Vilnius University

Dainius Dzindzalieta

TIGHT BERNOULLI TAIL PROBABILITY BOUNDS

Doctoral Dissertation

Physical Sciences, Mathematics (01 P)

Vilnius, 2014

Doctoral dissertation was prepared at the Institute of Mathematics and Informatics of Vilnius University in 2009–2013.

Scientific Supervisors:

Habil. Dr. Vidmantas Kastytis Bentkus (Vilnius University, Physical Sciences, Mathematics — 01 P) (2009–2010)

Prof. Dr. Habil. Saulius Norvidas (Vilnius University, Physical Sciences, Mathematics — 01 P) (2010–2013)

Prof. Dr. Friedrich Götze (University of Bielefeld, Germany, Physical Sciences, Mathematics — 01 P)

Vilniaus Universitetas

Dainius Dzindzalieta

TIKSLIOSIOS BERNULIO TIKIMYBIŲ NELYGYBĖS

Daktaro disertacija

Fiziniai mokslai, matematika (01 P)

Vilnius, 2014

Disertacija rengta 2009-2013 metais Vilniaus universiteto Matematikos ir informatikos institute.

Moksliniai vadovai:

habil. dr. Vidmantas Kastytis Bentkus (Vilniaus universitetas, fiziniai mokslai, matematika — 01 P) (2009–2010)

prof. habil. dr. Saulius Norvidas (Vilniaus universitetas, fiziniai mokslai, matematika — 01 P)

Mokslinis konsultantas:

prof. dr. Friedrich Götze (Bylefeldo universitetas, Vokietija, fiziniai mokslai, matematika — 01 P)

Acknowledgements

I will be always grateful to my first Advisor late Vidmantas Kastytis Bentkus. I will be always grateful for inviting me to join the group of students in Vilnius Institute of Mathematics and Informatics, for believing in me from the first years of studies in university. I will be always thankful for his hard work teaching me not only how to solve mathematical problems, but also how to present them. I am thankful for countless discussions on mathematics, history, politics, life, while drinking coffee in his house terrace.

I would like to express my sincere gratitude to my doctoral supervisors, Professors Saulius Norvidas and Friedrich Götze, for their priceless guidance and advices.

I would also like to thank Professor Sergey Bobkov for inviting me to Minnesota Institute for Mathematics and Its Applications and valuable advices.

I am extremely grateful to Petrė Grebeničenkaitė, my school teacher, who introduced me to mathematics and is still a person I can always call when I need an advice.

I would like to thank Tomas Juškevičius and Matas Šileikis for valuable discussions and help.

I am grateful to all the people who helped me during my visits to Bielefeld, where I was invited by Friedrich Götze. Especially I would like to thank to Anita Lydia Cole and Stephan Merkes.

I will be always thankful to Sophie Liu Othmer for her hospitality during my stay in Minnesota.

I would like to thank all my colleagues in Institute of Mathematics and Informatics of Vilnius University.

Finally, I wish to thank my family for the support during my studies. I thank my daughter Rugilė for her wonderful smile.

In memory of Vidmantas Kastytis Bentkus

Contents

Notations

$\boldsymbol{a}$ denotes a real vector $(a_1, \dots, a_n)$.

$\mathbb{R}^n$ denotes the set of real vectors $\boldsymbol{a}$.

$|\boldsymbol{a}|_2$ denotes an Eucleadian norm of the vector $\boldsymbol{a}$, i.e., $|\boldsymbol{a}|_2 = (a_1^2 + \cdots + a_n^2)^{1/2}$.

$|\boldsymbol{a}|_1$ denotes an l_1 norm of the vector $\boldsymbol{a}$, i.e., $|\boldsymbol{a}|_1 = |a_1| + \cdots + |a_n|$.

$(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + \cdots + a_n b_n$ denotes a scalar product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

$N(0,1)$ denotes a standard normal random variable.

$\mathbb{P}$ denotes the probability measure.

$\mathbb{E}\, X$ denotes an expectation of a random variable X.

$M_k = X_1 + \cdots + X_k$ denotes the martingale sequence with bounded differences $X_m = M_m - M_{m-1}$.

$\mathcal{M}$ denotes the class of martingales with bounded differences.

$\mathcal{SM}$ denotes the class of super-martingales with bounded differences.

$W_n = (M_0, M_1, \dots, M_n)$ denotes a random walk based on a martingale sequence M_k.

$[x]$ denotes integer part of real number x.

$\{x\}$ denotes a fractional part of a real number x.

$\lceil x \rceil$ denotes the smallest integer number not smaller than x.

$\lfloor x \rfloor$ denotes the largest integer number not larger than x.

$\{0,1\}^n$ denotes an n-dimensional discrete cube.

$\text{sgn}(x)$ denotes a signum function which is equal to 1 for $x \geq 0$ and -1 otherwise.

LTF denotes a linear threshold function.

$\text{Inf}_i(f)$ denotes an influence of the i'th variable.

$\hat{f}(A)$ denotes a Fourier coefficient.

τ_x denotes a stopping time.

(V, d, μ) denotes a probability metric space.

M_f denotes a median of the function f.

S^{n-1} denotes the Euclidean unit sphere.

Chapter 1

Introduction

In our work we consider an isoperimetric problem

$$D(\mathcal{F}, I, n) \stackrel{\text{def}}{=} \sup_{S_n \in \mathcal{F}} \mathbb{P}\{S_n \in I\}, \tag{1.0.1}$$

where $\mathcal{F}$ is a class of sums $S_n = X_1 + \cdots + X_n$ of Bernoulli's random variables, either independent or having a martingale type dependence, a set $I \subset \mathbb{R}$ is an interval, either bounded or unbounded. We find tight upper bounds for $D(\mathcal{F}, I, n)$ and extend the results to Lipschitz functions. We say that an upper bound is tight if we can construct a sequence of random variables, for which the sup in (1.0.1) is achieved. We say that a random variable is Bernoulli's, if it takes at most two values and have a mean equal to 0.

In order to illustrate the problems, we first introduce a special case of the problem considered in Chapter 2. Other problems are similar in formulation and spirit.

Let $\mathcal{F}$ be a class of sums $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$ of symmetric weighted independent Rademacher random variables ε_i, such that $\mathbb{P}\{\varepsilon_i = \pm 1\} = 1/2$. Rademacher's random variable (r.v.) is a special case of Bernoulli's random variables. Let $I = [x, \infty)$.

In a celebrated work Hoeffding 1963 showed that the following upper bound for the tail probability holds

$$D(\mathcal{F}, I, n) = \sup_{S_n \in \mathcal{F}} \mathbb{P}\{S_n \geq x\} \leq \exp\{-x^2/2n\}, \quad x \in \mathbb{R}. \tag{1.0.2}$$

Let $W_n = \varepsilon_1 + \cdots + \varepsilon_n$. If we take $S_n = W_n$, then in view of the Central Limit Theorem we can infer that the exponential function on the right-hand side is the minimal one. Yet a certain factor of order x^{-1} is missing, since $\Phi(x) \approx (\sqrt{2\pi}x)^{-1} \exp\{-x^2/2\}$ for large x.

Furthermore, it is possible to show that the random variable S_n is subgaussian in the sense that

$$D(\mathcal{F}, I, n) \leq c\,\mathbb{P}\left\{\sqrt{n}Z \geq x\right\}, \quad x \in \mathbb{R},$$

where Z is the standard normal random variable and c is some explicit positive constant.

Although there are numerous improvements of the Hoeffding inequality, to our knowledge there are no examples where tight bound for the tail probability is found. In our work we present a class of pairs $(\mathcal{F}, I)$ for which we can give a tight bound for $D(\mathcal{F}, I, n)$.

1.1 Aims and problems

We give a short summary of the problems considered in our thesis.

In Chapter 2 we solve an isoperimetric problem (1.0.1), where $\mathcal{F}$ is a class of sums $S_n = X_1 + \cdots + X_n$ of independent symmetric random variables and $I \subset \mathbb{R}$ is an interval (bounded or unbounded). Depending on I we consider two cases of boundedness conditions $|X_i| \geq 1$ and $|X_i| \leq 1$ separately. If $I = [x-k, x+k)$ and the bound for (1.0.1) depends only on k and n, then we assume that $|X_i| \geq 1$. is a classical Littlewood-Offord type problem of the 1940's. We give a short and self-contained proof of this problem based on an induction on dimension. If $I = [x, \infty)$ or $I = \{x\}$ then we consider a case $|X_i| \leq 1$. In this case we show that a probability (1.0.1) is maximized when $k = k(x)$ random variables X_i's are Rademacher random variables taking values ± 1 with equal probabilities and others are equal to 0 with probability 1. We give an explicit description of $k(x)$.

In Chapter 3 we consider a sum of weighted independent Rademacher random variables $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$. We assume that a variance of the sum S_n is bounded by 1. Our first result of Chapter 3 is an optimal subgaussian constant. The existence of such an absolute constant was first shown in 1994. Our second result of Chapter 3 is an improvement of a Chebyshev inequality $\mathbb{P}\{S_n \geq x\} \leq \frac{1}{2x^2}$ for all $x > 1$. We present an application of the results to the Student's statistics and to self normalized sums. Unlike to previous chapter, the supremum for the tail probability $\mathbb{P}\{S_n \geq x\}$ is not maximized when all non-zero coefficients are equal to each other as was shown by A. V. Zhubr [96].

In Chapter 4 we present an application of the results from Chapter 3 to investigate single coordinate influence of Boolean valued half-space functions on the Boolean cube, i.e., functions $f : \{-1, 1\}^n \to \{-1, 1\}$ such that

$f(x_1, \ldots, x_n) = \operatorname{sgn}(a_1 x_1 + \cdots + a_n x_n)$. We reformulate the problem in probabilistic terms and obtain conditional small ball type inequality for the sum of weighted independent Rademacher random variables. As a consequence we confirm a conjecture by Matulef, O'Donnell, Rubinfeld and Servedio [62] that the threshold function associated to a linear function with some large coefficient, if balanced, must have a large influence.

In Chapter 5 we assume that Bernoulli random variables X_i's are only bounded and have a martingale type dependence, i.e., $M_k = X_1 + \cdots + X_k$ is a martingale sequence. We find tight bounds for the probability that a random walk based on a martingale sequence M_k visits an interval $[x, \infty)$. We also show that the maximizing random walk is an inhomogeneous Markov chain. We present a full description of the maximizing random walk and give explicit expression for the maximal probability. We extend the results to random walks based on supermartingale sequences. Finally we show that maximal inequalities for martingales are equivalent to inequalities for tail probabilities. As far as we know our result gives the first known tight bounds for a Hoeffding type [47] inequalities for martingales with bounded differences.

In Chapter 6 we obtain an optimal deviation from the mean upper bound $D(x) \overset{\text{def}}{=} \sup_{f \in \mathcal{F}} \mu\{f - \mathbb{E}_\mu f \geq x\}$, for $x \in \mathbb{R}$, where $\mathcal{F}$ is the complete class of integrable, Lipschitz functions on probability metric (product) spaces. As corollaries we obtain $D(x)$ for Euclidean unit sphere S^{n-1} with a geodesic distance function and a normalized Haar measure, for $\mathbb{R}^n$ equipped with a Gaussian measure and for the multidimensional cube, rectangle, torus or Diamond graph equipped with uniform measure and Hamming distance function. We also prove that in general probability metric spaces extremal Lipschitz functions are from a family of negative distance functions.

1.2 Actuality and novelty

The thesis covers a major part of author's research carried out during his PhD studies. Chapters 2–6 contains some results appearing in the papers written by the author of these thesis in collaboration with T. Juškevičius and M. Šileikis [31] (Chapter 2), V. K. Bentkus [6] (a part of Chapter 3), F. Götze (Chapter 4) [30] and by the author individually [29, 33, 32] (a part of Chapter 3, Chapter 5, Chapter 6). Three papers are already published or accepted for publication, three are submitted for publication.

Chapter 2

Sums of symmetric random variables

In this chapter we consider the problem

$$D(\mathcal{F}, I, n) = \sup_{S_n \in \mathcal{F}} \mathbb{P}\{S_n \in I\}, \tag{2.0.1}$$

where $\mathcal{F}$ is a class of sums $S_n = X_1 + \cdots + X_n$ of independent symmetric random variables and $I \subset \mathbb{R}$ is an interval (bounded or unbounded including the special case when I is singleton). Depending on I we consider two cases of boundedness conditions $|X_i| \geq 1$ and $|X_i| \leq 1$ separately. If $I = [x-k, x+k)$ and the bound for (2.0.1) depends only on x and n, then we assume that $|X_i| \geq 1$. This is a classical Littlewood-Offord type problem of the 1940's. We give a short and self-contained proof of this problem based on an induction on dimension. If $I = [x, \infty)$ or $I = \{x\}$ then we consider a case $|X_i| \leq 1$. In this case we show that a probability (2.0.1) is maximized when $k = k(x)$ random variables X_i's are Rademacher random variables and others are equal to 0 with probability 1. We give an explicit description of k.

2.1 Introduction

Let $S_n = X_1 + \cdots + X_n$ be a sum of independent random variables X_i such that

$$|X_i| \leq 1 \quad \text{and} \quad \mathbb{E}\, X_i = 0. \tag{2.1.1}$$

Let $W_n = \varepsilon_1 + \cdots + \varepsilon_n$ be the sum of independent Rademacher random variables, i.e., such that $\mathbb{P}\{\varepsilon_i = \pm 1\} = 1/2$. We will refer to W_n as a simple random walk with n steps.

By a classical result of Hoeffding [47] we have the estimate

$$\mathbb{P}\left\{S_n \geq x\right\} \leq \exp\left\{-x^2/2n\right\}, \quad x \in \mathbb{R}. \tag{2.1.2}$$

If we take $S_n = W_n$ on the left-hand side of (2.1.2), then in view of the Central Limit Theorem we can infer that the exponential function on the right-hand side is the minimal one. Yet a certain factor of order x^{-1} is missing, since $\Phi(x) \approx (\sqrt{2\pi}x)^{-1} \exp\left\{-x^2/2\right\}$ for large x.

Furthermore, it is possible to show that the random variable S_n is subgaussian in the sense that

$$\mathbb{P}\left\{S_n \geq x\right\} \leq c\,\mathbb{P}\left\{\sqrt{n}Z \geq x\right\}, \quad x \in \mathbb{R},$$

where Z is the standard normal random variable and c is some explicit positive constant (see, for instance, [12]).

Perhaps the best upper bound for $\mathbb{P}\left\{S_n \geq x\right\}$ was given by Bentkus [10], which for integer x is optimal for martingales with differences X_i satisfying (2.1.1).

Although there are numerous improvements of the Hoeffding inequality, to our knowledge there are no examples where the exact bound for the tail probability is found. In this chapter we give an optimal bound for the tail probability $\mathbb{P}\left\{S_n \geq x\right\}$ under the additional assumption of symmetry.

We henceforth reserve the notation S_n and W_n for random walks with symmetric steps satisfying (2.1.1) and a simple random walk with n steps respectively.

Theorem 1 (Dzindzalieta, Juškevičius, Šileikis [31]). *For $x > 0$ we have*

$$\mathbb{P}\left\{S_n \geq x\right\} \leq \begin{cases} \mathbb{P}\left\{W_n \geq x\right\} & \text{if } \lceil x \rceil + n \in 2\mathbb{Z}, \\ \mathbb{P}\left\{W_{n-1} \geq x\right\} & \text{if } \lceil x \rceil + n \in 2\mathbb{Z}+1. \end{cases} \tag{2.1.3}$$

The latter inequality can be interpreted by saying that among bounded random walks the simple random walk is the "most stochastic".

Kwapień (see [88]) proved that for arbitrary independent symmetric random variables X_i and real numbers a_i with absolute value less than 1 we have

$$\mathbb{P}\left\{a_1X_1 + \ldots + a_nX_n \geq x\right\} \leq 2\,\mathbb{P}\left\{X_1 + \ldots + X_n \geq x\right\}, \quad x > 0. \tag{2.1.4}$$

In fact, Kwapień's inequality holds for X_i's in arbitrary Banach space. The case $n = 2$ with $X_i = \varepsilon_i$ shows that the constant 2 in (2.1.4) cannot be improved.

Theorem 1 improves Kwapień's inequality for Rademacher sequences. We believe that the inequality (2.1.3) with some conditioning arguments leads to better estimates for arbitrary symmetric random variables X_i under the assumptions of Kwapień's inequality, but we will not go into these details in this work.

We also consider the problem of finding the quantity

$$\sup_{S_n} \mathbb{P}\left\{S_n = x\right\},$$

which can be viewed as a non-uniform bound for the concentration of the random walk S_n at a point x.

Theorem 2 (Dzindzalieta, Juškevičius, Šileikis [31]). *For $x > 0$ and $k = \lceil x \rceil$ we have*

$$\mathbb{P}\left\{S_n = x\right\} \leq \mathbb{P}\left\{W_m = k\right\}, \tag{2.1.5}$$

where

$$m = \begin{cases} \min\left\{n, k^2\right\}, & \text{if } n + k \in 2\mathbb{Z}, \\ \min\left\{n-1, k^2\right\}, & \text{if } n + k \in 2\mathbb{Z} + 1. \end{cases}$$

Equality in (2.1.5) is attained for $S_n = \frac{x}{k} W_m$.

We give two different proofs for both inequalities. The first approach is based on induction on the number of random variables (§2.2). To prove Theorem 2 we also need the solution of the Littlewood-Offord problem.

Theorem 3 (Erdös [35]). *Let $a_1, \ldots, a_n$ be real numbers such that $|a_i| \geq 1$. We have*

$$\max_{x \in \mathbb{R}} \mathbb{P}\left\{S_n \in (x-k, x+k]\right\} \leq \mathbb{P}\left\{W_n \in (-k, k]\right\}.$$

That is, the number of the choices of signs for which S_n lies in an interval of length $2k$ does not exceed the sum of k largest binomial coefficients in n.

Theorem 3 was first proved by Erdős [35] using Sperner's Theorem. We give a very short solution which seems to be shorter than the original proof by Erdős. We only use induction on n and do not use Sperner's Theorem.

Surprisingly, Theorems 1 and 2 can also be proved by applying results from extremal combinatorics (Chapter 2.3). Namely, we use the bounds for the size of intersecting families of sets (hypergraphs) by Katona [51] and Milner [66].

Using a strengthening of Katona's result by Kleitman [53], we extend Theorem 1 to odd 1-Lipschitz functions rather than just sums of the random variables X_i (§2.4). It is important to note that the bound of Theorem

1 cannot be true for all Lipschitz functions since the extremal case is not provided by odd functions. We give the description of the extremal Lipschitz functions defined on general probability metric spaces in Chapter 6.

2.2 Proofs by induction on dimension

We will first show that it is enough to prove Theorems 1 and 2 in the case when S_n is a linear combination of independent Rademacher random variables ε_i with coefficients $|a_i| \le 1$.

