THESIS FOR THE DEGREE OF LICENTIATE OF PHILOSOPHY

On properties of generalizations of noise sensitivity

Malin Palö Forsström

CHALMERS
UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden 2015

On properties of generalizations of noise sensitivity

Malin Palö Forsström

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31 772 1000

Printed in Gothenburg, Sweden 2015

On properties of generalizations of noise sensitivity

Malin Palö Forsström

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

Abstract

In 1999, Benjamini et. al. published a paper in which they introduced two definitions, noise sensitivity and noise stability, as measures of how sensitive Boolean functions are to noise in their parameters. The parameters were assumed to be Boolean strings, and the noise consisted of each input bit changing their value with a small but positive probability. In the three papers appended to this thesis, we study generalizations of these definitions to irreducible and reversible Markov chains.

Keywords noise sensitivity, spectral analysis, Markov chains, exclusion processes.

List of appended papers

The following papers are included in this thesis.

A. Noise sensitivity and noise stability for Markov chains. Malin Palö Forsström. (preprint)

B. A noise sensitivity theorem for Schreier graphs. Malin Palö Forsström. (preprint)

C. Monotonicity properites of exclusion sensitivity. Malin Palö Forsström. (submitted)

Acknowledgements

First and foremost, I would like to thank my supervisor Jeff Steif, not only for discussions and suggestions concerning my research, but also for trying to teach me how to do research and think about mathematics. I am also very grateful to my co-supervisor Johan Jonasson for our discussions during the last six months, but also for your dedication and great enthusiasm during this time.

Further, I would like to thank Serik Sagitov, Anders Martinsson and Maria Roginskaya for, by asking about how my research is going and by giving advise on what is necessary to perform well, motivating me to try to perform better.

Many thanks also goes to all my colleagues at the department for making my workplace feel like a home. Especially, thank you Ivar for letting me borrow your sofa, Claes for letting me disturb you while you are working, Mariana for discussions about shoes and nailpolish and Roza for making our room such a nice place to work in.

Finally, I would like to thank Étienne, for your love, patience, encouragement and constant support.

Contents

Chapter One

Introduction

A Boolean function is a function which only assumes two values, minus one and one. Due to this quite restrictive choice of range, Boolean functions might not seem to be very interesting to study, but in fact, they are quite natural, and this is arguably one of the most frequently occurring type of function in real life. In fact, any question whose answer will be either one of two choices, such as yes or no, true or false, one out of two candidates, etc., can be described by a Boolean function. We think of the parameters of the functions as the information which is taken into account when making such a decision.

In any real life setting, a very sensible assumption is that not all available information is correct, as the information we have can have been misread, misinterpreted, or simply wrong. We think of this incorrect information as noise in our parameters. To minimize the effect of such noise, the function we use to make a decision should preferably maximize the probability of making the same decision as with completely correct information. In the very least, we would like such a function to allow us to make this probability arbitrarily close to one by taking more and more, possibly errornous, information into account. This thesis is about the extent to which there exists functions that are good in this regard for different types of information and noise, and about the properties of these functions when they exist.

The *noise sensitivity* of a sequence of Boolean functions was first defined in [1], although the same quantities in slightly different settings had been studied earlier. One reason for the interest in the subject that emerged from this paper was that the authors were able to show that percolation crossings of a triangular grid was highly sensitive to a small proportion of the edges being rerandomized. To do this, the authors proved a result which later became known as the noise sensitivity theorem.

In short, in [1], a sequence of Boolean functions $f_n \colon \{-1,1\}^n \to \{-1,1\}$ was said to be asymptotically noise sensitive if, in the limit, independently changing the sign of each parameter with a small probability caused all information about the initial value of the function to be lost. Asymptotic noise stability was defined to capture an opposite behaviour, namely that in the limit, independently resampling each parameter with a small probability ε was very unlikely to change the value of the function.

Since 1999, when [1] was published, several papers have been written that studies the noise sensitivity or noise stability of Boolean functions and related concepts, e.g. [13, 2], [14, 9, 11] [7, 12], [10] and [6].

