.0.2. Let ν be the probability distribution on $\{0,1\}^2$ which gives mass $1/2$ to $(0,1)$ and mass $1/2$ to $(1,0)$. Then ν have marginal distributions $(2^{-1}, 2^{-1})$ and hence $p_1 = 2^{-1}$. We claim that ν has no color representation. To see this, simply note that if ν had had a color representation μ, then we would have that

$$\nu((0,0)) = 2^{-1}\mu((12)) + 2^{-2}\mu((1,2)) > 0$$

which is a contradiction. By generalizing this argument, we see that if for any index set S, having $\nu(0^S) \geq p_0^{|S|}$ is in fact a necessary condition for ν having a color representation. Using this fact, it follows that for any index set S there are many distributions ν on $\{0,1\}^S$ which do not have a color representation. We will obtain many more interesting examples of distributions which do not have color representations in Paper D.

Now fix some index set S and $\sigma \in \mathcal{P}(S)$. If $\rho \in \{0,1\}^S$ can be obtained by assigning the value zero or one to each partition element in σ we say that ρ and σ are *compatible* and write $\rho \lhd \sigma$. Next, for $\sigma \in \mathcal{P}(S)$, let $\|\sigma\|$ be the number of partition elements of σ, and if $\rho \lhd \sigma$, let $c(\sigma, \rho)$ be the number of partition elements of σ which will have to be assigned the value one to obtain the binary string ρ. Using this notation, we see that a probability measure $\nu = (\nu(\rho))_{\rho \in \{0,1\}^S}$ has a color representation if and only if there is a non-negative solution $\mu = (\mu(\sigma))_{\sigma \in \mathcal{P}(S)}$ to the system of linear equations

$$\nu(\rho) = \sum_{\sigma\,:\,\sigma \lhd \rho} p_1^{c(\sigma,\rho)} p_0^{\|\sigma\| - c(\sigma,\rho)} \mu(\sigma), \qquad \rho \in \{0,1\}^S. \tag{4.1}$$

Since these equations are all linear, one can apply methods from linear algebra to find a solution. However, the problem of knowing if there is a non-negative solution is not naturally solved by the use of linear algebra, and this is also one of the main difficulties in answering (ii).

To make the discussion above more concrete, we now give another example.

Example 4.0.3. Let $S = \{1,2,3\}$. Then there are exactly five partitions of S, namely (123), $(12,3)$, $(13,2)$, $(1,23)$ and $(1,2,3)$. Let $\nu = (\nu(\rho))_{\rho \in \{0,1\}^3}$ be a probability vector with marginals (p_0, p_1). To simplify notation slightly, for $\rho \in \{0,1\}^3$ and $\sigma \in \mathcal{P}(\{1,2,3\})$, we write $\nu_\rho := \nu(\rho)$ and $\mu_\sigma := \mu(\sigma)$. Then the

system of equations in (4.1) reads

$$\begin{cases} \nu_{000} = p_0\,\mu_{123} + p_0^2\,\mu_{12,3} + p_0^2\,\mu_{13,2} + p_0^2\,\mu_{1,23} + p_0^3\,\mu_{1,2,3} \\ \nu_{001} = p_0 p_1\,\mu_{12,3} + p_0^2 p_1\,\mu_{1,2,3} \\ \nu_{010} = p_0 p_1\,\mu_{13,2} + p_0^2 p_1\,\mu_{1,2,3} \\ \nu_{100} = p_0 p_1\,\mu_{1,23} + p_0^2 p_1\,\mu_{1,2,3} \\ \nu_{011} = p_0 p_1\,\mu_{1,23} + p_0 p_1^2\,\mu_{1,2,3} \\ \nu_{101} = p_0 p_1\,\mu_{13,2} + p_0 p_1^2\,\mu_{1,2,3} \\ \nu_{110} = p_0 p_1\,\mu_{12,3} + p_0 p_1^2\,\mu_{1,2,3} \\ \nu_{111} = p_1\,\mu_{123} + p_1^2\,\mu_{12,3} + p_1^2\,\mu_{13,2} + p_1^2\,\mu_{1,23} + p_1^3\,\mu_{1,2,3}. \end{cases}$$