Lemma 4 (Dzindzalieta, Juškevičius, Šileikis [31])**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a bounded measurable function. Then we have*

$$\sup_{X_1,\dots,X_n} \mathbb{E}\, g(X_1,\dots,X_n) = \sup_{a_1,\dots,a_n} \mathbb{E}\, g(a_1\varepsilon_1,\dots a_n\varepsilon_n),$$

where the supremum on the left-hand side is taken over symmetric independent random variables $X_1,\dots,X_n$ such that $|X_i| \le 1$ and the supremum on the right-hand side is taken over numbers $-1 \le a_1,\dots,a_n \le 1$.

Proof. Define $S = \sup_{a_1,\dots,a_n} \mathbb{E}\, g(a_1\varepsilon_1,\dots a_n\varepsilon_n)$. Clearly

$$S \le \sup_{X_1,\dots,X_n} \mathbb{E}\, g(X_1,\dots,X_n).$$

By symmetry of $X_1,\dots,X_n$, we have

$$\mathbb{E}\, g(X_1,\dots,X_n) = \mathbb{E}\, g(X_1\varepsilon_1,\dots,X_n\varepsilon_n).$$

Therefore

$$\mathbb{E}\, g(X_1,\dots,X_n) = \mathbb{E}\,\mathbb{E}\left[g(X_1\varepsilon_1,\dots,X_n\varepsilon_n)\,|\,X_1,\dots,X_n\right] \le \mathbb{E}\, S = S. \qquad \square$$

Thus, in view of Lemma 4 we will henceforth write S_n for $a_1\varepsilon_1+\dots+a_n\varepsilon_n$ instead of a sum of arbitrary symmetric random variables X_i.

Proof of Theorem 1. First note that the inequality is true for $x \in (0,1]$ and all n. This is due to the fact that $\mathbb{P}\{S_n \ge x\} \le 1/2$ by symmetry of S_n and for all n the right-hand side of the inequality is given by the tail of an odd number of random signs, which is exactly $1/2$. We can also assume that the largest coefficient $a_i = 1$ as otherwise if we scale the sum by a_i then the tail of the this new sum would be at least as large as the former. We will thus assume, without loss of generality, that $0 \le a_1 \le a_2 \le \ldots \le a_n = 1$. Define a function $\mathbb{I}(x,n)$ to be 1 if $\lceil x\rceil + n$ is even, and zero otherwise. Then we can rewrite the right-hand side of (2.1.3) as

$$\mathbb{P}\left\{W_{n-1} + \varepsilon_n \mathbb{I}(x,n) \geq x\right\},$$

making an agreement $\varepsilon_0 \equiv 0$.

For $x > 1$ we argue by induction on n. Case $n = 0$ is trivial. Observing that $\mathbb{I}(x-1,n) = \mathbb{I}(x+1,n) = \mathbb{I}(x,n+1)$ we have

$$\begin{aligned}
\mathbb{P}\left\{S_{n+1} \geq x\right\} &= \tfrac{1}{2}\mathbb{P}\left\{S_n \geq x-1\right\} + \tfrac{1}{2}\mathbb{P}\left\{S_n \geq x+1\right\} \\
&\leq \tfrac{1}{2}\mathbb{P}\left\{W_{n-1} + \varepsilon_n \mathbb{I}(x-1,n) \geq x-1\right\} \\
&+ \tfrac{1}{2}\mathbb{P}\left\{W_{n-1} + \varepsilon_n \mathbb{I}(x+1,n) \geq x+1\right\} \\
&= \mathbb{P}\left\{W_n + \varepsilon_{n+1} \mathbb{I}(x,n+1) \geq x\right\}. \qquad \square
\end{aligned}$$

Proof of Theorem 3. We can assume that $a_1 \geq a_2 \geq \ldots \geq a_n \geq 1$. Without loss of generality we can also take $a_n = 1$. This is because

$$\begin{aligned}
\mathbb{P}\left\{S_n \in (x-k, x+k]\right\} &\leq \mathbb{P}\left\{S_n/a_n \in (x-k, x+k]/a_n\right\} \\
&\leq \max_{x\in\mathbb{R}} \mathbb{P}\left\{S_n/a_n \in (x-k, x+k]\right\}.
\end{aligned}$$

The claim is trivial for $n = 0$. Let us assume that we have proved the statement for $1, 2, ..., n-1$. Then

$$\begin{aligned}
&\mathbb{P}\left\{S_n \in (x-k, x+k]\right\} \\
=&\tfrac{1}{2}\mathbb{P}\left\{S_{n-1} \in (x-k-1, x+k-1]\right\} + \tfrac{1}{2}\mathbb{P}\left\{S_{n-1} \in (x-k+1, x+k+1]\right\} \\
=&\tfrac{1}{2}\mathbb{P}\left\{S_{n-1} \in (x-k-1, x+k+1]\right\} + \tfrac{1}{2}\mathbb{P}\left\{S_{n-1} \in (x-k+1, x+k-1]\right\} \\
\leq&\tfrac{1}{2}\mathbb{P}\left\{W_{n-1} \in (-k-1, k+1]\right\} + \tfrac{1}{2}\mathbb{P}\left\{W_{n-1} \in (-k+1, k-1]\right\} \\
=&\tfrac{1}{2}\mathbb{P}\left\{W_{n-1} \in (-k-1, k-1]\right\} + \tfrac{1}{2}\mathbb{P}\left\{W_{n-1} \in (-k+1, k+1]\right\} \\
=&\mathbb{P}\left\{W_n \in (-k, k]\right\}. \qquad \square
\end{aligned}$$

Note that we rearranged the intervals after the second equality so as to have two intervals of different lengths and this makes the proof work.

Before proving Theorem 2, we will obtain an upper bound for $\mathbb{P}\left\{S_n = x\right\}$ under an additional condition that all a_i are nonzero.

Lemma 5 (Dzindzalieta, Juškevičius, Šileikis [31]). *Let $x > 0$, $k = \lceil x \rceil$. Suppose that $0 < a_1 \leq \cdots \leq a_n \leq 1$. Then*

$$\mathbb{P}\left\{S_n = x\right\} \leq \begin{cases} \mathbb{P}\left\{W_n = k\right\}, & \text{if} \quad n+k \in 2\mathbb{Z}, \\ \mathbb{P}\left\{W_{n-1} = k\right\}, & \text{if} \quad n+k \in 2\mathbb{Z}+1. \end{cases} \tag{2.2.1}$$

Proof. We first prove the lemma for $x \in (0,1]$ and any n. By Theorem 3 we have

$$\mathbb{P}\left\{S_n = x\right\} \le 2^{-n}\binom{n}{\lceil n/2 \rceil}. \tag{2.2.2}$$

On the other hand, if $x \in (0,1]$, then $k = 1$ and

$$2^{-n}\binom{n}{\lceil n/2 \rceil} = \begin{cases} 2^{-n}\binom{n}{(n+1)/2} = \mathbb{P}\left\{W_n = 1\right\}, & \text{if} \quad n+1 \in 2\mathbb{Z}, \\ 2^{-n}\binom{n}{n/2} = \mathbb{P}\left\{W_{n-1} = 1\right\}, & \text{if} \quad n+1 \in 2\mathbb{Z}+1, \end{cases}$$

where the second equality follows by Pascal's identity:

$$2^{-n}\binom{n}{n/2} = 2^{-n}\left[\binom{n-1}{n/2} + \binom{n-1}{n/2-1}\right] = 2^{1-n}\binom{n-1}{n/2} = \mathbb{P}\left\{W_{n-1} = 1\right\}.$$

Let $\mathbb{N} = \{1, 2, \dots\}$ stand for the set of positive integers. Let us write $B_n(x)$ for the right-hand side of (2.2.1). Note that it has the following properties:

$$x \mapsto B_n(x) \text{ is non-increasing;} \tag{2.2.3}$$

$$x \mapsto B_n(x) \text{ is constant on each of the intervals } (k-1, k], \quad k \in \mathbb{N}; \tag{2.2.4}$$

$$B_n(k) = \tfrac{1}{2}B_{n-1}(k-1) + \tfrac{1}{2}B_{n-1}(k+1), \quad \text{if } k = 2, 3, \dots. \tag{2.2.5}$$

We proceed by induction on n. The case $n = 1$ is trivial. To prove the induction step for $n \ge 2$, we consider two cases: (i) $x = k \in \mathbb{N}$; (ii) $k - 1 < x < k \in \mathbb{N}$.

Case (i). For $k = 1$ the lemma has been proved, so we assume that $k \ge 2$. By the inductional hypothesis we have

$$\begin{aligned} \mathbb{P}\left\{S_n = k\right\} &= \tfrac{1}{2}\mathbb{P}\left\{S_{n-1} = k - a_n\right\} + \tfrac{1}{2}\mathbb{P}\left\{S_{n-1} = k + a_n\right\} \\ &\le \tfrac{1}{2}B_{n-1}(k - a_n) + \tfrac{1}{2}B_{n-1}(k + a_n). \end{aligned} \tag{2.2.6}$$

By (2.2.3) we have

$$B_{n-1}(k - a_n) \le B_{n-1}(k-1), \tag{2.2.7}$$

and by (2.2.4) we have

$$B_{n-1}(k + a_n) = B_{n-1}(k+1). \tag{2.2.8}$$

Combining (2.2.6), (2.2.7), (2.2.8), and (2.2.5), we obtain

$$\mathbb{P}\left\{S_n = k\right\} \le B_n(k). \tag{2.2.9}$$

Case (ii). For $x \in (0, 1]$ the lemma has been proved, so we assume $k \geq 2$. Consider two cases: (iii) $x/a_n \geq k$; (iv) $x/a_n < k$.

Case (iii). Define $S'_n = a'_1\varepsilon_1 + \cdots + a'_n\varepsilon_n$, where $a'_i = ka_i/x$, so that $S'_n = \frac{k}{x}S_n$. Recall that $a_n = \max_i a_i$, by the hypothesis of the lemma. Then $a'_i \leq ka_n/x$ and the assumption $x/a_n \geq k$ imply that $0 < a'_1, \ldots, a'_n \leq 1$. Therefore, by (2.2.9) and (2.2.4) we have

$$\mathbb{P}\{S_n = x\} = \mathbb{P}\{S'_n = k\} \leq B_n(k) = B_n(x).$$

Case (iv). Without loss of generality, we can assume that $a_n = 1$, since

$$\mathbb{P}\{S_n = x\} = \mathbb{P}\left\{\frac{a_1}{a_n}\varepsilon_1 + \cdots + \frac{a_n}{a_n}\varepsilon_n = \frac{x}{a_n}\right\}$$

and $k - 1 < x/a_n < k$, by the assumption of the present case. Sequentially applying the induction hypothesis, (2.2.4), (2.2.5), and again (2.2.4), we get

$$\begin{aligned}
\mathbb{P}\{S_n = x\} &= \tfrac{1}{2}\mathbb{P}\{S_{n-1} = x - 1\} + \tfrac{1}{2}\mathbb{P}\{S_{n-1} = x + 1\} \\
&\leq \tfrac{1}{2}B_{n-1}(x - 1) + \tfrac{1}{2}B_{n-1}(x + 1) \\
&= \tfrac{1}{2}B_{n-1}(k - 1) + \tfrac{1}{2}B_{n-1}(k + 1) \\
&= B_n(k) = B_n(x). \qquad \square
\end{aligned}$$

Proof of Theorem 2. Writing $B_n(k)$ for the right-hand side of (2.2.1), we have, by Lemma 5, that

$$\mathbb{P}\{S_n = x\} \leq \max_{j=k}^{n} B_j(k).$$

If $j + k \in 2\mathbb{Z}$, then $B_j(k) = \mathbb{P}\{W_j = k\} = B_{j+1}(k)$ and therefore

$$\max_{j=k}^{n} B_j(k) = \max_{\substack{k \leq j \leq n \\ k+j \in 2\mathbb{Z}}} \mathbb{P}\{W_j = k\}. \tag{2.2.10}$$

To finish the proof, note that the sequence $\mathbb{P}\{W_j = k\} = 2^{-j}\binom{j}{(k+j)/2}$, $j = k, k + 2, k + 4, \ldots$ is unimodal with a peak at $j = k^2$, i.e.,

$$\mathbb{P}\{W_{j-2} = k\} \leq \mathbb{P}\{W_j = k\}, \quad \text{if} \quad j \leq k^2,$$

and

$$\mathbb{P}\{W_{j-2} = k\} > \mathbb{P}\{W_j = k\}, \quad \text{if} \quad j > k^2.$$

Indeed, elementary calculations yield that the inequality

$$2^{-j+2}\binom{j-2}{(k+j)/2-1} \le 2^{-j}\binom{j}{(k+j)/2}, \qquad j \ge k+2,$$

is equivalent to the inequality $j \le k^2$. $\square$

2.3 Proofs based on results in extremal combinatorics

Let $[n]$ stand for the finite set $\{1, 2, \dots, n\}$. Consider a family $\mathcal{F}$ of subsets of $[n]$. We denote by $|\mathcal{F}|$ the cardinality of $\mathcal{F}$. The family $\mathcal{F}$ is called:

1. k**-intersecting** if for all $A, B \in \mathcal{F}$ we have $|A \cap B| \ge k$.
2. an **antichain** if for all $A, B \in \mathcal{F}$ we have $A \not\subseteq B$.

A well known result by Katona [51] (see also [23, p. 98, Theorem 4]) gives the exact upper bound for a k-intersecting family.

Theorem 6 (Katona [51]). *If $k \ge 1$ and $\mathcal{F}$ is a k-intersecting family of subsets of $[n]$ then*

$$|\mathcal{F}| \le \begin{cases} \sum_{j=t}^{n}\binom{n}{j}, & \text{if } k+n=2t, \\ \sum_{j=t}^{n}\binom{n}{j} + \binom{n-1}{t-1}, & \text{if } k+n=2t-1. \end{cases} \tag{2.3.1}$$

Notice that if $k+n=2t$, then

$$\sum_{j=t}^{n}\binom{n}{j} = 2^n \mathbb{P}\left\{W_n \ge k\right\}. \tag{2.3.2}$$

If $k+n=2t-1$, then using the Pascal's identity $\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}$ we get

$$\sum_{j=t}^{n}\binom{n}{j} + \binom{n-1}{t-1} = 2\sum_{j=t-1}^{n-1}\binom{n-1}{j} = 2^n \mathbb{P}\left\{W_{n-1} \ge k\right\}. \tag{2.3.3}$$

The exact upper bound for the size of a k-intersecting antichain is given by the following result of Milner [66].

Theorem 7 (Milner [66]). *If a family $\mathcal{F}$ of subsets of $[n]$ is a k-intersecting antichain, then*

$$|\mathcal{F}| \le \binom{n}{t}, \qquad t = \left\lceil \frac{n+k}{2} \right\rceil. \tag{2.3.4}$$

Note that we have

$$\binom{n}{t} = 2^n \mathbb{P}\left\{W_n = k\right\}, \quad \text{if} \quad n+k = 2t, \tag{2.3.5}$$

and

$$\binom{n}{t} = 2^n \mathbb{P}\left\{W_n = k+1\right\}, \quad \text{if} \quad n+k = 2t-1. \tag{2.3.6}$$

By Lemma 4 it is enough to prove Theorems 1 and 2 for the sums

$$S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n,$$

where $0 \le a_1, \ldots, a_n \le 1$. Denote as A^c the complement of the set A. For each $A \subset [n]$, write $s_A = \sum_{i\in A} a_i - \sum_{i\in A^c} a_i$. We define two families of sets:

$$\mathcal{F}_{\ge x} = \{A \subset [n] : s_A \ge x\}, \quad \text{and} \quad \mathcal{F}_x = \{A \subset [n] : s_A = x\}.$$

Proof of Theorem 1. We have

$$\mathbb{P}\left\{S_n \ge x\right\} = 2^{-n}|\mathcal{F}_{\ge x}|.$$

Let $k = \lceil x \rceil$. Since W_n takes only integer values, we have

$$\mathbb{P}\left\{W_n \ge k\right\} = \mathbb{P}\left\{W_n \ge x\right\} \qquad \text{and} \qquad \mathbb{P}\left\{W_{n-1} \ge k\right\} = \mathbb{P}\left\{W_{n-1} \ge x\right\}.$$

Therefore, in the view of (2.3.1), (2.3.2), and (2.3.3), it is enough to prove that $\mathcal{F}_{\ge x}$ is k-intersecting. Suppose that there are $A, B \in \mathcal{F}_{\ge x}$ such that $|A \cap B| \le k-1$. Writing $\sigma_A = \sum_{i\in A} a_i$, we have

$$s_A = \sigma_A - \sigma_{A^c} = (\sigma_{A\cap B} - \sigma_{A^c\cap B^c}) + (\sigma_{A\cap B^c} - \sigma_{A^c\cap B}) \tag{2.3.7}$$

and

$$s_B = \sigma_B - \sigma_{B^c} = (\sigma_{A\cap B} - \sigma_{A^c\cap B^c}) - (\sigma_{A\cap B^c} - \sigma_{A^c\cap B}). \tag{2.3.8}$$

Since

$$\sigma_{A\cap B} - \sigma_{A^c\cap B^c} \le \sigma_{A\cap B} \le |A\cap B| \le k-1 < x,$$

from (2.3.7) and (2.3.8) we get

$$\min\{s_A, s_B\} < x,$$

which contradicts the fact $s_A, s_B \geq x$. $\square$

The following lemma implies Theorem 2. It also gives the optimal bound for $\mathbb{P}\{S_n = x\}$ and thus improves Lemma 5.

Lemma 8 (Dzindzalieta, Juškevičius, Šileikis [31]). *Let $0 < a_1, \ldots, a_n \leq 1$ be strictly positive numbers, $x > 0$, $k = \lceil x \rceil$. Then*

$$\mathbb{P}\{S_n = x\} \leq \begin{cases} \mathbb{P}\{W_n = k\}, & \text{if} \quad n + k \in 2\mathbb{Z}, \\ \mathbb{P}\{W_n = k+1\}, & \text{if} \quad n + k \in 2\mathbb{Z}+1. \end{cases}$$

Proof. We have

$$\mathbb{P}\{S_n = x\} = 2^{-n}|\mathcal{F}_x|.$$

In the view of (2.3.4), (2.3.5), and (2.3.6), it is enough to prove that $\mathcal{F}_x$ is a k-intersecting antichain. To see that $\mathcal{F}_x$ is k-intersecting it is enough to note that $\mathcal{F}_x \subset \mathcal{F}_{\geq x}$. To show that $\mathcal{F}_x$ is an antichain is even easier. If $A, B \in \mathcal{F}_x$ and $A \subsetneq B$, then $s_B - s_A = 2\sum_{i \in B \setminus A} a_i > 0$, which contradicts the assumption that $s_B = s_A = x$. $\square$

Proof of Theorem 2. Lemma 8 gives

$$\mathbb{P}\{S_n = x\} \leq \max_{j=k}^{n} \mathbb{P}\{W_j = k + 1 - \mathbb{I}(k, j)\},$$

where again $\mathbb{I}(k,j) = \mathbb{I}\{k + j \in 2\mathbb{Z}\}$. Note that if $k + j \in 2\mathbb{Z}$ we have

$$\begin{aligned} \mathbb{P}\{W_j = k\} &\geq 1/2\mathbb{P}\{W_j = k\} + 1/2\mathbb{P}\{W_j = k+2\} \\ &= \mathbb{P}\{W_{j+1} = k+1\}, \qquad k > 0. \end{aligned}$$

Hence

$$\max_{j=k}^{n} \mathbb{P}\{W_j = k + 1 - \mathbb{I}(k,j)\} = \max_{\substack{k \leq j \leq n \\ k+j \in 2\mathbb{Z}}} \mathbb{P}\{W_j = k\},$$

the right-hand side being the same as the one of (2.2.10). Therefore, repeating the argument following (2.2.10) we are done. $\square$

2.4 Extension to Lipschitz functions

One can extend Theorem 1 to odd Lipschitz functions taken on n independent random variables. Consider the cube $C_n = [-1, 1]^n$ with the ℓ^1 metric d. We say that a function $f : C_n \to \mathbb{R}$ is K-Lipschitz with $K > 0$ if

$$|f(x) - f(y)| \le Kd(x, y), \qquad x, y \in C_n. \tag{2.4.1}$$

We say that a function $f : C_n \to \mathbb{R}$ is odd if $f(-x) = -f(x)$ for all $x \in C_n$. An example of an odd 1-Lipschitz function is the function mapping a vector to the sum of its coordinates:

$$f(x_1, \dots, x_n) = x_1 + \dots + x_n.$$

Note that the left-hand side of (2.1.3) can be written as $\mathbb{P}\{f(X_1, \dots, X_n) \ge x\}$.