Already in [1], the first paper on noise sensitivity, the authors mentioned that the noise sensitivity and noise stability of a Boolean function might be interesting also for other types of noise. In fact, the authors included a section on the noise identical to the noise described above, except that the number of flipped parameters was fixed instead of random. They also mentioned that it could be of interest to replace the original noise with some other model from statistical mechanics. In [2], one such noise was studied, namely an exclusion process on the complete graph. Apart from giving analogues to many of the results in [1], the authors also included results relating the noise sensitivity of a Boolean function with respect this noise to the noise sensitivity of the same function with respect to the original noise.

Closer in time to the original paper was [13], in which the notions of noise sensitivity and noise stability as given in [1] was studied for functions whose domains were more naturally defined as trees than as cubes. In particular, the authors asked which families of trees could be embedded into the hypercube in such a way that the notions of noise sensitivity and noise stability made sense in relation to the structure of the tree and the chosen function. Instead of altering the source of noise, the authors defined alternative notions of noise sensitivity in this setting, calling a sequence of functons tree sensitive if it was noise sensitive given some embedding of its tree domain onto the hypercube, and analogously, tree stable if it was noise stable for all such embeddings. The generalization given in this paper is not studied in any of the three appended papers, but provides an alternative solution of how to extend the definitions in [1], a topic with which we will be preoccupied through most of this thesis.

We will now briefly describe the mathematical tools and models which will be most frequently used in the appended papers, as well as give more formal descriptions of the problems studied witihin this thesis. In what follows, basic knowledge of mathematical ideas such as linear algebra, graphs, stochastic processes and discrete time Markov chains will be assumed.

1.1 CONTINUOUS TIME MARKOV CHAINS

At the heart of this thesis is the concept of *continuous time Markov chains*. A continuous time Markov chain differs from a discrete time Markov chain in that the time between two consequtive jumps is not constant, but instead a random variable whose law is an exponential distribution. The parameter of this exponential distribution might depend on the current state of the Markov chain. When we think about continuous time Markov chains, we often think about the exponential law as a set of clocks, with exactly one clock $C_{ss'}$ for each pair of distinct states (s, s') of the Markov chain. Each such clock $C_{ss'}$ has an associated *rate* $q_{ss'}$, and the time between two consequtive ticks of the clock has law $\exp q_{ss'}$. When $C_{ss'}$ ticks, if the current position of the Markov chain is s, it jumps to position s'. As the minimum of two exponential distributions with parameters λ_1 and λ_2 respectively is again an exponential distribution with parameter $\lambda_1 + \lambda_2$, it follows that the time until a Markov chain X moves from a state s has law $\exp(\sum_{s' \neq s} q_{ss'})$. It also follows that we could equivalently put a clock C_s on each state s of the Markov chain, whose time between two ticks is governed by an exponential distribution with rate $-q_{ss} \coloneqq \sum_{s' \neq s} q_{ss'}$, and let a Markov chain which is at position s when the corresponding clock ticks, move to state s' with probability $q_{ss'}/(-q_{ss})$. The ratios $p_{ss'} \coloneqq q_{ss'}/(-q_{ss})$ are called the *jump probabilities* of the Markov chain.

As an analogue of the transition matrix $P = (p_{ss'})_{s,s' \in S}$ of a discrete time Markov chain with state space S, we define the *generator* $Q = (q_{ss'})_{s,s' \in S}$ of a continuous time Markov chain X. To get something similar to the n-step transition probabilities for a discrete time Markov chain, we define

$$H_t(s, s') \coloneqq P(X_t = s' \mid X_0 = s),$$

where X_t denotes the position of the Markov chain X at time t. Similarly, for a function $f\colon S \to \mathbb{R}$, we define

$$H_t f(s) \coloneqq \mathbb{E}[f(X_t) \mid X_0 = s].$$

As for sufficiently small $\varepsilon > 0$,

$$H_{t+\varepsilon}(s, s') \approx H_t(s, s') \cdot e^{\varepsilon q_{s's'}} + \sum_{s'' \in S \setminus \{s'\}} H_t(s, s'')(1 - e^{-\varepsilon q_{s''s'}}),$$

it follows that

$$H_t = e^{Qt}.$$

That is, the matrix Q completely characterizes the evolution of the corresponding Markov chain.

Throughout this thesis, almost all our Markov chains will be irreducible, meaning that for all $s, s' \in S$,

$$P(X_t = s' \mid X_t = s) > 0$$

for all $t > 0$. From this it follows that there is always a unique stationary measure π with respect to X (see e.g. [8]). If not otherwise stated, we will choose X_0 according to this distribution.