One can quite easily verify that when $p_1 \notin \{0, 1/2, 1\}$ there is a unique (possible not non-negative) solution given by

$$\begin{cases} \mu_{1,2,3} &= \frac{\nu_{100}-\nu_{011}}{p_0 p_1 (p_0 - p_1)} \\ \mu_{12,3} &= \frac{p_0\nu_{110}-p_1\nu_{001}}{p_0 p_1 (p_0 - p_1)} \\ \mu_{13,2} &= \frac{p_0\nu_{101}-p_1\nu_{010}}{p_0 p_1 (p_0 - p_1)} \\ \mu_{1,23} &= \frac{p_0\nu_{011}-p_1\nu_{100}}{p_0 p_1 (p_0 - p_1)} \\ \mu_{123} &= 1 - \frac{p_1\nu_{000}-p_0\nu_{111}}{p_0 p_1 (p_0 - p_1)} \end{cases}$$

and that when $p = 1/2$ there are infinitely many (possible not non-negative) solutions given by

$$\begin{cases} \mu_{1,2,3} = 2 - 2t = 1 \\ \mu_{12,3} = 4\nu_{001} - 1 + t \\ \mu_{13,2} = 4\nu_{010} - 1 + t \\ \mu_{1,23} = 4\nu_{100} - 1 + t \\ \mu_{123} = 4\nu_{000} - t \end{cases}$$

where $t \in \mathbb{R}$ is a free variable. In [19], the authors show that there is a non-negative solution in this case if and only if ν has non-negative correlations. However, one can show that having non-negative correlations is not a sufficient condition on ν to have a color representation neither when $|S| \geq 4$ nor when $p \neq 1/2$.

One of the most important probability distributions is the *Gaussian distribution*. The Gaussian distribution is popular for two very good reasons. Firstly, a sum, and even a weighted sum, of two normally distributed random variables is again normally distributed, and secondly, because of the central limit theorem, stating in its simplest form that a sum of many small independent effects will often look like a Gaussian distribution. Each of these two properties, independently, when stated formally, completely characterizes the larger family of so called stable distributions.

The first to study the set of stable distributions as a concept was Lévy, in two monographs from 1925 and 1937 respectively. However, some of the probability distributions in this family had been studied much earlier. The Gaussian cumulative distribution function appeared as a function already in works by d'Moivre in 1733 as an approximation of certain binomial sums. It was first thought of as a probability distribution by Gauss in 1809. A first version of the central limit theorem appeared first in works by Laplace in 1810 (see e.g. [17]). Poisson, and later Cauchy, studied the distribution with density function $f_\lambda(x) = \frac{\lambda}{\pi(\lambda^2+x^2)}$ for $\lambda > 0$ and $x \in \mathbb{R}$, today known as the Cauchy distribution. Another early appearance of stable distributions was in the works by Holtsmark, a Danish astronomer. In 1919 Holtsmark observed patterns which followed a probability law with Fourier transform given by $\exp(-\lambda|t|^{3/2})$, $t \in \mathbb{R}^3$. This distribution is today known to be a spherically symmetric three-dimensional stable distribution with stability index $3/2$.

For a more thorough exposition of the history of the development of the theory of stable distributions, see e.g. [20]. In the remainder of this chapter, we will define the family of stable distributions and describe some of their properties.

5.1 One-dimensional stable distributions

We first consider stable distributions in a one-dimensional setting.

Definition 5.1.1 (Definition 1.1.1 on p. 2 in [16])**.** A real-valued random variable X is said to have a *stable distribution* if for any positive numbers A and B there is a positive number C and a real number D such that if X_1 and X_2 are two independent copies of X, then

$$AX_1 + BX_2 \stackrel{\mathcal{D}}{=} CX + D.$$

Since this is well known to hold for any random variable with a Gaussian distribution, it follows that the set of Gaussian distributions is a subset of the set of all stable distributions. One can show (see e.g. Theorem 1.1.2 on p. 3 in [16])

that for every stable random variable X there is a number $\alpha \in (0, 2]$ such that $C^\alpha = A^\alpha + B^\alpha$. The number α is called the *stability index* of X, and if X has stability index α we say that X is α-stable. If $\alpha = 2$, then X follows a Gaussian distribution.

An alternative way to characterize the family of stable distributions is the following, which can be said to generalize the central limit theorem.

Definition 5.1.2 (Definition 1.1.5 on p. 5 in [16])**.** If $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables, $(a_n)_{n\geq 1}$ and $(d_n)_{n\geq 1}$ are two sequences of real numbers and

$$\frac{X_1 + X_2 + \ldots + X_n}{d_n} + a_n$$

converges in distribution to a random variable X, then X is said to have a stable distribution.

Note in particular that if a sequence of normalized sums of i.i.d. real-valued random variables converges to another random variable in distribution, then the previous definition says that the limit is a stable random variable, and hence the collection of stable distributions is exactly the distributions that can arise as such limits. One can show that $d_n = n^{1/\alpha} h(n)$ for some slowly varying function $h(n)$ and some $\alpha \in (0, 2]$ which will be the stability index of the limiting stable distribution (see p. 575 in [7]).

A third way to define stable distributions is through their characteristic functions.