As in Theorems 1 and 2, the crux of the proof is dealing with two-valued random variables. The optimal bound for a k-intersecting family is not sufficient for this case, therefore we use the following generalization of Theorem 6 due to Kleitman [53] (see also [23, p. 102]) which we state slightly reformulated for our convenience. Let us define the diameter of a set family $\mathcal{F}$ by $\operatorname{diam} \mathcal{F} = \max_{A,B\in\mathcal{F}} |A \bigtriangleup B|$.

Theorem 9 (Kleitman [53])**.** *If $k \ge 1$ and $\mathcal{F}$ is a family of subsets of $[n]$ with $\operatorname{diam} \mathcal{F} \le n - k$, then*

$$|\mathcal{F}| \le \begin{cases} \displaystyle\sum_{j=t}^{n} \binom{n}{j}, & \text{if } \ k + n = 2t, \\ \displaystyle\sum_{j=t}^{n} \binom{n}{j} + \binom{n-1}{t-1}, & \text{if } \ k + n = 2t - 1. \end{cases} \tag{2.4.2}$$

To see that Theorem 9 implies Theorem 6, observe that $|A \cap B| \ge k$ implies $|A \bigtriangleup B| \le n - k$.

Theorem 10 (Dzindzalieta, Juškevičius, Šileikis [31])**.** *Suppose that a function $f : C_n \to \mathbb{R}$ is 1-Lipschitz and odd. Let $X_1, \dots, X_n$ be symmetric independent random variables such that $|X_i| \le 1$. Then, for $x > 0$, we have that*

$$\mathbb{P}\{f(X_1, \dots, X_n) \ge x\} \le \begin{cases} \mathbb{P}\{W_n \ge x\}, & \text{if } \ n + \lceil x \rceil \in 2\mathbb{Z}, \\ \mathbb{P}\{W_{n-1} \ge x\}, & \text{if } \ n + \lceil x \rceil \in 2\mathbb{Z} + 1. \end{cases} \tag{2.4.3}$$

Proof. Applying Lemma 4 with the function

$$g(y_1, \dots, y_n) = \mathbb{I}\{f(y_1, \dots, y_n) \ge x\},$$

we can see that it is enough to prove (2.4.3) with

$$X_1 = a_1\varepsilon_1, \ldots, X_n = a_n\varepsilon_n$$

for any 1-Lipschitz odd function f. In fact, we can assume that $a_1 = \cdots = a_n = 1$, since the function

$$(x_1, \ldots, x_n) \mapsto f(a_1x_1, \ldots, a_nx_n)$$

is clearly 1-Lipschitz and odd.

Given $A \subseteq [n]$, write f_A for $f(2\,\mathbb{I}_A(1) - 1, \ldots, 2\,\mathbb{I}_A(n) - 1)$, where $\mathbb{I}_A$ is the indicator function of the set A. Note that

$$|f_A - f_B| \leq 2|A \bigtriangleup B| \tag{2.4.4}$$

by the Lipschitz property. Consider the family of finite sets

$$\mathcal{F} = \{A \subseteq [n] : f_A \geq x\},$$

so that

$$\mathbb{P}\{f(\varepsilon_1, \ldots, \varepsilon_n) \geq x\} = 2^{-n}|\mathcal{F}|.$$

Write $k = \lceil x \rceil$. Note that W_{n-1} and W_n take only integer values. Therefore by (2.3.2) and (2.3.3) we see that the right-hand side of (2.4.2) is equal, up to the power of two, to the right-hand side of (2.4.3). Consequently, if $\operatorname{diam} \mathcal{F} \leq n - k$, then Theorem 9 implies (2.4.3). Therefore, it remains to check that for any $A, B \in \mathcal{F}$ we have $|A \bigtriangleup B| \leq n - k$.

Suppose that for some A, B we have $f_A, f_B \geq x$ but $|A \bigtriangleup B| \geq n - k + 1$. Then

$$|A \bigtriangleup B^c| = |(A \bigtriangleup B)^c| = n - |A \bigtriangleup B| \leq k - 1,$$

and hence by (2.4.4) we have

$$|f_A - f_{B^c}| \leq 2k - 2. \tag{2.4.5}$$

On the other hand we have that $f_{B^c} \leq -x$, as f is odd. Therefore

$$f_A - f_{B^c} \geq 2x > 2k - 2,$$

which contradicts (2.4.5). $\square$

Chapter 3

Weighted sum of Rademacher random variables

In this chapter we consider a sum of weighted independent Rademacher random variables $S_n = a_1\varepsilon_1 + \cdots + a_n\varepsilon_n$. We assume that a variance of S_n is bounded by 1. Our first result of this chapter is an optimal constant in the inequality $\mathbb{P}\{S_n \geq x\} \leq c\mathbb{P}\{\eta \geq x\}$, where $\eta \sim N(0,1)$ is a standard normal random variable. The existance of such an absolute constant was first shown in 1994. Our second result of this chapter is an improvement of a Chebyshev inequality $\mathbb{P}\{S_n \geq x\} \leq \frac{1}{2x^2}$ for $x \in (1, \sqrt{2})$. We provide an application of the results to the Student's statistics and to self normalized sums. Unlike to previous chapter, the supremum for the tail probability $\mathbb{P}\{S_n \geq x\}$ is not maximized when all non-zero coefficients are equal to each other as was shown by A. V. Zhubr [96].

3.1 Introduction and results

Let $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ be independent identically distributed Rademacher random variables, so that $\mathbb{P}\{\varepsilon = -1\} = \mathbb{P}\{\varepsilon = 1\} = 1/2$. Let $a = (a_1, a_2, \ldots)$ be a (weight) sequence of non-random real numbers. Write

$$S = a_1\varepsilon_1 + a_2\varepsilon_2 + \ldots.$$

Henceforth we assume that a has the l_2 norm $|a|_2 = (a_1^2 + a_2^2 + \ldots)^{1/2}$ bounded from above by 1, i.e., $|a|_2 \leq 1$. Furthermore, without loss of generality we assume that a is a non-increasing sequence of non-negative numbers,

$$a_1 \geq a_2 \geq a_2 \geq \cdots \geq 0.$$

It is well known that the random variable S is sub-gaussian, that is, there exists an absolute positive constant c such that

$$\mathbb{P}\{S \geq x\} \leq c\mathbb{P}\{\eta \geq x\} \tag{3.1.1}$$

for all $x \in \mathbb{R}$, where η is a standard normal random variable. The main result of the chapter is the following theorem.

Theorem 11 (Bentkus, Dzindzalieta [6]). *Let $\eta \sim N(0,1)$ be a standard normal random variable, then we have,*

$$\mathbb{P}\{S_n \geq x\} \leq c\mathbb{P}\{\eta \geq x\} \quad \textit{for all } x \in \mathbb{R} \tag{3.1.2}$$

with the constant c equal to

$$c_* := (4\mathbb{P}\{\eta \geq \sqrt{2}\})^{-1} \approx 3.178.$$

The value $c = c_*$ is the best possible in the sense that (11) becomes equality if $n \geq 2$, $S_n = (\varepsilon_1 + \varepsilon_2)/\sqrt{2}$ and $x = \sqrt{2}$. Let $\Phi(x) = \mathbb{P}\{\eta \leq x\}$ be the standard normal distribution function, and $I(x) = 1 - \Phi(x)$ be the standard normal survival function. Using the definition of $I(x)$, we can reformulate our result as

$$\mathbb{P}\{S \geq x\} \leq c_* I(x). \tag{3.1.3}$$

Inequalities of type (3.1.1) and (3.1.3) are related to the geometry of the Euclidean space $\mathbb{R}^n$. Let $C^n = [-1,1]^n$ be n-dimensional cube, and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ a unit vector, $|a|_2 = 1$. Then the half-space

$$\{z \in \mathbb{R}^n : \ (z,a) \geq x\}$$

contains at most $2^n c_* I(x)$ vertices of the cube C^n, and the bound $2^n c_* I(x)$ is the best possible among bounds expressed via Gaussian survival functions (the scalar product $(z,a) = z_1 a_1 + \cdots + z_n a_n$).

The value $c = c_*$ is the best possible since (3.1.3) turns to an equality if $S = (\varepsilon_1 + \varepsilon_2)/\sqrt{2}$ and $x = \sqrt{2}$. Indeed, in this special case

$$\mathbb{P}\{S \geq \sqrt{2}\} = 1/4 \equiv c_*\, I(\sqrt{2}).$$

For $x < \sqrt{2}$ inequality (3.1.3) is strict due to the following simple bounds

$$\begin{aligned}
\mathbb{P}\{S \geq x\} &\leq 1 && \text{for } x \leq 0, && (3.1.4)\\
\mathbb{P}\{S \geq x\} &\leq 1/2 && \text{for } 0 < x \leq 1, && (3.1.5)\\
\mathbb{P}\{S \geq x\} &\leq 1/(2x^2) && \text{for } x > 0. && (3.1.6)
\end{aligned}$$

Inspecting our proofs it is clear that (3.1.3) is strict for $x > \sqrt{2}$ as well. Altogether, (3.1.3) turns to equality if and only if $x = \sqrt{2}$ and $S = (\varepsilon_1 + \varepsilon_2)/\sqrt{2}$.

For $x \leq 1$ inequalities (3.1.4) and (3.1.5) are optimal (recall that we assume that $|a|_2 \leq 1$). Holzman and Kleitman [48] established a remarkable inequality

$$\mathbb{P}\{S > 1\} \leq \frac{1}{4} + \frac{1}{16}. \tag{3.1.7}$$

Together with the rather rough Chebyshev type inequality (3.1.6), this is all what is known for $1 < x \leq \sqrt{2}$. We prove the following result

Theorem 12 (Dzindzalieta [29]).

$$\mathbb{P}\{S \geq x\} \leq \frac{1}{4} + \frac{1}{8}\left(1 - \sqrt{2 - \frac{2}{x^2}}\right) \qquad \textit{for } x \in (1, \sqrt{2}). \tag{3.1.8}$$

For $1 < x < x_0$ with $x_0 = \sqrt{8/7} = 1.069\ldots$ the Holzman–Kleitman bound (3.1.7) is better than (3.1.8). For $x_0 < x < \sqrt{2}$ inequality (3.1.8) is better than (3.1.7) and (3.1.6). Inequality (3.1.8) can be improved, however an advantage of (3.1.8) is that it has a simple proof.

There is a number of open questions related to estimation of the function $\mathbb{P}\{S \geq x\}$, e.g., see [41, 48, 3, 45] and references therein, an important unsolved problems is to prove (or disprove) that

$$\mathbb{P}\{|S| \leq 1\} \geq 1/2, \tag{3.1.9}$$

and [25, 46, 72, 93]

$$\mathbb{P}\{|S| \geq 1\} \geq \frac{7}{64}, \tag{3.1.10}$$

An equivalent formulation of (3.1.9) is

$$\mathbb{P}\{S > 1\} \leq 1/4.$$

Assuming that $a_1 < 1$, Holzman and Kleitman [48] proved $\mathbb{P}\{|S| < 1\} \geq 3/8$, which is equivalent to $\mathbb{P}\{S \geq 1\} \leq 1/4 + 1/16$. Using an alternative approach, Ben-Tal et al [3] established the bound $\mathbb{P}\{|S| \leq 1\} \geq 1/3$, which is equivalent to $\mathbb{P}\{S > 1\} \leq 1/3$. Applying Lyapunov type bounds for the remainder term in the Central Limit Theorem with explicit absolute constant, say c_{L} (we introduce this bound in the section of proofs), the stronger inequality $\mathbb{P}\{S \geq 1\} \leq 1/4$ (hence, (3.1.9) as well) holds if $|a|_2 = 1$ and $a_1 \leq C$ with

$$C \stackrel{\text{def}}{=} (1 - 4I(1))/(4c_{\mathrm{L}}). \tag{3.1.11}$$

Using the best known bound $c_{\mathrm{L}} \leq 0.56$, we have $C \geq 0.1631\dots$. Furthermore, inequality $\mathbb{P}\{S > 1\} \leq 1/4$ holds if $S = (\varepsilon_1 + \cdots + \varepsilon_n)/\sqrt{n}$.

Using upper bounds for $\mathbb{P}\{S \geq x\}$ one can estimate the concentration of S around its mean. Namely, if $\mathbb{P}\{S \geq x\} \leq B(x)$ then $\mathbb{P}\{|S| < x\} \geq 1 - 2B(x)$. It follows that

$$\mathbb{P}\{|S| < x\} \geq \frac{1}{2} - \frac{1}{4}\left(1 - \sqrt{2 - \frac{2}{x^2}}\right) \qquad \text{for } 1 \leq x \leq \sqrt{2}, \tag{3.1.12}$$

and

$$\mathbb{P}\{|S| < x\} \geq 1 - 2c_*\, I(x) \qquad \text{for } x \geq \sqrt{2}.$$

We show that the conjecture (3.1.9) is equivalent to the following powerful concentration inequality

$$\mathbb{P}\{|S| \leq \delta\} \geq \mathbb{P}\{|S| \geq 1/\delta\} \qquad \text{for all } 0 \leq \delta \leq 1, \tag{3.1.13}$$

and/or to

$$\mathbb{P}\{|S| < \delta\} \geq \mathbb{P}\{|S| > 1/\delta\} \qquad \text{for all } 0 \leq \delta \leq 1, \tag{3.1.14}$$

where we define $1/0 = \infty$ and $\mathbb{P}\{|S| \geq \infty\} = 0$. Note, that (3.1.13)–(3.1.14) works well for all possible a_i's, that is, they cover both localized and delocalized cases. In some special cases it can be obtained using large deviation inequalities, but in general it can not be achieved from large deviation inequalities. In contrary, we could apply (3.1.13)–(3.1.14) to get large deviation results.

Proposition 13 (Dzindzalieta [29]). *Conjectured bound* (3.1.9) *is equivalent to* (3.1.13) *and/or to* (3.1.14).

Proof. Indeed, if (3.1.9) holds, then it holds for $S' = \frac{2\delta^2}{1+\delta^2} S + \frac{1-\delta^2}{1+\delta^2}\varepsilon'$ with any $\delta \in (0, 1]$ as well. Now integrating $\mathbb{P}\{S' \geq 1\}$ with respect to ε' we get that $\mathbb{P}\{S \geq \delta\} + \mathbb{P}\{S \geq 1/\delta\} \leq 1/2$. This is equivalent to $\mathbb{P}\{|S| \geq \delta\} + \mathbb{P}\{|S| \geq 1/\delta\} \leq 1$ or $\mathbb{P}\{|S| < \delta\} \geq \mathbb{P}\{|S| \geq 1/\delta\}$. Since $\mathbb{P}\{|S| < \delta\} \leq \mathbb{P}\{|S| \leq \delta\}$ and $\mathbb{P}\{|S| \geq 1/\delta\} \geq \mathbb{P}\{|S| > 1/\delta\}$ we get inequalities (3.1.13)–(3.1.14).

If (3.1.13) and/or (3.1.14) holds, we take $\delta = 1$ to get the bound (3.1.9). □

Inequalities 3.1.13 and (3.1.14) relate the concentration around zero to large deviations or the size of the support of S to concentration around zero. For example, (3.1.13) implies that for any unit vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, such that $|a|_2 = 1$, the strip

$$\{z \in \mathbb{R}^n : \ |(z, a)| \le \lambda\}$$

contains at least one vertex of the cube $[-1, 1]^n$, where

$$\lambda = \frac{1}{|a|_1}, \qquad |a|_1 = |a_1| + \cdots + |a_n|.$$

Another interpretation: among all 2^n sums

$$\pm a_1 \pm \cdots \pm a_n$$

at least one lies in the interval $[-\lambda, \lambda]$.

Write

$$Z_0 = 0, \qquad Z_k = \frac{\varepsilon_1 + \cdots + \varepsilon_k}{\sqrt{k}} \quad \text{for } k \ge 1,$$

and define the functions

$$M_n(x) = \max_{0 \le k \le n} \mathbb{P}\left\{Z_k \ge x\right\}, \qquad M(x) = \max_{k \ge 0} \mathbb{P}\left\{Z_k \ge x\right\} \equiv \max_{n \ge 1} M_n(x), \tag{3.1.15}$$

and

$$B_n(x) = \sup_{S_n} \mathbb{P}\left\{S_n \ge x\right\}, \qquad B(x) = \sup_{S} \mathbb{P}\left\{S \ge x\right\}, \tag{3.1.16}$$

where $\sup_{S_n}$ is taken over all $S_n = a_1\varepsilon_1 + \cdots + \varepsilon_n a_n$ such that $|a|_2 \le 1$, respectively $\sup_S$ is taken over all S such that $|a|_2 \le 1$.

In early 90's Sergey Bobkov asked whether

$$B_n(x) = M_n(x), \qquad B(x) = M(x). \tag{3.1.17}$$

It is clear that $B_n(x) \ge M_n(x)$ and $B(x) \ge M(x) \ge I(x)$. Conjecture (3.1.17) is much stronger than conjecture (3.1.9).

Conjecture (3.1.17) were disproved by Zhubr [96] using geometric interpretation of the problem. A. V. Zhubr gave counterexamples showing that the problem is much more complicated than expected. In [96] counterexamples shows directly that a weaker conjecture (3.1.9) is not true. Using Berry-Esseen bound (see discussion above the (3.1.11)) and counterexamples given by A. V. Zhubr [96] we see that the stronger version of the conjecture is false as well.

A standard application of inequalities of type (3.1.1) is to Student's statistic and to self-normalized sums. For example, if random variables $X_1, \ldots, X_n$ are independent, symmetric and not all identical zero, then the statistic

$$T = (X_1 + \cdots + X_n) / \sqrt{X_1^2 + \cdots + X_n^2}$$

is sub-gaussian,

$$\mathbb{P}\{T \geq x\} \leq c_* \, \mathbb{P}\{\eta \geq x\}, \tag{3.1.18}$$

and this inequality is optimal since (3.1.18) turns to an equality if $n = 2$ and $X_1 = \varepsilon_1$, $X_2 = \varepsilon_2$.