A very useful tool for analyzing continuous time Markov chain, which is not always mentioned in a first course on Markov chains, is the eigenvectors and eigenvalues of the generator Q, or analogously, of the transition matrix P in a discrete time setting. The purpose of this section is to briefly introduce these tools and derive basic equalities as examples of how they can be used. For more on the spectral analysis of Markov chains, we refer the reader to [3].

In what follows, let Q be the generator of a reversible and irreducible continuous time Markov chain X, i.e. an irreducible continuous time Markov chain whose generator$Q$ satisfies

$$\pi(s)q_{ss'} = \pi(s')q_{s's}$$

for all states s and s' in the state space S of X. Then the matrix

$$(-\pi(s)^{1/2}q_{ss'}\pi(s')^{-1/2})_{ss'\in S}$$

is symmetric, and consequently has an orthonormal basis of eigenvectors

$$\chi_1, \chi_2, \ldots, \chi_{|S|}$$

with corresponding real eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \ldots, \leq \lambda_{|S|}$$

with respect to any inner product on $\mathbb{R}^S = \{f\colon S \to \mathbb{R}\}$. In this thesis, for functions $f, g\colon S \to \mathbb{R}$, we will almost exclusively use the inner product given by

$$\langle f, g\rangle \coloneqq \mathbb{E}[f(X)g(X)] = \sum_{s\in S} \pi(s)f(s)g(s). \tag{1.1}$$

Note that irreducibility is necessary to get a unique stationary distribution π with full support, and reversibility is necessary to get the symmetry needed to get an orthogonal basis of eigenvectors.

If we define

$$\psi_i(S) \coloneqq \pi^{-1/2}(s)\chi_i(s)$$

for $s \in S$, then ψ_i is an eigenvector of $-Q$ with corresponding eigenvalue λ_i. As $H_t = e^{Qt}$, it follows that $H_t\psi_i = e^{-\lambda_i t}\psi_i$, that is ψ_i is an eigenvector to H_t with corresponding eigenvalue $e^{-\lambda_i t}$ for each $i \in \{1, 2, \ldots, |S|\}$.

Another way of characterizing the eigenvalues $\lambda_1, \ldots, \lambda_{|S|}$ is given by noting that

$$\lambda_i = \frac{\langle \psi_i, -Q\psi_i\rangle}{\langle \psi_i, \psi_i\rangle} = \inf_{\substack{f\colon S\to\mathbb{R},\\ f\perp\{\psi_1,\ldots,\psi_{i-1}\},\\ f\not\equiv 0}} \frac{\langle f, -Qf\rangle}{\langle f, f\rangle}. \tag{1.2}$$

The quotient in the last expression in (1.2) is called the *Rayleigh quotient* of Q. As

$$\begin{aligned}
\langle f, -Qf\rangle &= \sum_{s\in S} \pi(s) f(s) \sum_{s'\in S} -q_{ss'} f(s') \\
&= \sum_{s\in S} \pi(s) f(s) \sum_{s'\in S\setminus\{s\}} q_{ss'} \left(f(s) - f(s')\right) \\
&= \sum_{s\in S} \pi(s) \sum_{s'\in S\setminus\{s\}} q_{ss'} f(s) \left(f(s) - f(s')\right) \\
&= \frac{1}{2} \sum_{s\in S} \pi(s) \sum_{s'\in S\setminus\{s\}} q_{ss'} \left(f(s) - f(s')\right)^2
\end{aligned} \tag{1.3}$$

we can directly deduce that

(i) $\lambda_i \geq 0$ for all $i \in \{1, 2, \ldots, |S|\}$.

Using (1.2) and (1.3) for a $\{-1, 1\}$-valued function, we get that

(ii) $\lambda_i \leq 2 \sum_{s\in S} \pi(s) \cdot (-q_{ss})$ for all $i \in \{1, 2, \ldots, |S|\}$.

As the all ones vector is an eigenvector of $-Q$ with corresponding eigenvector 0, it follows from (i) that we can assume that

(iii) $\psi_1 \equiv 1$ and $\lambda_1 = 0$.

Further, it follows from (1.3) that given that X is irreducible, we must have that

(iv) $\lambda_2 > 0$.