Definition 5.1.3 (Definition 1.1.6, p. 5 in [16])**.** A real-valued random variable X is said to have a *stable distribution* if there is $\alpha \in (0, 2]$, $\sigma \geq 0$, $\beta \in [-1, 1]$ and $\mu \in \mathbb{R}$ such that the characteristic function of X is given by

$$\mathbb{E}[\exp i\theta X] = \begin{cases} \exp\big(-\sigma^\alpha |\theta|^\alpha (1 - i\beta \operatorname{sgn}\theta \tan\frac{\pi\alpha}{2}) + i\mu\theta\big) & \text{if } \alpha \neq 1 \\ \exp\big(-\sigma^\alpha |\theta| (1 + i\beta \frac{2}{\pi} \operatorname{sgn}\theta \log|\theta|) + i\mu\theta\big) & \text{if } \alpha = 1. \end{cases}$$

If X, α, β, μ and σ are as in Definition 5.1.3, we write $X \sim S_\alpha(\sigma, \beta, \mu)$. Here α is the stability index, μ is a *location parameter*, σ is a *scale parameter* and β is a *symmetry parameter*. If $\beta = 0$, then the distribution of X is symmetric. Note that if $\alpha = 2$, then the parameter β plays no role.

The probability density functions of stable distributions exist and are continuous, but are not known in closed form except in a few very special cases described in the examples below (see e.g. page 9 in [16]).

Example 5.1.4. When $\alpha = 2$, the characteristic function simplifies to $\mathbb{E}[\exp i\theta X] = \exp\big(-\sigma^2\theta^2 + i\mu\theta\big)$. This is exacty the characteristic function of a normally distributed random variable with mean μ and variance $2\sigma^2$. Hence $S_2(\sigma, \beta, \mu) \overset{\mathcal{D}}{=} N(\mu, 2\sigma^2)$ and it follows that the corresponding probability density function is given by

$$x \mapsto \frac{1}{\sqrt{4\pi\sigma^2}} e^{-(x-\mu)^2/4\sigma^2}, \qquad x \in \mathbb{R}.$$

Example 5.1.5. $S_1(\sigma, 0, \mu)$ is the so called Cauchy distribution, with probability density function given by

$$x \mapsto \frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)}, \qquad x \in \mathbb{R}.$$

The Cauchy distribution arises naturally as the distribution of intersection between the x-axis and a line passing through (μ, σ) at a uniformly distributed angle. The Cauchy distribution with $\mu = 0$ and $\sigma = 1$ is also the distribution of the ratio of two independent standard Gaussian random variables.

Example 5.1.6. $S_{1/2}(\sigma, 1, \mu)$ is the so called Lévy distribution, whose density is given by

$$x \mapsto \left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\sigma}{2(x-\mu)}\right), \qquad x > \mu.$$

If $\mu = 0$ and $\sigma = \ell^2$, then the corresponding Lévy distibution describes the hitting time of level ℓ by a Brownian motion starting at the origin.

In Figures 5.1 and 5.2 we draw the probability density functions of some stable distributions as α and β varies.

When $\alpha = 2$ it is well known that the tails of the cumulative density function of an α-stable random variable X decays at an exponential rate. When $\alpha < 2$ however, the tails decay much slower, and in fact if $X \sim S_\alpha(\sigma, \beta, \mu)$, then

$$\lim_{h\to\infty} h^\alpha P(X > h) = C_\alpha \frac{1+\beta}{2} \sigma^\alpha \tag{5.1}$$

and

$$\lim_{h\to\infty} h^\alpha P(X < -h) = C_\alpha \frac{1-\beta}{2} \sigma^\alpha$$

where $C_\alpha^{-1} := \int_0^\infty x^{-\alpha} \sin x \, dx$ (see Property 1.2.15 on p. 16 in [16]). This in particular implies that if X is α-stable and $\alpha < 2$, then $\mathbb{E}[|X^p|] < \infty$ if and only if $p \in (0, \alpha)$ (see Property 1.2.16 on p. 18 in [16]). Hence the variance of a α-stable random variable is only defined if $\alpha = 2$. Similarly, the expected value is only defined if $\alpha > 1$, in which case it is equal to μ (see Property 1.2.19 on p. 19 in [16]).

We now continue to the multivariate case.

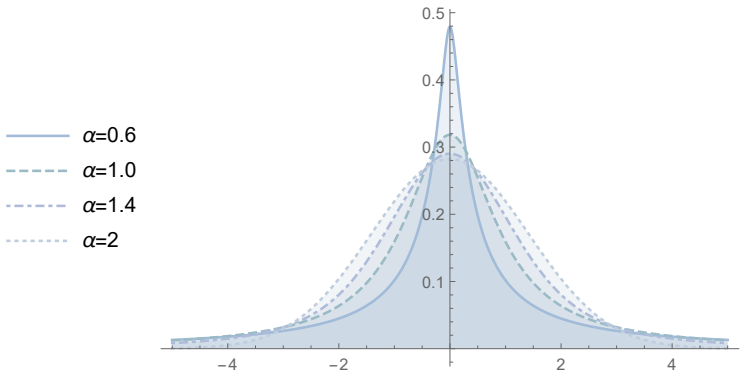


Figure 5.1: In the figure above we draw the probability density function of a stable distribution with $\mu = 0$, $\sigma = 1$ and $\beta = 0$ for four different values of α, including $\alpha = 2$ which corresponds to a Gaussian distribution with mean zero and variance two. Note in particular that the plot suggests that smaller α corresponds to fatter tails.