Let us describe the scheme of the proof of (3.1.3). For $x \geq \sqrt{3}$ analysis of $\mathbb{P}\{S \geq x\}$ is based on rather simple applications of the inequality

$$I(A) + I(B) \leq 2I(x), \quad A = \frac{x - \tau}{\vartheta}, \quad B = \frac{x + \tau}{\vartheta}, \quad \vartheta = \sqrt{1 - \tau^2}, \tag{3.1.19}$$

which holds in the region $\{0 \leq \tau \leq 1,\ x \geq \sqrt{3}\}$. Inequality (3.1.19) is a very special case of more general inequalities established in [12]. For $x \leq \sqrt{2}$, inequalities (3.1.4)–(3.1.6) suffice to prove (3.1.3). The most difficult is the case $\sqrt{2} < x \leq \sqrt{3}$. Surprisingly, this seemingly simple problem requires an indeed complicated proof.

Inequalities of type (3.1.1) are of considerable interest in probability, function theory and functional analysis, combinatorics, optimization, operations research etc, see e.g. [4, 3, 27, 28, 45]

Using exponential functions as upper bounds for indicator functions (or the so called Bernstein method), the inequality

$$\mathbb{P}\{S \geq x\} \leq \exp\{-x^2/2\}, \qquad x \geq 0, \tag{3.1.20}$$

is contained in Hoeffding 1963 [47], among others. Using Theorem 11, one can improve (3.1.20) to

$$\mathbb{P}\{S \geq x\} \leq c_2 \exp\{-x^2/2\}, \quad c_2 = 0.824\ldots, \quad \text{for } x > 0.$$

One of the related inequalities to our results is Khinchin's inequality. For a review of this type of inequalities and for other developments, see e.g. [54, 57, 73, 74, 94]. In the paper Latała [56] provided bounds on moments and tails of Gaussian chaoses. For general chaoses Berry-Esseen type bounds were obtained in [68].

Eaton [34] conjectured that

$$\mathbb{P}\{S \geq x\} \leq c_{\text{ep}} \exp\{-x^2/2\}/x, \qquad x \geq 0, \qquad c_{\text{ep}} = 4.46\ldots.$$

Pinelis [77] improved the form of Eaton's conjecture and established (3.1.1) with $c = c_{\text{ep}}$. He developed the so called moment comparison method (see [78, 76]). The method consists of proving the moment type inequalities $\mathbb{E}\,(S-h)_+^3 \leq \mathbb{E}\,(\eta-h)_+^3$ for $h \in \mathbb{R}$, where $x_+ = \max\{0, x\}$, and showing that these moment inequalities imply (3.1.1). It seems that the value $c = c_{\text{ep}} = 4.46\ldots$ is the best possible which can be obtained using this method. In the context of probabilities of large deviations, Bentkus 1986–2007[9, 7, 10, 5, 11, 13, 12, 8] (henceforth B 1986–2007) developed induction based methods. If it is possible to overcome related technical difficulties, these methods lead to the most tight known upper bounds for the tail probabilities. The paper [12] (without attempts to optimize the constants) contains the bound $c \leq (6I(\sqrt{3})^{-1} = 4.003\ldots$, as well as the lower bound $c \geq c_*$, and a conjecture that the optimal constant is c_*. Bobkov et al [18] using induction proved that $c \leq 12.01$. Pinelis [79] established the bound $c \leq c_* + 0.04$ (cf. the methods used in this chapter with those of B 2001–2007). Questions related to the conjecture (3.1.9) are considered in Derinkuyu and Pınar [27, 28], Nemirovski [69], He and Luo and Nie and Zhang [44], Ben-Tal and Nemirovski [4], So [86], among others. Summarizing, upper bounds for $\mathbb{P}\{S \geq x\}$ influence the quality of semi-definite programming algorithms, and there is a demand for explicit (non-asymptotic!) and as precise as possible such bounds.

3.2 Proofs

In this section we use the following notation

$$\tau = a_1, \quad \vartheta = \sqrt{1-\tau^2}, \quad I(x) = \mathbb{P}\{\eta \geq x\}, \quad \varphi(x) = -I'(x), \tag{3.2.1}$$

that is, $I(x)$ is the tail probability for standard normal random variable η and $\varphi(x)$ is the standard normal density. Without loss of generality we assume that $a_1^2 + \cdots + a_n^2 = 1$ and $a_1 \geq \cdots \geq a_n \geq 0$. Using (3.2.1) we have $S_n = \tau\varepsilon_1 + \vartheta X$ with $X = (a_2\varepsilon_2 + \cdots + a_n\varepsilon_n)/\vartheta$. The random variable X is symmetric and independent of ε_1. It is easy to check that $\mathbb{E}\,X^2 = 1$ and

$$\mathbb{P}\{S_n \geq x\} \;=\; \frac{1}{2}\mathbb{P}\{X \geq A\} + \frac{1}{2}\mathbb{P}\{X \geq B\}, \tag{3.2.2}$$

where $A = \frac{x-\tau}{\vartheta}$ and $B = \frac{x+\tau}{\vartheta}$.

We start with a simple Chebyshev type inequality.

Lemma 14 (Bentkus, Dzindzalieta [6])**.** *Let $s > 0$ and $0 \le a \le b$, then for any random variable Y we have*

$$a^s \mathbb{P}\{|Y| \ge a\} + (b^s - a^s)\mathbb{P}\{|Y| \ge b\} \le \mathbb{E}\,|Y|^s. \tag{3.2.3}$$

If Y is symmetric, then

$$a^s \mathbb{P}\{Y \ge a\} + (b^s - a^s)\mathbb{P}\{Y \ge b\} \le \mathbb{E}\,|Y|^s/2. \tag{3.2.4}$$

Proof. It is clear that (3.2.3) implies (3.2.4). To prove (3.2.3) we use the obvious inequality

$$a^s \mathbb{I}\{|Y| \ge a\} + (b^s - a^s)\mathbb{I}\{|Y| \ge b\} \le |Y|^s, \tag{3.2.5}$$

where $\mathbb{I}\{E\}$ stands for the indicator function of the event E. Taking expectation, we get (3.2.3). $\square$

Similarly to (3.2.3), one can derive a number of inequalities stronger than the standard Chebyshev inequality $\mathbb{P}\{S_n \ge x\} \le 1/(2x^2)$. For example, instead of $\mathbb{P}\{S_n \ge 1\} \le 1/2$ we have the much stronger

$$\mathbb{P}\{S_n \ge 1\} + \mathbb{P}\{S_n \ge \sqrt{2}\} + \mathbb{P}\{S_n \ge \sqrt{3}\} + \cdots \le 1/2.$$

We will make use of Lyapunov type bounds with explicit constants for the remainder term in the Central Limit Theorem. Let $X_1, X_2, \ldots$ be independent random variables such that $\mathbb{E}\,X_j = 0$ for all j. Denote $\beta_j = \mathbb{E}\,|X_j|^3$. Assume that the sum $Z = X_1 + X_2 + \ldots$ has unit variance. Then there exists an absolute constant, say c_{L}, such that

$$|\mathbb{P}\{Z \ge x\} - I(x)| \le c_{\mathrm{L}}(\beta_1 + \beta_2 + \ldots). \tag{3.2.6}$$

It is known that $c_{\mathrm{L}} \le 0.56\ldots$ [92, 84]. Note that we actually do not need the best known bound for c_L. Even $c_L = 0.958$ suffices to prove Theorem 11.

Replacing X_j by $a_j\varepsilon_j$ and using $\beta_j \le \tau a_j^2$ for all j, the inequality (3.2.6) implies

$$|\mathbb{P}\{S_n \ge x\} - I(x)| \le c_{\mathrm{L}}\tau. \tag{3.2.7}$$

Proof of Theorem 11. For $x \le \sqrt{2}$ Theorem 11 follows from the symmetry of S_n and Chebyshev's inequality (first it was implicitely shown in [12], later in [79]). In the case $x \geq \sqrt{2}$ we argue by induction on n. However, let us first provide a proof of Theorem 11 in some special cases where induction fails.

Using the bound (3.2.7), let us prove Theorem 11 under the assumption that

$$\tau \leq \tau_{\mathrm{L}} \stackrel{\mathrm{def}}{=} (c_* - 1) I(\sqrt{3})/c_{\mathrm{L}} \quad \text{and} \quad x \leq \sqrt{3}. \tag{3.2.8}$$

Using $c_{\mathrm{L}} = 0.56$, the numerical value of τ_{L} is $0.16\ldots$. In order to prove Theorem 11 under the assumption (3.2.8), note that the inequality (3.2.7) yields

$$\mathbb{P}\{S_n \geq x\} \leq I(x) + \tau c_{\mathrm{L}}. \tag{3.2.9}$$

If the inequality (3.2.8) holds, the right hand side of (3.2.9) is clearly bounded from above by $c_* \, I(x)$ for $x \leq \sqrt{3}$.

For x and τ such that (3.2.8) does not hold we use induction on n. If $n = 1$ then we have $S_n = \varepsilon_1$ and Theorem 11 is equivalent to the trivial inequality $1/2 \leq c_* I(1)$.

Let us assume that Theorem 11 holds for $n \leq k-1$ and prove it for $n = k$.

Firstly we consider the case $x \geq \sqrt{3}$. We replace S_n in (3.2.2) by S_k with $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$. We can estimate the latter two probabilities in (3.2.2) applying the induction hypothesis $\mathbb{P}\{X \geq y\} \leq c_* \, I(y)$. We get

$$\mathbb{P}\{S_k \geq x\} \leq c_* \, I(A)/2 + c_* \, I(B)/2. \tag{3.2.10}$$

In order to conclude the proof, it suffices to show that the right hand side of (3.2.10) is bounded from above by $c_* \, I(x)$, that is, that the inequality $I(A) + I(B) \leq 2I(x)$ holds. As $x \geq \sqrt{3}$ it follows by the inequality (3.1.19).

In the remaining part of the proof we can assume that $x \in (\sqrt{2}, \sqrt{3})$ and $\tau \geq \tau_{\mathrm{L}}$. In this case in order to prove Theorem 11 we have to improve the arguments used to estimate the right hand side of (3.2.2). This is achieved applying Chebyshev type inequalities of Lemma 14. By Lemma 14, for any symmetric X such that $\mathbb{E}\, X^2 = 1$, and $0 \leq A \leq B$, we have

$$A^2 \mathbb{P}\{X \geq A\} + (B^2 - A^2)\mathbb{P}\{X \geq B\} \leq 1/2. \tag{3.2.11}$$

By (3.2.1), we can rewrite (3.2.11) as

$$(x - \tau)^2 \mathbb{P}\{X \geq A\} + 4x\tau \mathbb{P}\{X \geq B\} \leq \vartheta^2/2. \tag{3.2.12}$$

For $x \in (\sqrt{2}, \sqrt{3})$ and $\tau \geq \tau_{\mathrm{L}}$ we consider the cases

$$\text{i) } (x-\tau)^2 \geq 4x\tau \quad \text{and} \quad \text{ii) } (x-\tau)^2 \leq 4x\tau$$

separately. We denote the sets of points (x, τ) such that $x \in (\sqrt{2}, \sqrt{3})$, $\tau \geq \tau_{\mathrm{L}}$ and (i) or (ii) holds by E_1 and E_2, respectively.

i) Using (3.2.2), (3.2.12) and the induction hypothesis we get

$$\mathbb{P}\{S_k \geq x\} \leq \frac{D\mathbb{P}\{X \geq B\} + \vartheta^2/2}{2(x-\tau)^2} \leq \frac{c_* \, DI(B) + \vartheta^2/2}{2(x-\tau)^2}, \tag{3.2.13}$$

where $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$ and $D = (x-\tau)^2 - 4x\tau$.

In order to finish the proof of Theorem 11 (in this case) it suffices to show that the right hand side of (3.2.13) is bounded above by $c_* \, I(x)$. In other words, we have to check that the function

$$f(x, \tau) \equiv f \ \stackrel{\text{def}}{=} \ \Big((x-\tau)^2 - 4x\tau\Big) \, c_* \, I(B) - 2c_* \, (x-\tau)^2 I(x) + \vartheta^2/2, \tag{3.2.14}$$

is negative on E_1, where $B = (x+\tau)/\vartheta$.

By Lemma 18 we have

$$f(x, \tau) \leq f(\sqrt{3}, \tau) =: g(\tau). \tag{3.2.15}$$

Since $\tau \leq (3 - 2\sqrt{2})x$ the inequality $f \leq 0$ on E_1 follows from Lemma 15, below.

ii) Using (3.2.2), (3.2.12) and induction hypothesis we get

$$\mathbb{P}\{S_k \geq x\} \leq \frac{C\mathbb{P}\{X \geq A\} + \vartheta^2/2}{8x\tau} \leq \frac{C/(2A^2) + \vartheta^2/2}{8x\tau}, \tag{3.2.16}$$

where $X = (a_2\varepsilon_2 + \cdots + a_k\varepsilon_k)/\vartheta$ and $C = 4x\tau - (x-\tau)^2$.

In order to finish the proof (in this case) it suffices to show that the right hand side of (3.2.16) is bounded from above by $c_* \, I(x)$. In other words, we have to check that

$$C/(2A^2) + \vartheta^2/2 \leq 8x\tau c_* \, I(x) \quad \text{on } E_2. \tag{3.2.17}$$

Recalling that $C = 4x\tau - (x-\tau)^2$, $\quad A = (x-\tau)/\vartheta$, inequality (3.2.17) is equivalent to

$$h(x, \tau) = h \ \stackrel{\text{def}}{=} \ \frac{1-\tau^2}{(x-\tau)^2} - 4c_* \, I(x) \leq 0 \quad \text{on } E_2. \tag{3.2.18}$$

Inequality (3.2.18) follows from Lemma 19, below. The proof of Theorem 11 is complete. □

Lemma 15. *The function g defined by (3.2.15) is negative for all $\tau \in [\tau_L, (3-2\sqrt{2})\sqrt{3}]$.*

Lemma 16. *$I'(B) \geq \vartheta I'(x)$ on E_1.*

Lemma 17. *$I(B) \geq I(x) + I'(x)\tau$ on E_1.*

Lemma 18. *The partial derivative $\partial_x f$ of the function f defined by (3.2.14) is positive on E_1.*

Lemma 19. *The function h defined by (3.2.18) is negative on E_2.*

Proof of Lemma 15. Since $g(\tau_{\mathrm{L}}) < 0$ it is sufficient to show that g is a decreasing function for $\tau_{\mathrm{L}} \leq \tau \leq (3-2\sqrt{2})\sqrt{3}$. Note that

$$g(\tau) \;=\; \left((\sqrt{3}-\tau)^2 - 4\sqrt{3}\tau\right) c_*\, I(B) + (1-\tau^2)/2 - 2c_*\,(\sqrt{3}-\tau)^2 I(\sqrt{3}),$$

We have

$$\begin{aligned} g'(\tau) = \quad & (2\tau - 6\sqrt{3})c_*\, I(B) - \left((\sqrt{3}-\tau)^2 - 4\sqrt{3}\tau\right) c_*\, \varphi(B)(1+\tau\sqrt{3})\vartheta^{-3} \\ & -\tau + 4c_*\,(\sqrt{3}-\tau)I(\sqrt{3}), \end{aligned}$$

where φ is the standard normal distribution. Hence

$$g'(\tau) \leq w(\tau) \;\stackrel{\text{def}}{=}\; (2\tau - 6\sqrt{3})c_*\, I(B) - \tau + 4c_*\,(\sqrt{3}-\tau)I(\sqrt{3}).$$

Note that the value of B in previous three displayed formulas should also be computed with $x = \sqrt{3}$. Using Lemma 17 we get

$$g'(\tau) \leq -2c_*\,(\sqrt{3}+\tau)I(\sqrt{3}) + 2c_*\,\tau(3\sqrt{3}-\tau)\varphi(\sqrt{3}) - \tau \;\stackrel{\text{def}}{=}\; Q(\tau)$$

with

$$Q(\tau) = -\alpha\tau^2 + \beta\tau - \gamma, \quad \alpha = 0.56\ldots, \quad \beta = 1.67\ldots, \quad \gamma = 0.45\ldots.$$

Clearly, Q is negative on the interval $[\tau_{\mathrm{L}}, (3-2\sqrt{2})\sqrt{3}]$. It follows that g' is negative, and g is decreasing on $[\tau_{\mathrm{L}}, (3-2\sqrt{2})\sqrt{3}]$. □

Proof of Lemma 16. Since $I' = -\varphi$ by (3.2.1), the inequality $I'(B) \geq \vartheta I'(x)$ is equivalent to

$$u(\tau) \;\stackrel{\text{def}}{=}\; (1-\tau^2)\exp\left\{\frac{(x+\tau)^2}{1-\tau^2} - x^2\right\} \geq 1.$$

Since $u(0) = 0$, it suffices to check that $u' \geq 0$. Elementary calculations show that $u' \geq 0$ is equivalent to the trivial inequality $x + \tau^2 x + \tau x^2 + \tau^3 \geq 0$. $\square$

Proof of Lemma 17. Set now $g(\tau) = I(B)$. Then the inequality

$$I(B) \geq I(x) + I'(x)\tau$$

turns into $g(\tau) \geq g(0) + g'(0)\tau$. The latter inequality holds provided that $g''(\tau) \geq 0$. Next, it is easy to see that

$$g'(\tau) = -\varphi(B)B' \quad \text{and} \quad g''(\tau) = (BB'^2 - B'')\varphi(B).$$

Hence, to verify that $g''(\tau) \geq 0$ we verify that $BB'^2 - B'' \geq 0$. This last inequality is equivalent to $-2 + 2x^2 + x^3\tau + x^2\tau^2 + x\tau + 2x\tau^3 + 3\tau^2 \geq 0$, which holds since $x \geq 1$. The proof of Lemma 17 is complete. $\square$

Proof of Lemma 18. We have

$$\partial_x f = 2\,(x - 3\tau)\, c_*\, I(B) + Dc_*\, I'(B)/\vartheta - 4c_*\,(x - \tau)I(x) - 2c_*\,(x - \tau)^2 I'(x).$$

We have to show that $\partial_x f \geq 0$ on E_1. Using Lemma 16, we can reduce this to the inequality

$$2\,(x - 3\tau)\, I(B) - (x + \tau)^2 I'(x) - 4(x - \tau)I(x) \geq 0. \tag{3.2.19}$$

On E_1 we have that $0 \leq \tau \leq (3 - 2\sqrt{2})x$, so $x - 3\tau \geq x - 3(3 - 2\sqrt{2})x = (6\sqrt{2} - 8)x > 0$. By Lemma 17 we have that l.h.s. of (3.2.19) is bigger than

$$\begin{aligned}&2(x - 3\tau)(I(x) + I'(x)\tau) - (x + \tau)I'(x) - 4(x - \tau)I(x) = \\ &-2(x + \tau)I(x) - (x^2 + 7\tau^2)I'(x).\end{aligned}$$

Inequality (3.2.19) follows by the inequality $-(x^2 + 7\tau^2)I'(x) \geq \alpha x(x + \tau)\varphi(x) > 2(x + \tau)I(x)$ on E_1 with $\alpha = 4\sqrt{14} - 14$, where the second inequality follows from the fact that $\varphi(x)x/I(x)$ increases for $x > 0$ and is larger than $2/\alpha$ for $x = \sqrt{2}$. The proof of Lemma 18 is complete.$\square$

Proof of Lemma 19. It is easy to see that the function h attains its maximal value at $\tau = 1/x$. Hence, it suffices to check (3.2.18) with $\tau = 1/x$, that is, that for $\sqrt{2} \leq x \leq \sqrt{3}$ the inequality $g(x) \overset{\text{def}}{=} 1 - 4c_*(x^2 - 1)I(x) \leq 0$ holds. Using $4c_*\, I(\sqrt{2}) = 1$, we have $g(\sqrt{2}) = 0$ and $g(\sqrt{3}) < 0$. Next, $g'(x) = -8c_* x I(x) + 4c_*(x^2 - 1)\varphi(x)$, so $g'(\sqrt{2}) < 0$ and $g'(\sqrt{3}) > 0$. We have that $g''(x) = 4c_*\,((5 - x^2)x\varphi(x) - 2I(x))$. Since $I(x) \leq \varphi(x)/x$ we have that $g''(x) \geq 4c_*\,((5 - x^2)x\varphi(x) - 2\varphi(x)/x) = 4c_*\varphi(x)/x((5 - x^2)x^2 - 2) \geq 8c_*\varphi(x)/x > 0$ for $x \in (\sqrt{2}, \sqrt{3})$. The proof of Lemma 19 is complete.$\square$

Proof of Theorem 12. We denote

$$g(x) \quad \stackrel{\text{def}}{=} \quad \frac{3}{8} - \frac{1}{8}\sqrt{2 - \frac{2}{x^2}}.$$

We consider the cases

$$\text{i) } 0 \le \tau \le (3 - 2\sqrt{2})\sqrt{2} \approx 0.243, \qquad \text{ii) } (3 - 2\sqrt{2})\sqrt{2} \le \tau \le 1$$

separately.