As the eigenvectors $\psi_1, \ldots, \psi_{|S|}$ constitute an orthonormal basis with respect to the inner product given by (1.1), any function $f\colon S \to \mathbb{R}$ can be written as

$$f = \sum_{i=1}^{|S|} \frac{\langle f, \psi_i\rangle}{\langle \psi_i, \psi_i\rangle} \psi_i.$$

Assuming that $\langle \psi_i, \psi_i\rangle = 1$, i.e. that the eigenvectors are normalized, this simplifies to

$$f = \sum_{i=1}^{|S|} \langle f, \psi_i\rangle \psi_i.$$

The terms $\hat{f}(i) \coloneqq \langle f, \psi_i\rangle$, $i = 1, 2, \ldots, |S|$, are called the Fourier coefficients of f with respect to the basis $\psi_1, \ldots, \psi_{|S|}$, and are useful for expressing several probabilistic quantities which will be of interest of us. One of the simplest such quantities is the expected value of $f(X_0)$ for function $f\colon S \to \mathbb{R}$. Using that $\psi_1 \equiv 1$, this can be expressed as

$$\mathbb{E}[f(X_0)] = \mathbb{E}[f(X_0) \cdot 1] = \langle f, 1 \rangle = \langle f, \psi_1 \rangle = \hat{f}(1).$$

Although a little bit more complicated, we also get that

$$\begin{aligned}
\mathbb{E}[f(X_0)f(X_t)] = \mathbb{E}[f(X_0)H_t f(X_0)] = \langle f, H_t f \rangle &= \Big\langle \sum_{i=1}^{|S|} \hat{f}(i)\psi_i, H_t \sum_{i=1}^{|S|} \hat{f}(j)\psi_j \Big\rangle \\
&= \Big\langle \sum_{i=1}^{|S|} \hat{f}(i)\psi_i, \sum_{i=1}^{|S|} \hat{f}(j) H_t \psi_j \Big\rangle = \Big\langle \sum_{i=1}^{|S|} \hat{f}(i)\psi_i, \sum_{i=1}^{|S|} \hat{f}(j) e^{-\lambda_j t} \psi_j \Big\rangle \\
&= \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} \hat{f}(i)\hat{f}(j) e^{-\lambda_j t} \langle \psi_i, \psi_j \rangle = \sum_{i=1}^{|S|} e^{-\lambda_i t} \hat{f}(i)^2
\end{aligned}$$

and consequently,

$$\mathrm{Cov}(f(X_0), f(X_t)) = \sum_{i=2}^{|S|} e^{-\lambda_i t} \hat{f}(i)^2.$$

From the expression for the covariance it directly follows that

$$\mathrm{Var}(f(X_0)) = \lim_{t \to 0} \mathrm{Cov}(f(X_0), f(X_t)) = \sum_{i=2}^{|S|} \hat{f}(i)^2, \tag{1.4}$$

but more interestingly, it also follows that

$$\mathrm{Cov}(f(X_0), f(X_t)) \leq e^{-\lambda_2 t}\, \mathrm{Var}(f(X_0)). \tag{1.5}$$

This suggests that if one want to study the decorrelations of a function f of a continuous time Markov chain, one should pick $t = O(1/\lambda_2)$. For this reason, $1/\lambda_2$ is called the *relaxation time* of the Markov chain and is denoted by t_{rel}. One thing which is important to keep in mind when talking about the relaxation time of a continuous time Markov chain is that its value heavily depends on the time scaling of the Markov chain. This is especially true when the Markov chain is a random walk on a connected graph, where it is not obvious how the rates rates $-q_{ii}$ should be chosen. Given a Markov chain X with generator Q, for any $\alpha > 0$ the matrix $Q' = \alpha Q$ defines a Markov chain X' which essentially is the same Markov chain as X, but for which time runs faster or slower depending on whether α is larger or smaller than one. It is easy to see that the eigenvalues of $-Q'$ will be given by $\alpha\lambda_1, \ldots, \alpha\lambda_{|S|}$ and consequently, the relaxation time of X' is given by $1/(\alpha\lambda_2)$.