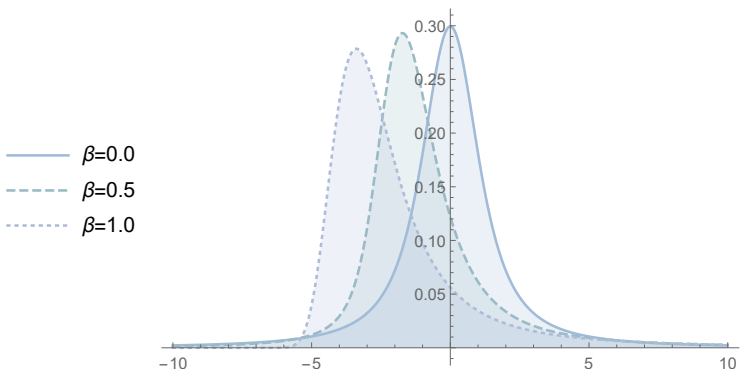


Figure 5.2: In the figure above we draw the probability density function of a stable distribution with $\mu = 0$, $\sigma = 1$ and $\alpha = 1.2$ for three different values of β, including $\beta = 0$ which corresponds to a symmetric distribution and $\beta = 1$ which corresponds to a so called *totally skewed distribution*. Interestingly, when $\alpha \geq 1$, a totally skewed stable distribution has full support and when $\alpha < 1$ it is supported only on $[\mu, \infty)$.

5.2 Multivariate stable distributions

The following generalizes Definition 5.1.1 to a multivariate setting.

Definition 5.2.1 (Definition 2.1.1 on p. 57 in [16])**.** A $\mathbb{R}^n$-valued random vector $X = (X_1, X_2, \ldots, X_n)$ is said to be a *stable random vector* if for any positive numbers A and B there is a positive number C and a vector $D \in \mathbb{R}^n$ such that if $X^{(1)}$ and $X^{(2)}$ are two independent copies of X, then

$$AX^{(1)} + BX^{(2)} \stackrel{\mathcal{D}}{=} CX + D.$$

One can show that there is a unique constant $\alpha \in (0, 2]$ such that $C^\alpha = A^\alpha + B^\alpha$, and this is called the *stability index* of X (see e.g. Theorem 2.1.2, p. 58 in [16]). If X has stability index α we say that X is an *α-stable random vector.*

The following theorem gives another characterization of an α-stable random vector.

Theorem 5.2.2 (Theorem 2.3.1 on p. 65 in [16])**.** *If $X = (X_1, \ldots, X_n)$ is an α-stable random vector then there exists a finite measure Λ on $\mathbb{S}^{n-1}$ and a vector $\mu \in \mathbb{R}^n$ such that*

$$\begin{aligned}&\mathbb{E}[\exp i\langle\theta, X\rangle] \\ &\quad = \exp\left(-\int_{\mathbb{S}^{n-1}} |\langle\theta, \mathbf{s}\rangle|^\alpha (1 + i\nu(\langle\theta, \mathbf{s}\rangle))\Lambda(d\mathbf{s}) + i\langle\theta, \mu\rangle\right),\end{aligned} \tag{5.2}$$

where

$$\nu(y, \alpha) = \begin{cases} -\operatorname{sgn} y \cdot \tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi}\operatorname{sgn} y \cdot \log|y|) & \text{if } \alpha = 1. \end{cases}$$

If $\alpha < 2$, the measure Λ is unique, and conversely, every finite measure on the unit sphere $\mathbb{S}^{n-1}$ and $\mu \in \mathbb{R}^{n-1}$ uniquely defines an α-stable distribution.

The measure Λ appearing in (5.2) is called the *spectral measure* of X. If $\mu = \mathbf{0}$ and X has spectral measure Λ, we write $X \sim S_\alpha(\Lambda)$.

Interestingly, Theorem 5.2.2 implies that when $\alpha < 2$ and $n \geq 2$, the parameter space of an n-dimensional α-stable random vector is infinite dimensional, contrary to when $\alpha = 2$ when it is well know to have dimension $2n + \binom{n}{2}$. Another consequence of (5.2) is that if $X_1 \sim S_\alpha(\Lambda_1)$ and $X_2 \sim S_\alpha(\Lambda_2)$, then $X_1 + X_2 \sim S_\alpha(\Lambda_1 + \Lambda_2)$. If Λ is finitely supported, it follows that we can represent the random variable $X \sim S_\alpha(\Lambda)$ as follows.