(i) Applying the Lyapunov bound (3.2.7), it is sufficient to show that

$$I(x) + c_{\mathrm{L}}\tau \le g(x).$$

In view of condition (i) this inequality is equivalent to

$$g(x) - I(x) \ge 0.243 c_{\mathrm{L}} \approx 0.136 \quad \text{for } x \in (1, \sqrt{2}).$$

Simple algebraic manipulations shows that $g''(x)$ has a sign of $3x^2 - 2$, so g is convex on interval $(1, \sqrt{2})$. Thus

$$g(x) \ge g(1.3) + g'(1.3)(x - 1.3) =: l_1(x)$$

on interval $(1, \sqrt{2})$. The function I is also convex on $(1, \sqrt{2})$, so

$$I(x) \ge \frac{I(1) - I(\sqrt{2})}{1 - \sqrt{2}}(x - 1) + I(1) =: l_2(x).$$

Since $l_1(x) - l_2(x)$ is increasing and $l_1(1) - l_2(1) \approx 0.141 > 0.136$ we derive a result. The proof in the case (i) is completed.

(ii) Elementary transformations show that in this case we have an inequality $B^2 \ge 2A^2$. Applying (3.2.2), we have

$$\mathbb{P}\{S \ge x\} = \frac{1}{2}\mathbb{P}\{X \ge A\} + \frac{1}{2}\mathbb{P}\{X \ge B\} = \mathbb{E}\, w(X), \tag{3.2.20}$$

where

$$X = (a_2\varepsilon_2 + a_3\varepsilon_3 + \dots)/\vartheta, \qquad w(t) = \frac{1}{2}\mathbb{I}\{t \ge A\} + \frac{1}{2}\mathbb{I}\{t \ge B\}.$$

In view of $B^2 \geq 2A^2$, the quadratic function

$$p(t) = c_0 + c_2 t^2, \quad c_0 = \frac{B^2 - 2A^2}{2(B^2 - A^2)}, \quad c_2 = \frac{1}{2(B^2 - A^2)}$$

is positive and satisfies $p(t) \geq w(|t|)$ for all t. Using $\mathbb{E}\, X^2 = 1$ and symmetry of X, we have

$$\begin{aligned} \mathbb{E}\, w(X) &= \frac{1}{2}\,\mathbb{E}\, w(|X|) \leq \frac{1}{2}\,\mathbb{E}\, p(|X|) = \frac{1}{2}\,\mathbb{E}\, p(X) = \frac{c_0 + c_2}{2} \\ &\equiv \frac{1}{4}\left(1 + \frac{1 - A^2}{B^2 - A^2}\right) \end{aligned} \tag{3.2.21}$$

Elementary transformations lead to

$$u(\tau) \overset{\text{def}}{=} \frac{1 - A^2}{B^2 - A^2} \equiv \frac{1 - 2\tau^2 - x^2 + 2\tau x}{4\tau x}.$$

It is easy to check that $u'(\tau)$ has the sign of $x^2 - 2\tau^2 - 1$. Hence, the stationary point of u is

$$\tau_s = \sqrt{\frac{x^2 - 1}{2}}.$$

Since u is an increasing function of $\tau \in [0, \tau_s]$, and u decreases for $\tau \in [\tau_s, 1]$, we derive

$$u(\tau) \leq u(\tau_s) \equiv \frac{x - \sqrt{2x^2 - 2}}{2x}. \tag{3.2.22}$$

Combining (3.2.20)–(3.2.22) we finish the proof of Theorem 12. $\square$

Chapter 4

Conditional tail probabilities

In this chapter we provide an application of the results from Chapter 3 for investigation of single coordinate influence of Boolean valued half-space functions of the form $f(x) = \operatorname{sgn}(a \cdot x - \theta)$ on the Boolean cube. We reformulate the problem in probabilistic terms and obtain conditional small ball inequality for the sum of weighted independent Rademacher's random variables. As a consequence we confirm a conjecture by Matulef, O'Donnell, Rubinfeld and Servedio [62] that the threshold function associated to a linear function with some large coefficient, if balanced, must have a large influence.

4.1 Introduction

In this chapter we investigate the *single coordinate influence* of Boolean valued half-space functions on the Boolean cube $\{-1, 1\}^n$, i.e. functions of the form $f(x) = \operatorname{sgn}(a \cdot x - \theta)$, with $a = (a_1, \ldots, a_n) \in \mathbb{R}^n, x \in \{-1, 1\}^n$, $a \cdot x = a_1x_1 + \cdots + a_nx_n$ and $\theta \in \mathbb{R}$. Here $\operatorname{sgn}(x)$ is equal to 1 for $x \geq 0$ and -1 otherwise. Without loss of generality, we assume everywhere that $a_1 \geq \cdots \geq a_n \geq 0$. Half-space functions are often called linear threshold functions thus for brevity we will refer to them as LTFs.

Although LTFs are simple functions, they played an important role in complexity theory, machine learning (e.g., computational learning theory) and optimization (see [17, 26, 42, 62, 67, 70, 95]).

In computational learning theory an important goal is to construct a test which, given access to an unknown function, say f, decides whether f belongs to some class, say $\mathcal{C}$, of functions or not. The class $\mathcal{C}$ is characterized by a certain number of properties. The test will check whether f has (with a given probability) these properties or f is "far" from the class $\mathcal{C}$. Such property testing procedures have been proposed in [39, 81] among others.

Recently Matulef, O'Donnell, Rubinfeld and Servedio [62] created a test which checks whether an unknown function belongs to the class of LTFs. In order to describe the structural properties of this class they considered the single coordinate influence of LTFs. The influence of a single variable is a well studied quantity in computer sciences since the late 80's [2, 49].

Since the quality of the test depends on non-asymptotic information on the influence, we not only present bounds using generic absolute constants, but give explicit values as well which are not far from the optimal ones. Furthermore, we reformulate the problem in probabilistic terms showing that it may be treated as a new class of conditional small ball inequalities for sums of weighted independent Rademacher's random variables.

In this chapter we shall need the notion of single variable influence only. (For more general notions and properties of influence functions we refer to e.g. [55, 71]). Let $\mathbb{P}$ denote the uniform probability measure on a discrete cube $\{-1,1\}^n$, i.e., $\mathbb{P}\{x\} = 1/2^n$ for all $x \in \{-1,1\}^n$. Given a Boolean-valued function $f : \{-1,1\}^n \to \{-1,1\}$ and $i \in \{1,\dots,n\}$, the influence of the i'th variable is defined as

$$\mathrm{Inf}_i(f) = \mathbb{P}\{f(x^{i+}) \neq f(x^{i-})\},$$

where x^{i+} and x^{i-} stands for the vector x with 1 and -1 in i'th coordinate, respectively. Equivalently, $\mathrm{Inf}_i(f) = 2^{-n}\#\{x : f(x^{i+}) \neq f(x^{i-})\}$. In other words, $\mathrm{Inf}_i(f)$ denotes the probability that changing a sign of the i'th coordinate of a randomly chosen x the function f changes it's sign.

The set of functions $f : \{-1,1\}^n \to \{-1,1\}$ is a 2^n-dimensional space over the reals with inner product given by $(f,g) = \mathbb{E}\, fg$. The set of functions $\{x_A\}_{A\subseteq[n]}$ defined by $\{x_A\} = \Pi_{i\in A}\, x_i$ forms a complete orthonormal basis in this space. Given a function $f : \{-1,1\}^n \to \{-1,1\}$ we define it's Fourier coefficients by $\hat{f}(A) = \mathbb{E}\, f x_A$. Thus we have $f = \sum_A \hat{f}(A) x_A$. A function f is said to be *unate* if it is monotone increasing or monotone decreasing as a function of the variable x_i for each i. In particular, LTF's are unate. It is well known that if f is unate, then $\mathrm{Inf}_i(f) = |\hat{f}(i)|$ (we write i instead of $\{i\}$). We will say that $f : \{-1,1\}^n \to \{-1,1\}$ is τ*-regular* if $|\hat{f}(i)| \leq \tau$ for all i. For this and more information on LTFs we refer the reader to, e.g., [63, 71].

Our main result of this chapter is the following theorem which proves a conjecture by Matulef, O'Donnell, Rubinfeld and Servedio [62].

Theorem 20 (Dzindzalieta, Götze [30])**.** *Let $f(x) = \mathrm{sgn}(a\cdot x - \theta)$ be a LTF. Assume that $\|a\|_2 = 1$, $a_1 \geq \cdots \geq a_n \geq 0$ and $\|a\|_\infty \geq \delta$ for some $0 < \delta < 1$ and $|\,\mathbb{E}\, f| \leq 1-\epsilon$ for some $0 \leq \epsilon \leq 1$. Then there exists an absolute constant*

$C > 0$ such that

$$|\hat{f}(1)| = \mathrm{Inf}_1(f) \geq C\,\delta\,\epsilon. \tag{4.1.1}$$

We prove that we can take $C \geq 3\sqrt{2}/64 \approx 0.066$. Furthermore, a simple example shows that an optimal C is not larger than $3\sqrt{2}/8$.

Note, that (4.1.1) do depend on θ. The dependence on θ is hidden in the inequality $|\,\mathbb{E}\, f| \leq 1 - \epsilon$. This can be clearly seen in the probabilistic reformulation of the result stated below.

In [36] it was shown that if f is LTF and $|a_i| \geq |a_j|$ then $\mathrm{Inf}_i(f) \geq \mathrm{Inf}_j(f)$, meaning that the variable with a largest weight is the most influential. We will provide a short proof of this result in Section 4.3. Since $a_1 \geq \cdots \geq a_n \geq 0$, it follows that $\mathrm{Inf}_1(f) \geq \mathrm{Inf}_i(f)$ for all i.

We need some additional definitions. Let φ denote the probability density of a standard normal random variable, i.e., $\varphi(t) = \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\}$ and introduce $\mu(\theta) = -1 + 2\int_\theta^\infty \varphi(t)\,dt$ as strictly decreasing map from $\mathbb{R}$ to $(-1, 1)$ as well as $W(\nu) = 2\varphi(\mu^{-1}(\nu))$.

Corollary 21. *Let f denote τ-regular LTF for some sufficiently small τ given by $f(x) = \mathrm{sgn}(a \cdot x - \theta)$. Then*

$$\left| \sum_{i=1}^{n} \hat{f}(i)^2 - W(\,\mathbb{E}\, f) \right| \leq \tau^{1/3}.$$

Let f_k, $k = 1, 2$, denote the τ-regular LTF's as above with $\theta = \theta_k$ and the same a. Then

$$\left| \left(\sum_{i=1}^{n} \hat{f}_1(i)\hat{f}_2(i) \right)^2 - W(\,\mathbb{E}\, f_1)W(\,\mathbb{E}\, f_2) \right| \leq \tau^{1/3}. \tag{4.1.2}$$

To prove Corollary 21 it is enough to use the bound (4.1.1) in the proof of [62, Theorem 48].

4.2 Probabilistic reformulation

In this section we will show that Theorem 20 is closely related to the results of the Chapter 3 of these thesis for sums of weighted independent Rademacher's random variables taking two values ± 1 with probabilities $1/2$. Since $\{-1, 1\}^n$ is equipped with a uniform measure, we consider the x_i's as independent Rademacher's random variables. Defining $S_n = a_1 x_1 + \cdots + a_n x_n$, and $S'_{n-1} = S_n - a_1 x_1$, we have $f = \mathrm{sgn}(S_n - \theta)$. Without loss of generality we assume that S_n does not take the value 0 and $\theta \geq 0$. Let us reformulate Theorem 20 using probabilistic terms. First of all, observe that

$$\mathrm{Inf}_1(f) = \mathbb{P}\left\{\left|S'_{n-1} - \theta\right| \leq a_1\right\}. \tag{4.2.1}$$

Furthermore, since S_n is a symmetric random variable and $\theta > 0$, we have by definition of sgn

$$\mathbb{E}\ \mathrm{sgn}(S_n - \theta) = \mathbb{P}\{S_n \geq \theta\} - \mathbb{P}\{S_n < \theta\} = -\mathbb{P}\{|S_n| < \theta\}.$$

Thus integrating w.r.t. x_1 we get again by symmetry of S'_{n-1}

$$|\,\mathbb{E}\ \mathrm{sgn}(S_n - \theta)| = \mathbb{P}\{|S'_{n-1} - a_1| < \theta\}. \tag{4.2.2}$$

Normalizing S'_{n-1} in (4.2.1)–(4.2.2) by $\sqrt{1-a_1^2}$ and replacing $n-1$ by n, a_1 by $\tau := a_1/\sqrt{1-a_1^2}$ and θ by $\theta/\sqrt{1-a_1^2}$ we see that Theorem 20 follows from the following theorem.

Theorem 22 (Dzindzalieta, Götze [30]). *Let $\|a\|_2 = 1$, $\|a\|_\infty \leq \tau$ for some $\tau > 0$ and*

$$\mathbb{P}\{|S_n - \tau| < \theta\} = 1 - \epsilon \tag{4.2.3}$$

for some $0 \leq \epsilon \leq 1$. There exists an absolute constant $C > 0$ such that

$$\mathbb{P}\{|S_n - \theta| \leq \tau\} \geq \frac{C\epsilon\tau}{\sqrt{1+\tau^2}}. \tag{4.2.4}$$

The constant C can be taken from Theorem 20 and thus the optimal constant in (4.2.4) is larger than $3\sqrt{2}/64 \approx 0.066$.

We can interpret Theorem 22 as a conditional small ball inequality for sums of weighted independent Rademacher's random variables.

The main strategy in the proof of Theorem 22 is to use the following lower bound for unconditional *small ball* probabilities of the sum of weighted independent Rademacher's random variables.

Lemma 23 (Dzindzalieta, Götze [30]). *Let $\|a\|_2 \leq 1$ and $\|a\|_\infty \leq \tau$, then there exists an absolute constant $c > 0$ such that*

$$\mathbb{P}\{|S_n| \leq \tau\} \geq \frac{c\tau}{\sqrt{1+\tau^2}}. \tag{4.2.5}$$

Furthermore, the optimal constant in (4.2.5), say c_*, is larger than $3\sqrt{2}/16 \approx 0.26$.

As can be seen from the proof, the bounds for an optimal constant c_* in Lemma 23 depend on bounds for $\mathbb{P}\{S_n \geq \tau\}$ for τ close to 1. Taking $\tau = 1-\gamma$ with sufficiently small $\gamma > 0$ and $S_n = \frac{1}{2}(x_1+x_2+x_3+x_4)$ we get that $c_* \leq 3\sqrt{2}/8 \approx 0.53$. We believe that the optimal constant in (4.2.5) is equal to $3\sqrt{2}/8$. Inequality (4.2.5) has been proved in Khot et. al. [52] with a constant $c \approx 8.6 \cdot 10^{-40}$.

4.3 Proofs

In our proofs in this chapter we will use again Lyapunov type Berry-Esseen bounds with explicit constants for the remainder term in the Central Limit Theorem(see Chapter 3 for more details).

In this section w.l.o.g. we assume that $\|a\|_2 = 1$ and $a_1 \geq \cdots \geq a_n \geq 0$.

Proof of Lemma 23. Before proving this lemma it is worth to note, that the proof would be much easier, if one wanted to prove only the existence of an absolute constant c. Our goal is to find the best possible bound for the constant using our method, therefore the proof becomes a bit cumbersome.

For $\tau \in (1, \sqrt{2}]$ Lemma 23 follows by (3.1.7) and Theorem 12 from Chapter 3. Note that Theorem 12 allows us to obtain a better bound for the constant c in (4.2.5) for $\tau \in (\sqrt{8/7}, \sqrt{2})$.

For $\tau \geq \sqrt{2}$ Lemma 23 follows from Theorem 11 from Chapter 3.