A rich source of examples of continuous time Markov chains is so called *continuous time random walks*. To define what we mean by a continuous time random walk, let G be any connected graph. We say that a continuous time Markov chain with state space $S = V(G)$ is a continuous time random walk on G if for any to distict states $s, s' \in S$,

1. $q_{ss'} = 0$ whenever $(s, s') \notin E(G)$ and
2. there is $\alpha > 0$, such that either
   a) $q_{ss'} = \alpha/\deg(s)$ for all $(s, s') \in E(G)$ or
   b) $q_{ss'} = \alpha$ for all $(s, s') \in E(G)$.

The only difference between a continuous time random walk and a discrete time random walk is thus that the time between two steps in the random walk is random with an exponential distribution, whose rate is either the same for all vertices in the graph (condition 2a) or the same for all edges in the graph (condition 2b). For regular graphs, conditions 2a and 2b coincide, but for general graphs they result in different Markov chains.

We now give three specific examples of continuous time Markov chains walks which will appear in the appended papers.

Example 1.3.1. Fix $n \in \mathbb{N}$ and let $S = \{-1, 1\}^n$. Pick a clock C_j for each coordinate $j \in \{1, 2, \ldots, n\}$ and assume that all the clocks have the same rate. When a clock C_j ticks, change the sign of the jth coordinate of the current state. This defines a Markov chain, which can also be thought of as describing a continuous time random walk on a n-dimensional Hamming cube.

If we let each clock have rate 1, then the relaxation of this Markov chain is of order one (see eg. [5]). Moreover, if we pick a one-to-one mapping $i \mapsto S_i \subset \{1, 2, \ldots, n\}$ such that $|S_i| < |S_j|$ whenever $i < j$, then it is well known (see e.g. [5]) that we can pick the eigenvectors $\{\psi_i\}_i$ such that $\lambda_i = |S_i|$ and

$$\psi_i(s) = \prod_{i \in S_i} (-1)^{s(i)}.$$

If we think of this Markov chain as a continuous time random walk on the Hamming cube, it might be more natural to let the Markov chain jump with rate one. This would give each clock rate $1/n$, $t_{rel} = n$ and $\lambda_i = |S|/n$.

Example 1.3.2. Fix $n \in \mathbb{N}$ and let $S = \mathbb{Z}_n$. For all $i, j \in \mathbb{Z}_n$, let

$$q_{ij} = \begin{cases} 1/2 & \text{if } |i-j| = 1 \bmod n \\ -1 & \text{if } i = j \\ 0 & \text{else.} \end{cases}$$

This defines a continuous time Markov chain X which can be though of as a random walk on a circle. The eigenvectors of the generator fo this Markov chain are known to be given by

$$\psi_i(j) = \cos\left(\frac{2\pi i j}{n}\right)$$

with corresponding eigenvalues $\lambda_i = 1 - \cos\left(\frac{2\pi i}{n}\right)$, implying that the relaxation time is given by $(1 - \cos(2\pi/n))^{-1}$ (see e.g. [8]).

Example 1.3.3. Let G be any connected graph on n vertices, and let $\ell \in \{1, 2, \ldots, n-1\}$. Imagine a bag containing exactly ℓ black marbles and $n - \ell$ white marbles. Pick X_0 by randomly placing one marble from the bag on each vertex of the graph. Now for each $e \in E(G)$, put a clock C_e on e with rate $1/\max_{v \in V(G)} \deg v$. When a clock C_e ticks, we interchange the marbles on the endpoints of the edge e. Then X is said to be a (simple symmetric) *exclusion process* on G. The relaxation time for this Markov chains clearly depends on the choice of graph G, and in general neither the eigenvectors nor the eigenvalues are known. However, when G is e.g. the complete graph on n vertices, both the eigenvectors and eigenvalues can be calculated (see e.g. [4] and Paper C). Also, when $\ell = 1$, we recover a random walk on the graph G, where explicit formulas for both eigenvalues and eigenvectors are known for e.g. Hamming cubes and circles.

1.4 NOISE SENSITIVITY AND NOISE STABILITY

The terms *noise sensitivity* and *noise stability* were first mentioned in [1], where the authors gave the following definitions.