Theorem 5.2.3 (Example 2.6.3 on p. 69 in [16])**.** *Let Λ be a finitely supported symmetric measure on $\mathbb{S}^{n-1}$ with weight γ_j at $\pm\mathbf{x}_j \in \mathbb{S}^{n-1}$ for $j = 1, 2, \ldots, m$,*

and let A_Λ be a matrix with columns given by $\{(2\gamma_j)^{1/\alpha}\mathbf{x}_j\}_{1\le j\le m}$. Moreover, let $\mathbf{S} = (S_1, S_2, \dots, S_m)$ be a random vector with i.i.d. entries, each with distribution $S_\alpha(1,0,0)$. Then

$$X := A_\Lambda \mathbf{S} \sim S_\alpha(\Lambda).$$

For general spectral measures, we have the following result, which is called the *Poisson representation* of a stable random vector. We present it here only for symmetric spectral measures Λ, but similar representations exist also in the general case.

Theorem 5.2.4 (Theorem 3.10.1 on p. 149 in [16])**.** *Fix $\alpha \in (0,2)$ and $n \ge 1$ and let Λ be a symmetric measure on $\mathbb{S}^{n-1}$. Further, let C_α be as in* (5.1)*, let (Γ_i) be the arrival times of a rate one Poisson process and let $\mathbf{W}_i$ be chosen according to the normalized spectral measure $\Lambda/\Lambda(\mathbb{S}^{n-1})$. Then*

$$C_\alpha^{1/\alpha}\Lambda(\mathbb{S}_1)^{1/\alpha}\sum_{i=1}^{\infty}\Gamma_i^{-1/\alpha}\mathbf{W}_i$$

converges almost surely to an α-stable random vector with spectral measure Λ.

Yet another way to understand stable random vectors is as follows. If X is a n-dimensional α-stable random vector then for any $\mathbf{u} \in \mathbb{R}^n\backslash\{0\}^n$ the projection $\langle \mathbf{u}, X\rangle$ is an α-stable random variable with some location parameter $\mu(\mathbf{u})$, scale parameter $\sigma(\mathbf{u})$ and skewness $\beta(\mathbf{u})$. One can show that (see e.g. Example 2.3.4 on p. 67 in [16])

$$\sigma(\mathbf{u}) = \left(\int_{\mathbb{S}^{n-1}} |\langle \mathbf{u}, \mathbf{s}\rangle|^\alpha \Lambda(d\mathbf{s})\right)^{1/\alpha} \tag{5.3}$$

$$\beta(\mathbf{u}) = \sigma(\mathbf{u})^{-\alpha}\int_{\mathbb{S}^{n-1}} |\langle \mathbf{u}, \mathbf{s}\rangle|^\alpha \operatorname{sgn}\langle \mathbf{u}, \mathbf{s}\rangle \Lambda(d\mathbf{s})$$

and

$$\mu(\mathbf{u}) = \begin{cases} \langle \mathbf{u}, \mu\rangle & \text{if } \alpha \ne 1 \\ \langle \mathbf{u}, \mu\rangle - \frac{\pi}{2}\int_{\mathbb{S}^{n-1}} \langle \mathbf{u}, \mathbf{s}\rangle \log|\langle \mathbf{u}, \mathbf{s}\rangle| \Lambda(d\mathbf{s}) & \text{if } \alpha = 1. \end{cases}$$

This is called the *projection parametrization* of X, and one can show that these functions determine Λ (see e.g. [1]). Using this parametrization with $n = 1$ and $\mathbf{u} = 1$, it follows that if $\mu = 0$, then the spectral measure with mass λ_- at -1 and mass λ_+ at 1 corresponds to

$$\sigma = (\lambda_+ + \lambda_-)^{1/\alpha}$$

and

$$\beta = \frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-}.$$

We now give a few examples.

Example 5.2.5. Suppose that $X_1, X_2, \ldots, X_n \sim S_\alpha(1,0,0)$ are independent. Then $X := (X_1, X_2, \ldots, X_n)$ has independent components, and using Theorem 5.2.3 we see that the corresponding spectral measure has mass $1/2$ at $\pm e_i$ for $i = 1, 2, \ldots, n$. Clearly the distribution of X is symmetric and permutation invariant. However, using (5.3) one sees that the probability density function is not rotation invariant unless $\alpha = 2$. We draw the density of (X_1, X_2) in Figure 5.3.