For $\tau < 1$ the proof is not so straightforward. In this case we consider two cases where a is localized, i.e., all a_i's are "small" and where a is delocalized, i.e., at least one a_i is "large". This method has been found very effective and is used by many authors, e.g., [6, 29, 38, 61, 82]. We denote $\delta := a_1$. We integrate S_n with respect to the first variable obtaining

$$\mathbb{P}\{|S_n| \leq \tau\} = \frac{1}{2}\mathbb{P}\{|S'_{n-1} - \delta| \leq \tau\} + \frac{1}{2}\mathbb{P}\{|S'_{n-1} + \delta| \leq \tau\}.$$

Since $\{x \in \mathbb{R} : |x| \leq \tau + \delta\} \subset \{x \in \mathbb{R} : |x - \delta| \leq \tau\} \cup \{x \in \mathbb{R} : |x + \delta| \leq \tau\}$ we have

$$\mathbb{P}\left\{|S_n| \leq \tau\right\} \geq \frac{1}{2}\mathbb{P}\left\{|S'_{n-1}| \leq \tau + \delta\right\}.$$

Let $S^*_{n-1} = S'_{n-1}/\sqrt{1-\delta^2}$. It is clear that $S^*_{n-1} = b_2 x_2 + \cdots + b_n x_n$ such that $\delta/\sqrt{1-\delta^2} \geq b_2 \geq \cdots \geq b_n$ and $b_2^2 + \cdots + b_n^2 = 1$. We have

$$\mathbb{P}\{|S_n| \leq \tau\} \geq \frac{1}{2}\mathbb{P}\left\{|S^*_{n-1}| \leq \frac{\tau+\delta}{\sqrt{1-\delta^2}}\right\} = \frac{1}{2} - \mathbb{P}\left\{S^*_{n-1} > \frac{\tau+\delta}{\sqrt{1-\delta^2}}\right\}.$$

If $(\tau+\delta)/\sqrt{1-\delta^2} \geq 1$ then using (3.1.7) we have that

$$\mathbb{P}\{|S_n| \leq \tau\} \geq \frac{1}{2} - \frac{5}{16} = \frac{3}{16} \geq \frac{3\sqrt{2}}{16} \frac{\tau}{\sqrt{1+\tau^2}}.$$

Let $(\tau+\delta)/\sqrt{1-\delta^2} < 1$. Since $\delta > 0$ we obtain that $\delta < (\sqrt{2-\tau^2}-\tau)/2$. Let A be a set of points $(\delta, \tau) \in \mathbb{R}^2$ such that $0 < \delta \leq \min\{\tau, (\sqrt{2-\tau^2}-\tau)/2\}$ and $\tau \in (0,1)$. Using the Berry-Esseen bounds (3.2.7) we get that

$$\mathbb{P}\{|S_n| \leq \tau\} \geq 1 - 2I\left(\frac{\tau+\delta}{\sqrt{1-\delta^2}}\right) - \frac{2c_{\mathrm{L}}\delta}{\sqrt{1-\delta^2}}. \tag{4.3.1}$$

Thus in a view of (4.2.5) and (4.3.1) it is enough to prove that

$$g(\delta,\tau) = g \overset{\text{def}}{=} \left(1 - 2I\left(\frac{\tau+\delta}{\sqrt{1-\delta^2}}\right) - \frac{2c_{\mathrm{L}}\delta}{\sqrt{1-\delta^2}}\right)\frac{\sqrt{1+\tau^2}}{\tau} \geq \frac{3\sqrt{2}}{16} \tag{4.3.2}$$

on A.

We consider the two cases $\tau > (\sqrt{2-\tau^2}-\tau)/2$ and $\tau \leq (\sqrt{2-\tau^2}-\tau)/2$ separately.

Let $\tau < (\sqrt{2-\tau^2}-\tau)/2$, i.e., $\tau < 1/\sqrt{5}$, then by Sublemma 1 it is enough to consider the case $\delta = \tau$. In this case

$$g = \left(1 - 2I\left(\frac{2\tau}{\sqrt{1-\tau^2}}\right) - \frac{2\,c_{\mathrm{L}}\tau}{\sqrt{1-\tau^2}}\right)\frac{\sqrt{1+\tau^2}}{\tau}.$$

It is easy to verify that $I\left(\frac{2\tau}{\sqrt{1-\tau^2}}\right)$ is convex, hence

$$I\left(\frac{2\tau}{\sqrt{1-\tau^2}}\right) \leq 1 - \sqrt{5}(1-2I(1))\tau,$$

and therefore

$$g(\tau,\tau) \geq \left(\sqrt{5}(1-2I(1)) - \frac{2c_{\mathrm{L}}}{\sqrt{1-\tau^2}}\right)\sqrt{1+\tau^2}. \tag{4.3.3}$$

Furthermore, since $\sqrt{1-\tau^2} \geq \sqrt{1-1/5} = 2/\sqrt{5}$ and $\sqrt{1+\tau^2} \geq 1$, we have

$$g(\tau,\tau) \geq \left(\sqrt{5}(1-2I(1)) - \sqrt{5}c_{\mathrm{L}}\right) > \frac{3\sqrt{2}}{16}.$$

Let $\tau \geq (\sqrt{2-\tau^2}-\tau)/2$. By Sublemma 1 we can take $\delta = (\sqrt{2-\tau^2}-\tau)/2$.

In this case after simple algebraic manipulation we obtain

$$g \equiv g(\delta,\tau) = \left(1 - 2I(1) - 2c_L \frac{\sqrt{2-\tau^2}-\tau}{\sqrt{2-\tau^2}+\tau}\right) \frac{\sqrt{1+\tau^2}}{\tau}.$$

Next since $\sqrt{2-\tau^2}+\tau \geq 4/\sqrt{5}$, $\sqrt{2-\tau^2}-\tau \geq \sqrt{1-\tau^2}$ and $\sqrt{1-t^4} \geq 1$ we have

$$g \geq (1-2I(1))\frac{\sqrt{1+\tau^2}}{\tau} - \frac{\sqrt{5}c_L}{2\tau} =: h(\tau).$$

It is easy to show that the derivative of h satisfies for all $\tau \geq 0$

$$\begin{aligned} h'(\tau) &= \frac{1}{\tau^2}\left(-\frac{1-2I(1)}{\sqrt{1+\tau^2}} + \frac{\sqrt{5}c_L}{2}\right) \\ &\geq \frac{1}{\tau^2}\left(-\frac{\sqrt{5}(1-2I(1))}{\sqrt{6}} + \frac{\sqrt{5}c_L}{2}\right) \approx \frac{0.003}{\tau^2}. \end{aligned}$$

Thus $h'(\tau) > 0$ and $g \geq h\left(1/\sqrt{5}\right) \approx 0.27 > \frac{3\sqrt{2}}{16}$. □

Sublemma 1. *The partial derivative $\partial_\delta g(\delta,\tau)$ of the function g defined by (4.3.2) is negative on A.*

Proof of Sublemma 1. It is easy to see that $\partial_\delta g(\delta,\tau) < 0$ is equivalent to

$$\varphi\left(\frac{\tau+\delta}{\sqrt{1-\delta^2}}\right) + \varphi\left(\frac{\tau+\delta}{\sqrt{1-\delta^2}}\right)\delta\tau - c_L < 0. \tag{4.3.4}$$

It is clear that $\varphi\left(\frac{\tau+\delta}{\sqrt{1-\delta^2}}\right) \leq \varphi(0)$ and $\frac{\tau+\delta}{\sqrt{1-\delta^2}} \geq 2\delta$ on A. Thus the left hand side of (4.3.4) is not larger than

$$\varphi(0) + \varphi(2\delta)\delta - c_L \leq \varphi(0) + \varphi(1)/2 - c_L < 0.$$

This ends the proof of Sublemma 1. □ □

Proof of Theorem 22. We consider the two cases, $\theta \leq \tau$ and $\theta > \tau$, separately.

Case $\theta \leq \tau$. Since

$$\mathbb{P}\{|S_n - \theta| \leq \tau\} \geq \mathbb{P}\{0 \leq S_n \leq \tau\} \geq \frac{1}{2}\mathbb{P}\{|S_n| \leq \tau\}.$$

Theorem 22 follows by Lemma 23.

Case $\theta > \tau$. By (4.2.3) we get that

$$\mathbb{P}\{S_n \geq \theta - \tau\} \geq \epsilon/2. \tag{4.3.5}$$

Given $a_1, \ldots, a_n$ we denote

$$A_k = \{\{-1,1\}^n \ni x : S_0(x) < \theta - \tau, \ldots, S_{k-1}(x) < \theta - \tau, S_k(x) \geq \theta - \tau\}.$$

Note that the sets $A_1, \ldots, A_n$ are disjoint and $\{x : S_n \geq \theta - \tau\} \subset \bigcup_{i=1}^n A_i$.

From (4.3.5) it follows that $\mathbb{P}\{A_1 \cup \cdots \cup A_n\} \geq \epsilon/2$.

Due to symmetry of $S_n - S_k = a_{k+1}x_{k+1} + \cdots + a_n x_n$ we have

$$\mathbb{P}\{|S_n - \theta| \leq \tau | A_k\} \geq \frac{1}{2}\mathbb{P}\{|S_n - S_k| \leq \tau\}. \tag{4.3.6}$$

By Lemma 23 and (4.3.5)–(4.3.6) we get that

$$P\{|S_n - \tau| \leq \theta\} \geq \frac{c}{4}\frac{\epsilon\tau}{\sqrt{1+\tau^2}}$$

with the same constant c as in *Lemma* 23, so $C \geq c/4 \geq 0.066$. This ends the proof of the *Theorem* 22.

□

Lemma 24. *Let f be a LTF and $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$. If $i < j$, then* $\mathrm{Inf}_i(f) \geq \mathrm{Inf}_j(f)$.

Proof. Let $S_n^{i,j} = S_n - a_i x_i - a_j x_j$, where $S_n = a_1 x_1 + \cdots + a_n x_n$. By definition we have

$$\begin{aligned}
& \mathrm{Inf}_i(f) = \mathbb{P}\{S_n^{i,j} + a_j x_j + \theta \in [-a_i, a_i)\} \\
= \;& \frac{1}{2}\mathbb{P}\{S_n^{i,j} + \theta \in [-a_i - a_j, a_i + a_j)\} + \frac{1}{2}\mathbb{P}\{S_n^{i,j} + \theta \in [a_j - a_i, a_i - a_j)\} \\
\geq \;& \frac{1}{2}\mathbb{P}\{S_n^{i,j} + \theta \in [-a_i - a_j, a_j - a_i) \cup [a_i - a_j, a_i + a_j)\} = \mathrm{Inf}_j(f).
\end{aligned}$$

□

Chapter 5

Maximal random walks based on martingales

In this chapter we assume that Bernoulli random variables X_i's are only bounded and have a martingale type dependence, i.e., $M_k = X_1 + \cdots + X_k$ is a martingale sequence. We find tight upper bounds for the probability that a random walk based on a martingale sequence M_k visits an interval $[x, \infty)$. As far as we know our result gives the first known tight bounds for a Hoeffding type [47] inequalities for martingales with bounded differences. We also show that the maximizing random walk is an inhomogeneous Markov chain. We present a full description of the maximizing random walk and give an explicit expression for the maximal probability. We extend the results to random walks based on supermartingale sequences.

5.1 Introduction and results

We consider random walks, say $W_n = (M_0, M_1, \ldots, M_n)$ of length n starting at 0 and based on a martingale sequence $M_k = X_1 + \cdots + X_k$ (assume $M_0 = 0$) with differences $X_m = M_m - M_{m-1}$. Let $\mathcal{M}$ be the class of martingales with bounded differences such that $|X_m| \leq 1$ and $\mathbb{E}\left(X_m | \mathcal{F}_{m-1}\right) = 0$ with respect to some increasing sequence of σ-algebras $\emptyset \subset \mathcal{F}_0 \subset \cdots \subset \mathcal{F}_n$. We assume that a class $\mathcal{M}$ is such that maximal random walks belongs to this class. If a random walk W_n is based on a martingale sequence of the class $\mathcal{M}$ then we write symbolically $W_n \in \mathcal{M}$. Extensions to super-martingales are provided at the end of the Introduction.

In this chapter we provide a solution of the problem

$$D_n(x) \stackrel{\text{def}}{=} \sup_{W_n \in \mathcal{M}} \mathbb{P}\left\{W_n \text{ visits an interval } [x, \infty)\right\}, \quad x \in \mathbb{R}. \tag{5.1.1}$$

In particular, we describe random walks which maximize the probability in (5.1.1) and give an explicit expression of the upper bound $D_n(x)$. It turns out that the random walk maximizing the probability in (5.1.1) is an inhomogeneous Markov chain, i.e., given x and n, the distribution of kth step depends only on M_{k-1} and k. For an integer $x \in \mathbb{Z}$, the maximizing random walk is a simple symmetric random walk (that is, a symmetric random walk with independent steps of length 1) stopped at x. For non-integer x, the maximizing random walk makes some steps of smaller sizes. Smaller steps are needed to make the remaining distance an integer number. When the remaining distance becomes integer number the random walk continues as a simple random walk. The average total number of the smaller steps is bounded by 2. For martingales our result can be interpreted as a maximal inequality

$$\mathbb{P}\left\{\max_{1\le k\le n} M_k \ge x\right\} \le D_n(x).$$

The maximal inequality is optimal since the equality is achieved by martingales related to the maximizing random walks, that is,

$$\sup_{W_1,\dots,W_n\in\mathcal{M}} \mathbb{P}\left\{\max_{1\le k\le n} M_k \ge x\right\} = D_n(x), \tag{5.1.2}$$

where we denote by W_k a random walk $(M_0, M_1, \dots, M_k, M_k, \dots, M_k) \in \mathcal{M}$.

To prove the result we formulate a general principle of maximal inequalities for (natural classes of) martingales which reads as

$$\sup_{W_1,\dots,W_n\in\mathcal{M}} \mathbb{P}\left\{\max_{1\le k\le n} M_k \ge x\right\} = \sup_{M_n\in\mathcal{M}} \mathbb{P}\{M_n \ge x\} \tag{5.1.3}$$

in our case. It means that for martingales, the solutions of problems of type (5.1.1) are *inhomogeneous Markov chains*, i.e., the problem of type (5.1.1) can always be reduced to finding a solution of (5.1.1) in a class of inhomogeneous Markov chains.

Our methods are similar in spirit to a method used in [10], where problem (5.1.1) was solved for integer x. Namely, it was shown that if $R_n = \varepsilon_1 + \cdots + \varepsilon_n$ is a sum of independent Rademacher's random variables such that $\mathbb{P}\{\varepsilon_i = -1\} = \mathbb{P}\{\varepsilon_i = -1\} = 1/2$ and $B(n,k)$ is a normalized sum of $n-k+1$ smallest binomial coefficients, i.e.,

$$B(n,k) = 2^{-n} \sum_{i=0}^{n-k} \binom{\left\lfloor \frac{i}{2} \right\rfloor}{n} \tag{5.1.4}$$

where $\lfloor x \rfloor$ denotes an integer part of x, then for all $k \in \mathbb{Z}$

$$D_n(k) = B(n,k) = \begin{cases} 2\,\mathbb{P}\{R_n \geq k+1\} + \mathbb{P}\{R_n = k\} & \text{if } n+k \in 2\mathbb{Z}, \\ 2\,\mathbb{P}\{R_n \geq k+1\} & \text{if } n+k \in 2\mathbb{Z}+1. \end{cases}$$

In Chapter 2 we solved the problem (5.1.1) in the case of sums of bounded independent symmetric random variables. We reformulate the result using notation of this section and get that if $S_n = X_1 + \cdots + X_n$ is a sum of independent symmetric random variables such that $|X_i| \leq 1$ then

$$\mathbb{P}\{S_n \geq x\} \leq \begin{cases} \mathbb{P}\{R_n \geq x\} & \text{if } n + \lceil x \rceil \in 2\mathbb{Z}, \\ \mathbb{P}\{R_{n-1} \geq x\} & \text{if } n + \lceil x \rceil \in 2\mathbb{Z}+1, \end{cases}$$

where $\lceil x \rceil$ denotes the smallest integer number greater or equal to x. We note that for integer x the random walk based on the sequence R_k stopped at a level x is a solution of (5.1.1).

To best of our knowledge, the statement below is the first result where problems for martingales of type (5.1.1) and (5.1.2) are solved for all $x \in \mathbb{R}$.

Let us turn to more detailed formulations of our results. For a martingale $M_n \in \mathcal{M}$ and $x \in \mathbb{R}$, we introduce the stopping time

$$\tau_x = \min\{k : \ M_k \geq x\}. \tag{5.1.5}$$

The stopping time τ_x is a non-negative integer valued random variable possibly taking the value $+\infty$ in cases where $M_k < x$ for all $k = 0, 1, \ldots$. For a martingale $M_n \in \mathcal{M}$, define its version stopped at level x as

$$M_{n,x} = M_{\tau_x \wedge n}, \quad a \wedge b = \min\{a, b\}. \tag{5.1.6}$$

Given a random walk $W_n = \{0, M_1, \ldots, M_n\}$ its stopped version is denoted as $W_{n,x} = \{0, M_{1,x}, \ldots, M_{n,x}\}$.

Fix n and $x > 0$. The maximizing random walk $RW_n = \{0, M_1^x, \ldots, M_n^x\}$ is defined as follows. We start at 0. Suppose that after k steps the remaining distance to the target $[x, \infty)$ is ρ_k. The distribution of the next step is a Bernoulli random variable (which takes only two values), say $X^* = X^*(k, \rho_k, n)$, such that

$$\sup_X \mathbb{E}\, D_{n-k}(\rho_k - X) = \mathbb{E}\, D_{n-k}(\rho_k - X^*) \tag{5.1.7}$$

where sup is taken over all random variables X such that $|X| \leq 1$ and $\mathbb{E}\, X = 0$.

The distribution of the next step X^* depends on four possible situations.

$i)$ ρ_k is integer;
$ii)$ $n-k$ is odd and $0<\rho_k<1$;
$iii)$ the integer part of ρ_k+n-k is even;
$iv)$ the integer part of ρ_k+n-k is odd and $\rho_k>1$.

After k steps we make a step of length s_l or s_r to the left or right with probabilities $p_i=\frac{s_r}{s_r+s_l}$ and $q_i=\frac{s_l}{s_r+s_l}$ respectively. Let $\{x\}$ denote the fractional part of a number x. Depending on (i)-(iv) we have.

$i)$ $s_l=s_r=1$ with equal probabilities $p_1=q_1=\frac{1}{2}$, i.e., we continue as a simple random walk;

$ii)$ $s_l=\rho_k$ and $s_r=1-\rho_k$ with $p_2=1-\{\rho_k\}$ and $q_2=\{\rho_k\}$, i.e., we make a step so that the remaining distance ρ_{k+1} becomes equal either to 0 or 1;

(iii) $s_l=\{\rho_k\}$ and $s_r=1$ with $p_3=\frac{1}{1+\{\rho_k\}}$ and $q_3=\frac{\{\rho_k\}}{1+\{\rho_k\}}$, i.e., we make a step to the left so that ρ_{k+1} is of the same parity as $n-k-1$ or to the right side as far as possible ;

(iv) $s_l=1$ and $s_r=1-\{\rho_k\}$ with $p_4=\frac{1-\{\rho_k\}}{2-\{\rho_k\}}$ and $q_4=\frac{1}{2-\{\rho_k\}}$, i.e., we make a step to the left so that ρ_{k+1} is of the same parity as $n-k-1$ or to the right side as far as possible.

In other words if ρ_k is non-integer then the maximizing random walk jumps so that ρ_{k+1} becomes of the same parity as the remaining number of steps $n-k-1$ or the step $\min\{x,1\}$ to the other side. If the remaining distance ρ_k is integer, then it continues as a simple random walk.

The main result of the chapter is the following theorem.

Theorem 25 (Dzindzalieta [33])**.** *The random walk RW_n stopped at x maximizes the probability to visit an interval $[x,\infty)$ in first n steps, i.e., the following equalities hold*

$$D_n(x)=\mathbb{P}\{RW_{n,x} \text{ visits an interval } [x,\infty)\}=\mathbb{P}\{M_{n,x}^x\geq x\}, \tag{5.1.8}$$

for all $x\in\mathbb{R}$ and $n=0,1,2,\ldots$.