Definition 1.4.1. *For each $n \geq 1$, let $X^{(n)}$ be the continuous time Markov chain given by Example 1.3.1. A sequence $(f_n)_{n \geq 1}$, $f_n \colon \{-1,1\}^n \to \{-1,1\}$ of Boolean functions is said to be* noise sensitive *if for every $\varepsilon > 0$,*

$$\lim_{n \to \infty} \operatorname{Cov}(f_n(X_0^{(n)}), f_n(X_\varepsilon^{(n)})) = 0. \tag{1.6}$$

Definition 1.4.2. *For each $n \geq 1$, let $X^{(n)}$ be the continuous time Markov chain with state space $\{-1,1\}^n$ given by Example 1.3.1. A sequence $(f_n)_{n \geq 1}$, $f_n \colon \{-1,1\}^n \to \{-1,1\}$ of Boolean functions is said to be* noise stable *if*

$$\lim_{\varepsilon \to 0} \sup_n P(f_n(X_0^{(n)}) \neq f_n(X_\varepsilon^{(n)})) = 0. \tag{1.7}$$

The two definitions above are meant to capture two opposite behaviours of sequences of Boolean functions with domain $\{-1,1\}^n$, when their domains are exposed to the noise described in Example 1.3.1. Common examples of Boolean functions that are noise stable is the *Dictator function*;

$$\textsc{Dict}_n(s) = s(1)$$

and the *Majority function*;

$$\textsc{Maj}_n(s) = \begin{cases} 1 & \text{if } \sum_{i=1}^n s(i) > n/2 \\ -1 & \text{else.} \end{cases}$$

As an example of a noise sensitive function, a common examples is the *Parity function*

$$\textsc{Parity}_n(s) = \prod_{i=1}^n s(i).$$

A noise sensitive function which might be more interesting, defined for $n = 3^m$ and $m \in \mathbb{N}$, is called *Iterated-3-Majority*;

$$\text{I3}\textsc{Maj}_{3^m}(s) = \begin{cases} \textsc{Maj}_3(s) & \text{if } n = 3 \\ \textsc{Maj}_3(\text{I3}\textsc{Maj}_{3^{m-1}}((s(1), \dots s(3^{m-1}))) & \\ \qquad + \text{I3}\textsc{Maj}_{3^{m-1}}((s(3^{m-1}+1), \dots s(2*3^{m-1}))) & \\ \qquad + \text{I3}\textsc{Maj}_{3^{m-1}}((s(2*3^{m-1}+1), \dots, s(3*3^{m-1})))) & \text{else.} \end{cases}$$

In all papers appended to this thesis, we are studying questions related to the following two definitions, which generalize Definitions 1.4.1 and 1.4.2.

Definition 1.4.3. *For each $n \geq 1$, let $X^{(n)}$ be a continuous time Markov chain with corresponding state space S_n, and let $(t_n)_{n\geq 1}$ be a sequence of positive real numbers. A sequence $(f_n)_{n\geq 1}$, $f_n \colon \{-1,1\}^n \to \{-1,1\}$ of Boolean functions is said to be* noise sensitive *with respect to $(X^{(n)}, t_n)_{n\geq 1}$ if for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \operatorname{Cov}(f_n(X_0^{(n)}), f_n(X_{\varepsilon t_n}^{(n)})) = 0. \tag{1.8}$$

Definition 1.4.4. *For each $n \geq 1$, let $X^{(n)}$ be a continuous time Markov chain with corresponding state space S_n, and let $(t_n)_{n\geq 1}$ be a sequence of positive real numbers. A sequence $(f_n)_{n\geq 1}$, $f_n \colon \{-1,1\}^n \to \{-1,1\}$ of Boolean functions is said to be* noise stable *with respect to $(X^{(n)}, t_n)_{n\geq 1}$ if*

$$\lim_{\varepsilon\to 0} \sup_n P(f_n(X_0^{(n)}) \neq f_n(X_{\varepsilon t_n}^{(n)})) = 0. \tag{1.9}$$

In both Definition 1.4.1 and Definition 1.4.2 we look at the position of the Markov chain at time ε which, given the definition of this Markov chain, is a fraction ε of the corresponding relaxation time. At this time, the probability that the value of the first coordinate is not the same as at time zero is the same as the probability that the clock C_1 has ticked an odd number of times. As the times between two ticks are exponentially distributed, this has probability $(1-e^{-\varepsilon})/2$. As the clocks are independent, we obtain the following alternative interpretation of X_ε, namely that we can generate X_ε from X_0 by changing the value at each coordinate of X_0 with probability $(1-e^{-\varepsilon})/2$.