Example 5.2.6. Let Λ be the uniform measure on $\mathbb{S}_{n-1}$ with

$$\int_{\mathbb{S}_{n-1}} |\mathbf{s}(1)|^\alpha d\Lambda(\mathbf{s}) = 1.$$

Then $X_\alpha(\Lambda)$ is a stable random vector whose distribution is permutation invariant and symmetric. Interestingly, this example coincides with the previous example when $\alpha = 2$, but not when $\alpha < 2$.

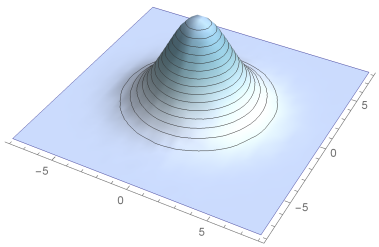
Example 5.2.7. Let $S_0, S_1, \ldots, S_n \sim S_\alpha(1,0,0)$ be independent, let $a \in (0,1)$, and for $i = 1, 2, \ldots, n$ define X_i by

$$X_i := aS_0 + (1 - a^\alpha)^{1/\alpha} S_i.$$

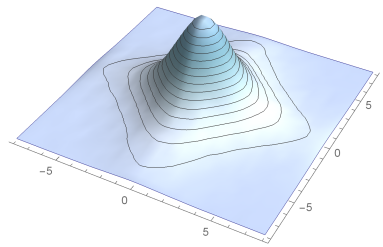
By definition, the distribution of $X = (X_1, X_2, \ldots, X_n)$ is symmetric and permutation invariant. Using Theorem 5.2.3 we see that the corresponding spectral measure has finite support, with mass $(a\sqrt{n})^\alpha/2$ at $\pm(1,1,\ldots,1)/\sqrt{n}$ and mass $(1-a^\alpha)/2$ at $\pm e_i$ for $i = 1, 2, \ldots, n$. When $\alpha = 2$, this corresponds to the multivariate Gaussian vector with covariance matrix $A = (a_{ij})$ with $a_{ii} = 2$ and $a_{ij} = 2a^2$ for $i, j = 1, 2, \ldots, n$.

In Figures 5.3 and 5.4 we draw the probability density function for a few different spectral measures to illustrate how the stability index and spectral measure affects the shape of the probability density function of the corresponding random vector. In both figures, we use only finitely supported spectral measures. One can show that if we fix $\alpha \in (0,2]$, then the set of probability density functions corresponding of to α-stable random vectors and finitely supported spectral measures is in fact dense in the set of all probability density functions corresponding to α-stable random vectors, both when compared pointwise, but also when integrated over Borel sets ([6]).

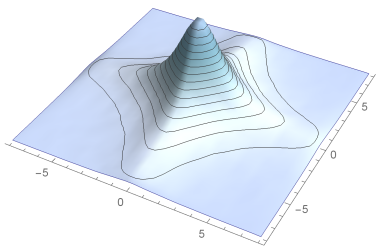
If X is a multivariate stable vector with stability index $\alpha < 2$ and spectral measure Λ, then Figures 5.3 and 5.4 suggest that if $r > 0$ is large then $\frac{X}{\|X\|} \mid \|X\| > r$ is concentrated near the points where Λ has support. This observation turns out



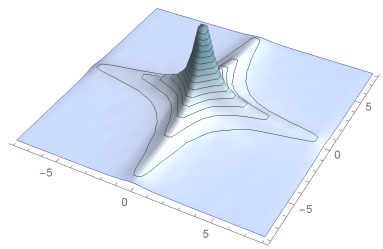
(a) $\alpha = 2$



(b) $\alpha = 1.2$

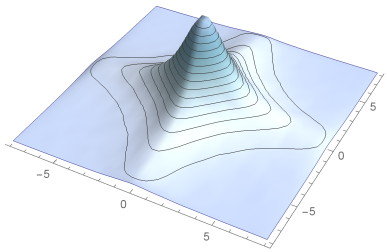


(c) $\alpha = 0.8$

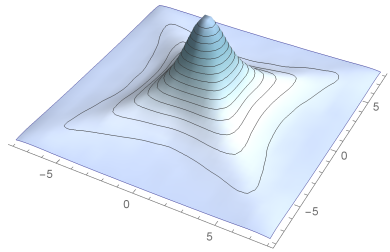


(d) $\alpha = 0.4$

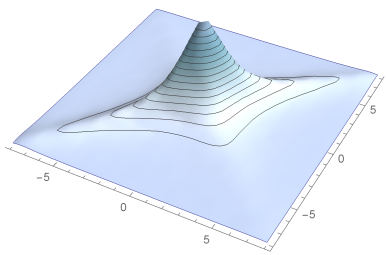
Figure 5.3: The four figures above illustrate how the shape of the probability density function is affected by the stability index α. In all four figures above we have drawn the probability density function of a multivariate stable vector (X_1, X_2) with independent components having distribution $S_\alpha(2, 0, 0)$, corresponding to a spectral measure with unit mass at $\pm(0, 1)$ and $\pm(1, 0)$. Note in particular that the probability density function is not rotation invariant unless $\alpha = 2$.