An explicit definition of $D_n(x)$ depends on the parity of n. Namely, let $x=m+\alpha$ with $m\in\mathbb{Z}$ and $0\leq\alpha<1$.

If $m+n$ is odd then

$$D_n(x)=\sum_{i=0}^{h}a_i\,B(n-i-1,m+i),\quad a_i=\frac{\alpha^i}{(1+\alpha)^{i+1}}, \tag{5.1.9}$$

where $h = (n - m - 1)/2$.

If $m + n$ is even then

$$D_n(x) = \sum_{i=0}^{m+1} b_i \, B(n - i - 1, m - i + 1), \tag{5.1.10}$$

where $b_i = \frac{(1-\alpha)^i}{(2-\alpha)^{i+1}}$, for $i < m$, $b_m = \alpha \left(\frac{1-\alpha}{2-\alpha}\right)^m$ and $b_{m+1} = (1-\alpha)\left(\frac{1-\alpha}{2-\alpha}\right)^m$.

It is easy to see from (5.1.9) and (5.1.10) that D_n is decreasing and continuous for all $x \in \mathbb{R}$ except at $x = n$ it has a jump. In particular we have that $D_n(x) = 1$ for $x \le 0$ and $D_n(x) = 0$ for $x > n$. In *Section* 5.3 we prove that the function D_n is piecewise convex and piecewise continuously differentiable. We also give the recursive definition of the function D_n.

A great number of papers is devoted to construction of *upper* bounds for tail probabilities of sums of random variables. The reader can find classical results in books [75, 85]. One of the first and probably the most known non-asymptotic bound for $D_n(x)$ was given by Hoeffding in 1963 [47]. He proved that for all x the function $D_n(x)$ is bounded by $\exp\{-x^2/2n\}$. Hoeffding's inequalities remained unimproved until 1995 when Talagrand [90] inserted certain missing factors. Bentkus 1986–2007 [7, 10, 11, 13] developed induction based methods. If it is possible to overcome related technical difficulties, these methods lead to the best known upper bounds for the tail probabilities (see [6, 31] for examples of tight bounds received using these methods). In [10] first tight bound for $D_n(x)$ for integer x was obtained. To overcome technical difficulties for non-integer x in [10] the linear interpolation between integer points was used, thus losing precision for non-integer x. Our method is similar in spirit to [10].

5.1.1 An extension to super-martingales

Let $\mathcal{SM}$ be the class of super-martingales with bounded differences such that $|X_m| \le 1$ and $\mathbb{E}\,(X_k|\mathcal{F}_{k-1}) \le 0$ with respect to some increasing sequence of σ-algebras $\emptyset \subset \mathcal{F}_0 \subset \cdots \subset \mathcal{F}_n$. We show that

Theorem 26 (Dzindzalieta [33])**.** *For all $x \in R$ we have*

$$\sup_{SW_n \in \mathcal{SM}} \mathbb{P}\left\{SW_n \textit{ visits an interval } [x, \infty)\right\} = D_n(x). \tag{5.1.11}$$

For super-martingales *Theorem* 26 can also be interpreted as the maximal inequality

$$\mathbb{P}\left\{\max_{1\le k\le n} M_k \ge x\right\} \le D_n(x),$$

where $M_k \in \mathcal{SM}$, and furthermore, the sup over the class of super-martingales is achieved on a martingale class.

Proof of *Theorem* 26**.** Suppose that sup in (5.1.11) is achieved with some super-martingale $SM_n = X_1 + \cdots + X_n$. Let $M_n = Y_1 + \cdots + Y_n$ be a sum of random variables, such that

$$(Y_k|\mathcal{F}_{k-1}) = ((X_k|\mathcal{F}_{k-1}) - 1)\frac{\mathbb{E}\,(X_k|\mathcal{F}_{k-1})}{1 - \mathbb{E}\,(X_k|\mathcal{F}_{k-1})}.$$

It is easy to see that $Y_k \geq 0$, $|X_k + Y_k| \leq 1$ and $\mathbb{E}\,(X_k + Y_k|\mathcal{F}_{k-1}) = 0$, so $SM_n + M_n \in \mathcal{M}$. Since $Y_k \geq 0$ we have that $M_n \geq 0$, so $\mathbb{P}\,\{SM_n + M_n \geq x\}$ is greater or equal to $\mathbb{P}\,\{SM_n \geq x\}$. This proves the theorem. □

5.2 Maximal inequalities for martingales

Let $\mathcal{M}$ be a class of martingales. Introduce the upper bounds for tail probabilities and in the maximal inequalities as

$$B_n(x) \stackrel{\text{def}}{=} \sup_{M_n \in \mathcal{M}} \mathbb{P}\{M_n \geq x\}, \quad B_n^*(x) \stackrel{\text{def}}{=} \sup_{M_n \in \mathcal{M}} \mathbb{P}\left\{\max_{0 \leq k \leq n} M_k \geq x\right\}$$

for $x \in \mathbb{R}$ (we define $M_0 = 0$).

Let as before τ_x be a stopping time defined by

$$\tau_x = \min\{k : \ M_k \geq x\}. \tag{5.2.1}$$

Then we have the following result.

Theorem 27 (Dzindzalieta [33])**.** *If a class $\mathcal{M}$ of martingales is closed under stopping at level x, then*

$$B_n(x) \equiv B_n^*(x).$$

We can interpret *Theorem* 27 by saying that inequalities for tail probabilities for natural classes of martingales imply (seemingly stronger) maximal inequalities. This means that maximizing martingales are *inhomogeneous Markov chains*. Assume that for all $M_n \in \mathcal{M}$ we have

$$\mathbb{P}\{M_n \geq x\} \leq g_n(x)$$

with some function g which depends only on n and the class $\mathcal{M}$. Then it follows that

$$\mathbb{P}\left\{\max_{0 \leq k \leq n} M_k \geq x\right\} \leq g_n(x).$$

In particular, equalities (5.1.1)–(5.1.3) are equivalent.

Proof of *Theorem* 27**.** It is clear that $B_n \le B_n^*$ since $M_n \le \max_{0\le k\le n} M_k$. Therefore it suffices to check the opposite inequality $B_n \ge B_n^*$. Let $M_n \in \mathcal{M}$. Using the fact that $M_{\tau_x\wedge n} \in \mathcal{M}$, we have

$$\mathbb{P}\left\{\max_{0\le k\le n} M_k \ge x\right\} = \mathbb{P}\left\{M_{\tau_x\wedge n} \ge x\right\} \le B_n(x). \tag{5.2.2}$$

Taking in (5.2.2) sup over $M_n \in \mathcal{M}$, we derive $B_n^* \ge B_n$. $\square$

In general conditions of *Theorem* 27 are fulfilled under usual moment and range conditions. That is, conditions of type

$$\mathbb{E}\left(|X_k|^{\alpha_k} \mid \mathcal{F}_{k-1}\right) \le g_k, \quad \left(X_k \mid \mathcal{F}_{k-1}\right) \in I_k,$$

with some $\mathcal{F}_{k-1}$-measurable $\alpha_k \ge 0$, $g_k \ge 0$, and intervals I_k with $\mathcal{F}_{k-1}$-measurable endpoints. One can use as well assumptions like symmetry, unimodality, etc.

5.3 Proofs

In order to prove *Theorem* 25 we need some additional lemmas.

Lemma 28. *Suppose $f \in C^1(0,2)$ is a continuously differentiable, non-increasing, convex function on $(0,2)$. Suppose that f is also two times differentiable on intervals $(0,1)$ and $(1,2)$. The function $F : (0,2) \to R$ defined as*

$$\begin{aligned} F(x) &= \frac{1}{x+1} f(0) + \frac{x}{x+1} f(x+1) \qquad \textit{for } x \in (0,1];\\ F(x) &= \frac{2-x}{3-x} f(x-1) + \frac{1}{3-x} f(2) \qquad \textit{for } x \in (1,2) \end{aligned}$$

is convex on intervals $(0,1)$ and $(1,2)$.

Proof. Since the function f is decreasing and convex, we have that

$$f'(x+1) \ge \frac{f(x+1)-f(0)}{x+1} \qquad \text{for } x \in (0,1); \tag{5.3.1}$$

$$f'(x-1) \le \frac{f(2)-f(x-1)}{3-x} \qquad \text{for } x \in (1,2). \tag{5.3.2}$$

For $x \in (0,1)$ simple algebraic manipulations gives

$$F''(x) = \frac{x}{x+1} f''(x+1) + \frac{2}{(x+1)^2} \left(f'(x+1) - \frac{f(x+1) - f(0)}{x+1} \right). \quad (5.3.3)$$

By (5.3.1) the second term in right hand side of (5.3.3) is non-negative. Thus $F''(x) \geq 0$ for all $x \in (0,1)$.

For $x \in (1,2)$ similar algebraic manipulation gives

$$F''(x) = \frac{2-x}{3-x} f''(x-1) - \frac{2}{(3-x)^2} \left(f'(x-1) - \frac{f(2) - f(x-1)}{3-x} \right). \quad (5.3.4)$$

By (5.3.2) the second term in right hand side of (5.3.4) is non-negative. Thus $F''(x) \geq 0$ for all $x \in (1,2)$. □ □

We use *Lemma* 28 to prove that the function $x \to D_n(x)$ satisfies the following analytic properties.

Lemma 29. *The function D_n is convex and continuously differentiable on intervals $(n-2,n), (n-4,n-2), \ldots, (0, 2\{n/2\})$.*

Proof. In order to prove this lemma it is very convenient to use a recursive definition of the function $D_n(x)$ which easily follows from the the description of the maximizing random walk $RW_{n,x}$. We have $D_0(x) = \mathbb{I}\{x \leq 0\}$ and

$$D_{n+1}(x) = \begin{cases} 1 & \text{if } x \leq 0, \\ p_1 D_n(x-1) + q_1 D_n(x+1) & \text{if } x \in \mathbb{Z} \text{ and } x > 0, \\ p_2 D_n(0) + q_2 D_n(1) & \text{if } n \in 2\mathbb{Z}+1 \text{ and } x < 1, \\ p_3 D_n(\lfloor x \rfloor) + q_3 D_n(x+1) & \text{if } \lfloor x \rfloor + n \in 2\mathbb{Z} \text{ and } x > 0, \\ p_4 D_n(x-1) + q_4 D_n(\lceil x \rceil) & \text{if } \lceil x \rceil + n \in 2\mathbb{Z} \text{ and } x > 1, \\ 0 & \text{if } x > n. \end{cases} \quad (5.3.5)$$

where $p_i + q_1 = 1$ with $p_1 = 1/2$, $p_2 = 1 - \{x\}$, $p_3 = \frac{1}{1+\{x\}}$ and $p_4 = \frac{1-\{x\}}{2-\{x\}}$.

To prove *Lemma* 28 we use induction on n. If $n = 0$ then the function $D_n(x) = \mathbb{I}\{x \leq 0\}$ clearly satisfies Lemma 29. Suppose that *Lemma* 28 holds for $n = k-1 \geq 0$. Assume $n = k$.

First we prove that D_k is convex and continuously differentiable on intervals $(0,1), (1,2), \ldots, (k-1,k)$. Since D_k is rational by construction and do not have discontinuities between integer points, it is clearly continuously differentiable on intervals $(0,1), (1,2), \ldots, (k-1,k)$. If $x \in (k-1,k]$ then by (5.1.9) we have that $D_k(x) = 2^{-k+1}/(x-k+1)$. Thus the function D_k

is clearly convex on interval $(k-1,k)$. The convexity of D_k on intervals $(0,1),(1,2),\ldots,(k-2,k-1)$ follows directly from *Lemma* 28 and recursive definition (5.3.5). To prove that the function D_k is also continuously differentiable on intervals $(k-2,k),(k-4,k-2),\ldots,(0,2\{k/2\})$ it is enough to show that $D_k'(m-0)=D_k'(m+0)$ for all $m\in\mathbb{N}$ such that $k+m\in 2\mathbb{Z}+1$. Here we write $D_k'(m-0)$ to denote $\lim_{x\uparrow m}D_k'(x)$, etc. We also write $x=m-0$ for $x=m-\epsilon$ for some vanishing $\epsilon>0$. The limit exists, since $D_n'(x)$ is continuous between integer points. If $x=m-0$ (we consider only the case $m>0$, since for $m=0$ the function $D_k(x)$ is linear), then by (5.3.5) we have

$$D_k(x)=p_4D_{k-1}(x-1)+q_4D_{k-1}(m) \tag{5.3.6}$$

and since D_{k-1} is continuously differentiable at $x-1$ we have

$$D_k'(x)=q_4^2D_{k-1}(x-1)+p_4D_{k-1}'(x-1)-q_4^2D_{k-1}(m).$$

Since $x=m-0$ we get that $D_k'(x)=D_{k-1}(x-1)-D_{k-1}(m)$. Similarly we have that if $x=m+0$ then $D_k'(x)=D_{k-1}(m)-D_{k-1}(x-1)$. Since $D_{k-1}(m-1)-D_{k-1}(m)=D_{k-1}(m)-D_{k-1}(m-1)$ we get that $D_k'(m-0)=D_k'(m+0)$. Since D_k is continuously differentiable on intervals $(k-2,k),(k-4,k-2),\ldots,(0,2\{k/2\})$ and D_k is convex on intervals $(0,1),(1,2),\ldots,(k-1,k)$ we have that $D_k(x)\geq 0$ is convex on $x=m$ for all $m\in\mathbb{N}$ such that $k+m\in 2\mathbb{Z}+1$. This ends the proof of *Lemma* 28.

□

We also need the following lemma, which is used to find the minimal dominating linear function in the proof of Theorem 25.

Lemma 30. *The function D_n satisfies the following inequalities.*
a) If $n\in 2\mathbb{Z}+1$ and $0<x<1$ then

$$p_2D_n(0)+q_2D_n(1)-p_3D_n(0)-q_3D_n(x+1)\geq 0. \tag{5.3.7}$$

b) If $\lfloor n+x\rfloor\in 2\mathbb{Z}$ then

$$p_3D_n(\lfloor x\rfloor)+q_3D_n(x+1)-p_1D_n(x-1)-q_1D_n(x+1)\geq 0. \tag{5.3.8}$$

c) If $\lfloor n+x\rfloor\in 2\mathbb{Z}+1$ and $x>1$

$$p_4D_n(x-1)+q_4D_n(\lfloor x\rfloor+1)-p_1D_n(x-1)-q_1D_n(x+1)\geq 0. \tag{5.3.9}$$

Here p_i and q_i are the same as in *Lemma* 28.

Proof. We prove this lemma by induction on n. If $n=0$ then *Lemma* 30 is equivalent to the trivial inequality $1-1 \geq 0$. Suppose that the properties (a)–(c) hold for $n = k-1 \geq 0$. Assume $n = k$.

Proof of (a). We use the following equalities which directly follow from the definition of the function D_k. If $k \in 2\mathbb{Z}+1$ and $x \in (0,1)$ then

$$\begin{aligned}
D_k(0) &= D_{k-1}(0);\\
D_k(1) &= p_1 D_{k-1}(0) + q_1 D_{k-1}(2);\\
D_k(x+1) &= p_4 D_{k-1}(2) + q_4 D_{k-1}(x);\\
D_{k-1}(x) &= D_{k-1}(0) + x\left(D_{k-1}(1) - D_{k-1}(0)\right);
\end{aligned}$$

We insert all these equalities in (5.3.7) and get that the left hand side of (5.3.7) is equal to

$$q_2 p_3 p_4 x\left(p_1 D_{k-1}(0) + q_1 D_{k-1}(2) - D_{k-1}(1)\right).$$

Now, the inequality (5.3.7) follows from the inequality

$$D_{k-1}(1) \leq D_k(1) = p_1 D_{k-1}(0) + q_1 D_{k-1}(2).$$

Proof of (b). We rewrite each term in the inequality (5.3.8) using the definition of the function D_k to get

$$\begin{aligned}
&p_1 p_3(D_{k-1}(\lfloor x\rfloor - 1) + D_{k-1}(\lfloor x\rfloor + 1))\\
&+q_3(p_3 D_{k-1}(\lfloor x\rfloor + 1) + q_3 D_{k-1}(x+2)) -\\
&p_1 p_3(D_{k-1}(\lfloor x\rfloor - 1) + D_{k-1}(\lfloor x\rfloor + 1)) -\\
&p_1 q_3(D_{k-1}(x) + D_{k-1}(x+2)).
\end{aligned}$$

The inequality

$$p_3 D_{k-1}(\lfloor x\rfloor + 1) + q_3 D_{k-1}(x+2) \geq p_1 D_{k-1}(x) + q_1 D_{k-1}(x+2)$$

follows from the inductive assumption (5.3.8) for $n = k-1$.

Proof of (c). In this case we have to consider two separate cases.

Case $x > 2$. We again rewrite each term in the inequality (5.3.9) using the definition of the function D_k to get

$$\begin{aligned}
&p_4(p_4 D_{k-1}(x-2) + q_4 D_{k-1}(\lfloor x\rfloor)) + q_4 p_1(D_{k-1}(\lfloor x\rfloor) + D_{k-1}(\lfloor x\rfloor + 2)) -\\
&q_4 p_1(D_{k-1}(\lfloor x\rfloor) + D_{k-1}(\lfloor x\rfloor + 2)) - p_4 p_1(D_{k-1}(x-2) + D_{k-1}(x)).
\end{aligned}$$

The inequality

$$p_4 D_{k-1}(x-2) + q_4 D_{k-1}(\lfloor x \rfloor) \geq p_1 D_{k-1}(x-2) + q_1 D_{k-1}(x)$$

follows from the inductive assumption (5.3.9) for $n = k-1$.

Case $1 < x < 2$. First let us again rewrite the inequality (5.3.9) using the recursive definition of D_k. After combining the terms we get that (5.3.9) is equivalent to

$$x\, D_{k-1}(1) + (1-x) D_{k-1}(0) - D_{k-1}(x) \geq 0. \tag{5.3.10}$$

Now we use the inequality

$$D_{k-1}(x) \leq (2-x) D_{k-1}(1) + (x-1) D_{k-1}(2).$$

to get that

$$\begin{aligned}
&x D_{k-1}(1) + (1-x) D_{k-1}(0) - D_{k-1}(x) \geq \\
&2(x-1) D_{k-1}(1) + (1-x) D_{k-1}(0) - (x-1) D_{k-1}(2) = \\
&(x-1)\left(2D_{k-1}(1) - D_{k-1}(0) - D_{k-1}(2)\right).
\end{aligned}$$

Equality $(2D_{k-1}(1) - D_{k-1}(0) - D_{k-1}(2)) = 0$ follows from inductive assumption. □

Now we are ready to prove *Theorem* 25.

Proof of Theorem 25. For $x \leq 0$ to achieve sup in (5.1.8) take $M_n \equiv 0$. For $x > n$ the sup in (5.1.8) is equal to zero since $M_n \leq n$ for all $n = 0, 1, \ldots$. To prove *Theorem* 25 for $x \in (0, n]$ we use induction on n.