The discussion in the previous paragraph suggests two possible ways to generalize the definitions of noise sensitivity and noise stability to general sequences of Markov chains; either we compare $f_n(X_0^{(n)})$ with $f_n(X_{\varepsilon t_{rel}}^{(n)})$, or we compare $f_n(X_0^{(n)})$ with $f_n(X_{\varepsilon t}^{(n)})$ with some choice of t which is more relevant given some description of the Markov chain $X^{(n)}$. However, as it follows directly from (1.5) that all functions will be noise sensitive when $t_n = \omega(t_{rel})$, the relaxation time is the latest possibly interesting time to study. Also, as both the property of being noise sensitive and being noise stable is monotone in $(t^{(n)})_{n\geq 1}$, if one are interested in noise stable functions, this is a natural time to consider.

1.5 THE NOISE SENSITIVITY THEOREM

In [1], where noise sensitivity was first defined, the authors proved a result which connected noise sensitivity to the structure of the sets $\{s \in \{0,1\}^n : f_n(s) = 1\}$. Such a result is sometimes very useful when proving that a given sequence of functions in noise sensitive, as it removes what is often a major complication, namely understanding the diffusive behaviour of the Markov chain. The result, which is often called the noise sensitivity theorem, is formulated below. Before stating the theorem, we will need to introduce some additional notation, which will provide us with a measure of the size of the boundary of the sets $\{s \in \{-1,1\}^n : f_n(s) = 1\}$.

Given $s \in \{-1,1\}^n$ and $i \in \{1,2,\ldots,n\}$, let $s \oplus e_i$ be the element in $\{-1,1\}^n$ which is obtained by changing the sign of the ith coordinate in s. As an example, when $n=3$, this means that we can summarize the action of $s \mapsto s \oplus e_i$ by the following table.

s	$s \oplus e_1$	$s \oplus e_2$	$s \oplus e_3$
000	100	010	001
001	101	011	000
010	110	000	011
011	111	001	010
100	000	110	101
101	001	111	100
110	010	100	111
111	011	101	110

For $i \in \{1, 2, \ldots, n\}$, we define the *influence* of the ith variable on a function $f\colon \{-1,1\}^n \to \{-1,1\}$ to be

$$I_i(f) \coloneqq P\left(f(X_0) \neq f(X_0 \oplus e_i)\right),$$

where we assume that X_0 is chosen according to the stationary distribution of X and X is the continuous time Markov chain defined in Example 1.3.1.

Theorem 1.5.1 (The noise sensitivity theorem)**.** *Let $X^{(n)}$ be the continuous time Markov chain defined in Example 1.3.1 and let $(f_n)_{n\geq 1}$ be a sequence of Boolean functions with $f_n\colon \{-1,1\}^n \to \{-1,1\}$. If*

$$\lim_{n\to\infty} \sum_{i=1}^{n} I_i(f_n)^2 = 0$$

then $(f_n)_{n\geq 1}$ is noise sensitive with respect to $(X^{(n)}, 1)_{n\geq 1}$.

Apart from being a valuable tool when trying to show that more complicated sequences of functions are noise sensitivite, the noise sensitivity theorem is interesting in itself in that it at a first glace seems counter-intuitive. Given that the right hand side is very small, the boundary between the set $\{s \in \{0,1\}^n \colon f(s) = 1\}$ and its complement is in some sense small, why one could guess that the probability that a random walker will cross it should be small as well, rendering the function to be noise stable. This is however not the case, as concluded by the theorem.

The main objective of Paper B is to try to give an analogue of the noise sensitivity theorem in a more general setting. However, it should be noted that we cannot expect such a generalization to be valid for all Markov chains, as any such generalization requires that there is something in the definition of the Markov chain which can serve as an analogue of the coordinates of a Hamming cube to yield an analogue of the influences in the original statement of the theorem.

Chapter Two

Summary of appended papers

2.1 Paper A – Noise sensitivity and noise stability for Markov chains.

The main objective of this paper is to study the extensions of noise sensitivity and noise stability given by Definitions 1.4.3 and 1.4.4. In particular, this paper contains several results providing conditions on the sequence of Markov chains $(X^{(n)})_{n\geq 1}$ for which such an extension makes sense, in that both noise stable and noise sensitive sequences of Boolean functions exist at the relaxation time. The same questions for $t_n \leq t_{rel}^{(n)}$ are also discussed briefly.