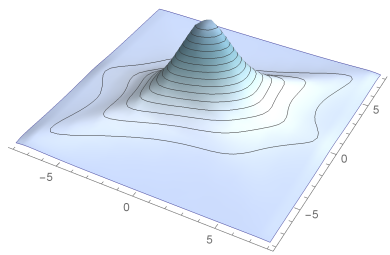
(a) Unit mass at $\pm(1,0)$ and $\pm(0,1)$.



(b) Unit mass at $\pm(1,1)/\sqrt{2}$ and $\pm(1,-1)/\sqrt{2}$.



(c) Unit mass at $\pm(1,1)/\sqrt{2}$ and mass 0.5 at $\pm(1,-1)/\sqrt{2}$.



(d) Unit mass at $\pm(1,1)/\sqrt{2}$, mass 0.4 at $\pm(1,-1)/\sqrt{2}$ and mass 0.6 at $\pm(1,0)$.

Figure 5.4: The four figures above illustrate how the spectral measure affects the shape of the probability density function of a multivariate stable vector $X = (X_1, X_2)$. In all cases above $\mu = \mathbf{0}$ and $\alpha = 0.8$. Note that the figures suggests that if we would condition X on $\|X\|$ being large, then X would be close to a multiple of a vector where the spectral measure has support.

to be valid, and in fact for any $E \subseteq \mathbb{S}^{n-1}$ and $\alpha \in (0,2)$, we have that (see e.g. Corollary 6.20 in [3])

$$\lim_{r\to\infty} P(X/\|X\| \in E \mid \|X\| > r) = \Gamma(E)/\Gamma\big(\mathbb{S}^{n-1}\big).$$

5.3 Properties of threshold stable vectors

Our main interest in Gaussian and stable random vectors comes from using them as a source of positively correlated probability measures on $\{0,1\}^n$. We obtain such measures by fixing some $h \in \mathbb{R}$ and defining a random vector $X^h := (X_1^h, X_2^h, \ldots, X_n^h)$ by, for $i \in [n]$, letting

$$X_i^h := I(X_i > h).$$

If X is a stable random vector, we say that X^h is a *threshold stable vector*, and similarly, if X is a Gaussian random vector we say that X^h is a *threshold Gaussian vector*. The idea is that for any fixed $\alpha \in (0,2]$, the set of measures on $\{0,1\}^n$ which can be obtained as the law of a threshold α-stable random vectors is a natural subclass of all Bernoulli measures on $\{0,1\}^n$ which one might be able to say someting about for certain questions due to the underlying dependency structure.

If we fix the marginals (p_i) of a Bernoulli measure μ, it is easy to see that we get $2^n - n - 1$ degrees of freedom. If we further assume that μ is symmetric, in that $\mu(x) = \mu(-x)$ for any $x \in \{-1,1\}^n$, this reduces to $2^{n-1} - 1$ degrees of freedom. If we consider the set of Gaussian random vectors whose marginals have mean zero and variance one, then we get exactly $\binom{n}{2}$ parameters, namely the covariances between the marginals. From this it follows that the set of probability measures on $\{-1,1\}^n$ which can be constructed as threshold Gaussian vectors is a proper subset of the set of all Bernoulli measures on the same space. If we instead consider threshold stable random vectors for some fixed index of stability $\alpha \in (0,2]$, one can show that the set of probability distributions one can obtain from threshold stable vectors is increasing as α decreases. Moreover, one can show that by taking α sufficiently close to zero, we can get arbitrarily close to any Bernoulli disribution. In particular, this shows that the family of threshold stable distributions is much richer than the set of threshold Gaussian distributions.

Paper A: Monotonicity properties of exclusion sensitivity

When studying whether or not a particular sequence of functions is exclusion sensitive, it is natural to ask how dependent this property is on the sequence of graphs on which we run the exclusion process. In particular, one could ask whether a sequence of functions which is exclusion sensitive given a sequence of graphs is still exclusion sensitive if we add edges to that graph. One reason to suspect that this would be true is that adding more edges to a graph intuitively should make the Markov chain more diffusive. In [5], the authors asked if this would be true if one adds all possible edges to obtain a complete graph, and conversely if any sequence of Boolean functions which is exclusion stable with respect to a sequence of complete graphs will be exclusion stable for any sequence of graphs with the same number of vertices but possibly fewer edges. In this paper, we answer both these questions affirmatively. Conversely, we show that if we add only some, but not all, edges then the answer to both questions is no in general.

Paper B: The spectrum and convergence rates of exclusion and interchange processes on the complete graph

In this paper, we further develop the ideas used in Paper A to give bounds on the convergence rates of exclusion and interchange processes on the complete graph. These bounds were already known, but our proof is more elementary than previously known proofs.

Paper C: Denseness of volatile and nonvolatile sequences of functions

In this paper, we investigate the relationship between the concepts of volatility, noise stability and noise sensitivity of sequences of Boolean functions. In particular, we show that the set of volatile sequences of Boolean functions is dense in the set of all sequences of Boolean functions, as well as in the set of noise stable sequences of Boolean functions. Conversely, by constructing a sequence of functions inspired

by the Cantor set, we also show that the set of non-volatile sequences of Boolean functions is not dense in the set of volatile sequences of Boolean functions.

Paper D: Fortuin-Kastelyn representations for threshold Gaussian and stable vectors

In this paper, our main objective is to investigate which threshold Gaussian and threshold stable random vectors have color representations. While this is the goal of the paper, by considering several different approaches to this problem, we also obtain methods which can be used to tackle similar questions for other families of processes and also give some necessary and sufficient conditions for a color representation to exist. The results of this paper include formulas for the rank of the corresponding linear operator, a formula for the tail behavious of multivariate stable random vectors, necessary conditions to have a color representation in a non-fully dimensional setting and positive results for discrete Gaussian free fields — both for when the threshold is zero and when it is very large.

BIBLIOGRAPHY

[1] Husein Abdul-Hamid and John P. Nolan. *Multivariate stable densities as functions of one-dimensional projections.* Journal of Multivariate Analysis, Vol 67, pp. 80-89 (1998).

[2] David Aldous and James Allen Fill. *Reversible Markov chains and random walks on graphs.* Unfinished monograph, recompiled 2014, available at https://www.stat.berkeley.edu/users/aldous/RWG/book.html, (2002).

[3] Aloisio Araujo and Evarist Giné. *The central limit theorem for real and Banach valued random variables.* Wiley, NY (1980).

[4] Itai Benjamini, Gil Kalai and Oded Schramm. *Noise sensitivity of Boolean functions and applications to Percolation.* Publications Mathématiques de l'Institut des Hautes Études Scientifiques 90:5–43, (1999).

[5] Erik I. Broman, Christophe Garban and Jeffrey E. Steif. *Exclusion sensitivity of Boolean functions.* Probability Theory and Related Fields, 155, pp. 621–663 (2013).

[6] Tomasz Byczkowski, John P. Nolan and Balram Rajput. *Approximations of multivariate stable densities.* Journal of Multivariate Analysis, Vol. 46, Issue 1, pp. 13–31 (1993).

[7] William Feller. *An introduction to probability theory and its applications, Volume II.* Wiley series in probability and mathematical statistics, Second edition, John Wiley & Sons (1966).

[8] Geoffrey G. Grimmett. *The random-cluster model.* Springer (2006).

[9] Olle Häggström. *Coloring percolaiton clusters at random.* Stochastic Proceses and their Applications, Vol. 96, pp. 213–242 (2001).

[10] Johan Jonasson and Jeffrey E. Steif. *Volatility of Boolean functions.* Stochastic Processes and their Applications 126(10):2956–2975, (2016).

[11] Hubert Lacoin and Rémi Leblond. *Cutoff phenomenon for the simple exclusion process on the complete graph.* ALEA, Latin American Journal of Probability and Mathematical Statistics, Number 8, pp. 285-301 (2011).

[12] David A. Levin, Yuval Peres and Elizabeth L. Wilmer. *Markov chains and mxing times.* American Mathematical Society (2009).

[13] Thomas M. Liggett. *Continuous time Markov processes.* American Mathematical Societt (2010).

[14] Ilya Molchanov. *Convex and star-shaped sets associated with multivariate stable distributions, I: Moments and densities.* Journal of Multivariate Analysis, Volume 100, Issue 10, pp 2195–2213 (2009).

[15] Ryan O'Donnell. *Analysis of Boolean functions.* Cambridge Univesity Press (2014).

[16] Gennady Samorodnitsky and Murad S. Taqqu. *Stable non-Gaussian random processes*, (1995).

[17] Saul Stahl. *The evolution of the normal distribution.* Matematics Magazine, Volume 79, pp. 96-113 (2006).

[18] Anatoliy V. Skorokhod. *Basic principles and applications of probability theory.* Springer-Verlag Berlin Heidelberg 2005.

[19] Jeffrey E. Steif and Johan Tykesson. *Generalized divide and color models.* preprint available at arXiv:1702.04296 (2017).

[20] Vladimir M. Zolotarev. *One-dimensional stable distributions.* Volume 65 of "Translations of mathematical monographs", American Mathematical Society. Translation from the 1983 Russian edition, (1986).