Apart from its relevance for noise sensitivity and noise stability, many of the questions asked within the paper concerns just as much the definition of the relaxation time and its relation to the geometry of the corresponding Markov chain.

2.2 Paper B – A noise sensitivity theorem for Schreier graphs.

In this paper, we extend the noise sensitivity theorem from the setting of Definition 1.4.1 to the setting of Definition 1.4.3 in the special case when $X^{(n)}$ is a random walk on a Schreier graph. We then use this result to give an alternative proof of an extension of the noise sensitivity theorem to exclusion processes which was given in [2]. One reason that this might be interesting, is that it relates this version of the noise sensitivity theorem to the corresponding result in the original setting by giving a similar proof, which sheds some light on the differences and similarities of the two versions.

The proof of the noise sensitivity theorem for Schreier graphs closely follows the proof of the noise sensitivity theorem as given in [5], although new complications arise due to the more general setting.

2.3 Paper C – Monotonicity properites of exclusion sensitivity.

When studying noise sensitivity and noise stability for exclusion processes, a natural question is whether a sequence of functions $(f_n)_{n\geq 1}$ which is noise sensitive with respect to exclusion processes on graphs $(G_n)_{n\geq 1}$ will be noise sensitive for any sequence of graphs $(G'_n)_{n\geq 1}$ with $V(G_n) = V(G'_n)$ and $E(G_n) \subset E(G'_n)$

for all n. 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 [2], the authors asked if this would be true if G'_n is the complete graph on the vertex set of G_n, and also if any sequence of Boolean functions $(f_n)_{n\geq 1}$ which is noise stable with respect to an exclusion process on a sequence of complete graphs will be noise stable on any sequence of graphs with the same number of vertices but possible fewer edges.

Naturally, the answer to these questions depends on for how long the exclusion process is run for each n, given the time scaling. In [2], the authors gave each clock a rate which was at most $1/\max_{v\in V(G_n)} \deg v$ and let $t_n = 1$. Using these choices of parameters in the model, we prove that both questions can be answered yes when G'_n is the complete graph. Conversely, we show that if G'_n is allowed to be any graph satisfying $V(G_n) = V(G'_n)$ and $E(G_n) \subset E(G'_n)$ for all n, the answer is no. We also show that this holds even if stronger conditions, such as vertex transitivity, is imposed on the graphs G_n and G'_n. The proof of the positive result uses spectral analysis of continuous time Markov chains as well as a number results about the eigenvectors and eigenvalues of an exclusion process on a complete graph, whose proofs are also included in the paper.

Chapter Three

References

[1] 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.

[2] Erik Broman, Christophe Garban, and Jeffrey Steif. Exclusion sensitivity of Boolean functions. *Probability Theory and Related Fields*, 155:621–663, 2013.

[3] Fan Chung. *Spectral graph Theory*. American Mathematical Society, 1996.

[4] Yuval Filmus. Orthogonal basis for functions over a slice of the Boolean hypercube. arXiv:1406.0142, January 2014.

[5] Christophe Garban and Jeffrey Steif. *Noise Sensitivity of Boolean Functions and Percolation*. Cambridge University Press, first edition, 2014.

[6] Allan Hammond, Elchanan Mossel, and Pete Gabor. Exit time tails from pairwise decorrelation in hidden markov chains, with applications to dynamical percolation. *Electronic Journal of Probability*, 17(68):1–16, 2012.

[7] Nathan Keller and Guy Kindler. Quantitative relation between noise sensitivity and influences. *Combinatorica*, 33(1):45–71, 2013.

[8] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.

[9] Elchanan Mossel and Ryan O'Donnel. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Structures and Algorithms*, 26(4):418–436, 2005.

[10] Elchanan Mossel, Ryan O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010.

[11] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E. Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–366, 2006.

[12] Oded Schramm and Jeffrey Steif. Quantitative noise sensitivity and exceptional times for percolation. *Annals of Mathematics*, 171:619–672, 2010.

[13] Oded Schramm and Boris Tsirelson. Trees, not cubes: hypercontractivity, cosiness and noise stability. *Electronic communications in probability*, 4:39–49, 1999.

[14] Ke Yang. On the (im)-possibility of non-interactive correlation distillation. *Theoretical Computer Science*, 382(2):157–166, August 2007.