For $n = 0$ the statement is obvious since $\mathbb{P}\{M_0 \geq x\} = \mathbb{I}\{x \leq 0\} = D_0(x)$. Suppose that *Theorem* 25 holds for $n = k > 0$. Assume $n = k+1$. In order to prove *Theorem* 25 it is enough to prove that D_{k+1} satisfies the recursive relations (5.3.5). We have

$$\begin{aligned}
\mathbb{P}\{M_{k+1} \geq x\} &= \mathbb{P}\{X_2 + \cdots + X_{k+1} \geq x - X_1\} \\
&= \mathbb{E}\,\mathbb{P}\{X_2 + \cdots + X_k \geq x - X_1 | X_1\} \\
&\leq \mathbb{E}\, D_k(x - X_1).
\end{aligned}$$

Now for every x we find a linear function $t \mapsto f(t)$ dominating the function $t \mapsto D_k(x-t)$ on interval $[-1, 1]$ and touching it at two points, say x_1 and x_2, on the different sides of zero. After this we consider a random variable, say

$X \in \{x_1, x_2\}$ with mean zero. It is clear that $\mathbb{E}\, D_k(x - X_1) \leq \mathbb{E}\, D_k(x - X)$. We show that the numbers x_1 and x_2 are such that (5.3.5) holds.

Since D_k is piecewise convex between integer points, the points where $f(t)$ touches $D_k(x-t)$ can be only the endpoints of interval $[-1, 1]$ or the points where $D_k(x-t)$ is not convex.

For $x > 0$ we consider four separate cases.

$$\begin{array}{llll} i) & x \in \mathbb{Z}; & & \\ ii) & x \notin \mathbb{Z}, & k \in 2\mathbb{Z}+1 & \text{and} \quad x < 1; \\ iii) & x \notin \mathbb{Z}, & \lfloor x \rfloor + k \in 2\mathbb{Z}; & \\ iv) & x \notin \mathbb{Z}, & \lfloor x \rfloor + k \in 2\mathbb{Z} & \text{and} \quad x > 1. \end{array}$$

Note that all these four cases covers all the positive real numbers.

Case (i)**.** Since $x \in \mathbb{Z}$ the dominating linear function touches $D_k(x-t)$ at integer points. So maximizing $X_1 \in \{-1, 0, 1\}$.
If $x + k \in 2\mathbb{Z}+1$ then the function $D_k(x-t)$ is convex on $(-1, 1)$ so maximizing X takes values 1 or -1 with equal probabilities $1/2$.
If $x + k \in 2\mathbb{Z}$, then

$$D_k(x) = \frac{1}{2}\left(D_{k-1}(x-1) + D_{k-1}\left(x+1\right)\right) = \frac{1}{2}(D_k(x-1) + D_k(x+1)),$$

so the dominating function touches $D_k(x-t)$ at all three points $-1, 0, 1$. Taking $X \in \{-1, 1\}$ we end the proof of the case (i).

The case (i) was first considered in [10].

Case (ii)**.** Since D_k is convex on intervals $(0, 1)$ and $(1, 3)$ the dominating minimal function can touch $D_k(x-t)$ only at $x, x-1, -1$. But due to an inequality (5.3.7) the linear function $f(t)$ going through $(x, D_k(0))$ and $(x-1, D_k(1))$ is above the point $(-1, D_k(x+1))$.

Case (iii)**.** Since the function D_k is convex on intervals $(\lfloor x \rfloor - 1, \lfloor x \rfloor)$ and $(\lfloor x \rfloor, \lfloor x \rfloor + 2)$ the dominating minimal function can touch $D_k(x-t)$ only at $-1, \{x\}, 1$. But due to an inequality (5.3.8) the linear function $f(t)$ going through $(\{x\}, D_k(\lfloor x \rfloor))$ and $(-1, D_k(x+1))$ is above the point $(1, D_k(x-1))$.

Case (iv)**.** Since the function D_k is convex on intervals $(\lfloor x \rfloor - 1, \lfloor x \rfloor + 1)$ and $(\lfloor x \rfloor + 1, \lfloor x \rfloor + 3)$ the dominating minimal function can touch $D_k(x+t)$ only at $-1, \{x\} - 1, 1$. But due to an inequality (5.3.9) the linear function $f(t)$ going through $(1, D_k(x-1))$ and $(\{x\} - 1, D_k(\lfloor x \rfloor + 1))$ is above the point $(-1, D_k(x+1))$. $\square$

Chapter 6

Extremal Lipschitz functions

In this chapter we obtain an optimal deviation from the mean upper bound

$$D(x) \stackrel{\text{def}}{=} \sup_{f \in \mathcal{F}} \mu\{f - \mathbb{E}_\mu f \geq x\}, \qquad \text{for } x \in \mathbb{R} \tag{6.0.1}$$

where $\mathcal{F}$ is the complete class of integrable, Lipschitz functions on probability metric (product) spaces. As corollaries we get exact solutions of (6.0.1) for Euclidean unit sphere S^{n-1} with a geodesic distance function and a normalized Haar measure, for $\mathbb{R}^n$ equipped with a Gaussian measure and for the multidimensional cube, rectangle, torus or Diamond graph equipped with uniform measure and Hamming distance function. We also prove that in general probability metric spaces the sup in (6.0.1) is achieved on a family of negative distance functions.

6.1 Introduction and results

Let us recall a well known result for Lipschitz functions on probability metric spaces, (V, d, μ). Here a probability metric space means that a measure μ is Borel and normalized, $\mu(V) = 1$. Given a measurable non-empty set $A \in V$ we denote a distance function by $d(A, u) = \min\{d(u, v), v \in A\}$. We denote by $\mathcal{F} = \mathcal{F}(V)$ the class of integrable, i.e., $f \in L_1(V, d, \mu)$, 1-Lipschitz functions, i.e., $f : V \to \mathbb{R}$ such that $|f(u) - f(v)| \leq d(u, v)$ for all $u, v \in V$. We will write in short $\{f \in A\}$ instead of $\{u : f(u) \in A\}$, etc. We will say that $\mathcal{F}(V, d, \mu)$ is *complete* if it contains all 1-Lipschitz functions f defined on (V, d, μ). Note that completeness in our sense just means that the distance function $d(x, x_0)$ is μ-integrable (this property does not depend on x_0). Let M_f be a median of the function f, i.e., a number such that $\mu\{f(x) \leq M_f\} \geq \frac{1}{2}$ and $\mu\{x : f(x) \geq M_f\} \geq \frac{1}{2}$. Given probability metric space (V, d, μ), the sup in

$$\sup_{f \in \mathcal{F}} \mu\{f - M_f \geq x\} \qquad \text{for } x \in \mathbb{R}$$

is achieved on a family, say $\mathcal{F}^*$, of distance functions $f(u) = -d(A, u)$ with a measurable set $A \subset V$ (for a nice exposition of the results we refer reader to [59, 65]). From this, it is easy to deduce that this problem is equivalent to the following isoperimetric problem. Given $t \geq 0$ and $h \geq 0$,

$$\text{minimize } \mu(A^h) \text{ over all } A \subset V \text{ with } \mu(A) \geq t, \tag{6.1.1}$$

where $A^h = \{u \in V : d(u, A) \leq h\}$ is an h-enlargement of A.

Following [16] we say that a space (V, d, μ) is *isoperimetric* if for every $t \geq 0$ there exists a solution, say A_{opt}, of (6.1.1) which does not depend on h.

However, as was pointed out by Talagrand [91] in practice it is easier to deal with expectation $\mathbb{E}\, f$ rather than median M_f. In order to get results for the mean instead of median two different techniques were usually used. One way was to evaluate the distance between median and mean, another was to use a martingale technique (see [12, 59, 64, 91] for more detailed exposition of the results). Unfortunately, none of them could lead to tight bounds for the mean.

In this chapter we find tight deviation from the mean bounds

$$D(x) \stackrel{\text{def}}{=} \sup_{f \in \mathcal{F}} \mu\{f - \mathbb{E}_\mu f \geq x\}, \qquad \text{for } x \in \mathbb{R} \tag{6.1.2}$$

for the complete class $\mathcal{F} = \mathcal{F}(V, d, \mu)$. If we change f to $-f$ we get that

$$D(x) = \sup_{f \in \mathcal{F}} \mu\{f - \mathbb{E}_\mu f \leq -x\} \qquad \text{for } x \in \mathbb{R}.$$

Note that the function $D(x)$ depends also on (V, d, μ).

We first state a general result for probability metric spaces.

Theorem 31 (Dzindzalieta [32]). *If $\mathcal{F}(V, d, \mu)$ is complete, then* sup *in* (6.1.2) *is achieved on a family of negative distance functions, i.e.,*

$$\sup_{f \in \mathcal{F}} \mu\{f - \mathbb{E}_\mu f \geq x\} = \sup_{f \in \mathcal{F}^*} \mu\{f - \mathbb{E}_\mu f \geq x\} \qquad x \in \mathbb{R}.$$

Note that $\mathcal{F}^* \subset \mathcal{F}$.

Proof. Fix $x \in \mathbb{R}$. Let $f \in \mathcal{F}$ and $B = \{f - \mathbb{E}_\mu f \geq x\}$. If $B = \emptyset$, then $\mu\{f - \mathbb{E}_\mu f \geq x\} = 0$ and thus $\mu\{f - \mathbb{E}_\mu f \geq x\} \leq \mu\{f^* - \mathbb{E}_\mu f^* \geq x\}$ for any function $f^* \in \mathcal{F}^*$. Let $B \neq \emptyset$. Since $\mathbb{E}_\mu f < \infty$ and f is bounded from below by $\mathbb{E}_\mu f + x$ on B we have that $-\infty < \mathbb{E}_\mu f + x \leq \inf_{u \in B} f(u) < \infty$. Thus without loss of generality we can assume that $\inf_{u\in B} f(u) = 0$. Let g be a function such that $g(u) = 0$ on B and $g(u) = f(u)$ on B^c. It is clear that $g \in \mathcal{F}$ and $f \geq g$ on V and thus $\mathbb{E}_\mu f \geq \mathbb{E}_\mu g$. Next, $x \leq \inf_{u\in B} f(u) - \mathbb{E}_\mu f = g - \mathbb{E}_\mu f \leq g - \mathbb{E}_\mu g$ on B, so $B \subset \{g - \mathbb{E}_\mu g \geq x\}$. Let $f^*(u) = -d(B, u)$. Since g is Lipschitz function, $|g(u)| = |g(u) - g(v)| \leq d(u, v)$ for all $v \in B$, so $|g(u)| \leq d(u, B)$ and thus $g(u) \geq -d(u, B) = f^*(u)$. Again, for all $u \in B$ we have $x \leq g(u) - \mathbb{E}\, g = f^*(u) - \mathbb{E}\, g \leq f^*(u) - \mathbb{E}\, f^*$, so $B \subset \{f^* - \mathbb{E}_\mu f^* \geq x\}$. Thus, $\mu\{f - \mathbb{E}_\mu f \geq x\} \leq \mu\{g - \mathbb{E}_\mu g \geq x\} \leq \mu\{f^* - \mathbb{E}_\mu f^* \geq x\}$. Since f is arbitrary the statement of *Theorem* 31 follows. □

In the special case when $V = \mathbb{R}^n$ and $\mu = \gamma_n$ is a standard Gaussian measure, *Theorem* 31 was proved by Bobkov [20].

Our main result of this chapter is the following theorem.

Theorem 32 (Dzindzalieta [32])**.** *If (V, d, μ) is isoperimetric and $\mathcal{F}$ is complete, then*

$$D(x) = \mu\{f^*_{opt} - \mathbb{E}_\mu f^*_{opt} \geq x\} \qquad \text{for } x \in \mathbb{R},$$

where $f^*_{opt}(u) = -d(A_{opt}, u)$ is a negative distance function from some extremal set A_{opt}. It turns out that $\mu(A_{opt}) = D(x)$.

Proof. For any measurable set $A \subset V$ we have

$$\begin{aligned} \mathbb{E}_\mu d(A, \cdot) &= \int_0^\infty \left(1 - \mu\left\{A^h\right\}\right) \mathrm{dh} \\ &\leq \int_0^\infty \left(1 - \mu\left\{A^h_{\text{opt}}\right\}\right) \mathrm{dh} = \mathbb{E}_\mu d(A_{\text{opt}}, \cdot). \end{aligned} \tag{6.1.3}$$

Let $f^* \in \mathcal{F}^*$ and $A = \{f^* - \mathbb{E}_\mu f^* \geq x\}$. From (6.1.3) we get that for all $u \in A$

$$x \leq f^*(u) - \mathbb{E}_\mu f^* \leq f^*(u) - \mathbb{E}_\mu f^*_{\text{opt}},$$

where $f^*_{\text{opt}}(u) = -d(A_{\text{opt}}, u)$. Since $f^*_{\text{opt}}(u) = 0$ for all $u \in A_{\text{opt}}$ we have that $x \leq -\mathbb{E}_\mu f^*_{\text{opt}} = f^*_{\text{opt}}(u) - \mathbb{E}_\mu f^*_{\text{opt}}$ for all $u \in A_{\text{opt}}$ as well. Since f^* (or the set A) is arbitrary and $\mu\{A_{\text{opt}}\} \geq \mu\{A\}$, by Theorem 31 we have

$$\sup_{f\in\mathcal{F}} \mu\{f - \mathbb{E}_\mu f \leq -x\} = \sup_{f\in\mathcal{F}^*} \mu\{f - \mathbb{E}_\mu f \geq x\} = \mu\{f^*_{\text{opt}} - \mathbb{E}_\mu f^*_{\text{opt}}\},$$

which completes the proof of *Theorem* 32. □

6.2 Isoperimetric spaces and corollaries

In this section we present a short overview of the results on the isoperimetric problem described by (6.1.1). We also state a number of corollaries implied by *Theorem* 31 and *Theorem* 32.

A typical and basic example of isoperimetric spaces is the Euclidean unit sphere $S^{n-1} = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} |x_i|^2 = 1\}$ with a geodesic distance function ρ and a normalized Haar measure σ_{n-1}. P. Lévy [60] and E. Schmidt [83] showed that if A is a Borel set in S^{n-1} and H is a cap (ball for geodesic distance function ρ) with the same Haar measure $\sigma_{n-1}(H) = \sigma_{n-1}(A)$, then

$$\sigma_{n-1}(A^h) \geq \sigma_{n-1}(H^h) \qquad \text{for all} \quad h > 0. \tag{6.2.1}$$

Thus A_{opt} for the space $(S^{n-1}, \rho, \sigma_{n-1})$ is a cap. We refer readers for a short proof of (6.2.1) to [14, 37]. The extension to Riemannian manifolds with strictly positive curvature can be found in [40]. Note that if H is a cap, then H^h is also a cap, so we have an immediate corollary.

Corollary 33 (Dzindzalieta [32])**.** *For a unit sphere S^{n-1} equipped with normalized Haar measure σ_{n-1} and geodesic distance function we have*

$$D(x) = \sigma_{n-1}\{f^* - \mathbb{E}_\mu f^* \geq x\} \qquad \text{for } x \in \mathbb{R},$$

where $f^*(u) = -d(A_{opt}, u)$ and A_{opt} is a cap.

Probably the most simple non-trivial isoperimetric space is n-dimensional discrete cube $C_n = \{0,1\}^n$ equipped with uniform measure, say μ, and Hamming distance function. Harper [43] proved that some number of the first elements of C_n in the simplicial order is a solution of (6.1.1). Bollobas and Leader [22] extended this result to multidimensional rectangle. Karachanjan and Riordan [50, 80] solved the problem (6.1.1) for multidimensional torus. Bezrukov considered powers of the Diamond graph [15] and powers of cross-sections [16]. We state the results for discrete spaces as corollary.

Corollary 34 (Dzindzalieta [32])**.** *For discrete multidimensional cube, rectangle, torus and Diamond graph equipped with uniform measure and Hamming distance function we have*

$$D(x) = \mu\{f^* - \mathbb{E}_\mu f^* \geq x\} \qquad \text{for } x \in \mathbb{R},$$

where $f^*(u) = -d(A_{opt}, u)$ and A_{opt} are the sets of some first elements in corresponding orders. In particular, for n-dimensional discrete cube with Hamming distance function, A_{opt} is a set of some first elements of C_n in simplicial order.

There is a vast of papers dedicated to bound $D(x)$ for various discrete spaces. We mention only [21, 58, 89] among others. In [15, 59] a nice overview of isoperimetric spaces and bounds for $D(x)$ is given.

Another important example of isoperimetric spaces comes from Gaussian isoperimetric problem. Sudakov and Tsirel'son [87] and Borell [24] discovered that if γ_n is a standard Gaussian measure on $\mathbb{R}^n$ with a usual Euclidean distance function d, then $(\mathbb{R}^n, d, \gamma_n)$ is isoperimetric. In [87] and [24] it was shown that among all subsets A of $\mathbb{R}^n$ with $t \geq \gamma_n(A)$, the minimal value of $\gamma_n(A^h)$ is attained for half-spaces of measure t. Thus we have the following corollary of *Theorem* 31 and *Theorem* 32.

Corollary 35 (Dzindzalieta [32]). *For a Gaussian space* $(\mathbb{R}^n, d, \gamma_n)$ *we have*

$$D(x) = \gamma_n\{f^* - \mathbb{E}_{\gamma_n} f^* \geq x\} \qquad \textit{for } x \in \mathbb{R},$$

where $f^*(u) = -d(A_{opt}, u)$ *is a negative distance function from a half-space of space* $\mathbb{R}^n$*.*

The latter result was firstly proved by Bobkov [20]. We also refer for further investigations of extremal sets on $\mathbb{R}^n$ for some classes of measures to [1, 19] among others.

Chapter 7

Conclusions

During the doctoral studies at Vilnius University, we have studied the tail probabilities for sums of random variables possessing various boundedness, dependence, moment restrictions. In this last Chapter, a brief summary of these results is given.

- For the sum $S = X_1 + \cdots + X_n$ of symmetric weighted independent Bernoulli random variables, such that $|X_i| \leq 1$ we solved two main problems. First one is to find a tight upper bound for $\mathbb{P}\{S_n \geq x\}$ and the second one is to find a tight upper bound for $\mathbb{P}\{S_n = x\}$. We described the random variables which gives an optimal results. We also gave a short proof of Littlewood-Offord problem 40's using mathematical induction.

- For the sum of weighted independent Rademacher random variables with unit variance we found an optimal subgaussian constant and improved Chebyshev's bound for all non-trivial arguments.

- We applied the results to Boolean valued linear threshold functions and showed that confirmed a conjecture by Matulef, O'Donnell, Rubinfeld and Servedio that the threshold function associated to a linear function with some large coefficient, if balanced, must have a large influence.

- We considered a class of martingales $M_k = X_1 + \cdots + X_n$ with bounded differences, $|X_i| \leq 1$ and described random walks based on such martingales which maximizes the probability to visit an interval $[x, \infty)$. We showed that maximizing martingales are from the class of inhomogeneous Markov processes.

- We found extremal Lipschitz function which maximizes the probability, that a Lipschitz function is larger than it's average by some given number. We showed that maximizing Lipschitz functions are negative distance functions from some sets which are solution of well known isoperimetric problem.

Bibliography

[1] F. Barthe, P. Cattiaux, and C. Roberto. Isoperimetry between exponential and Gaussian. *Electron. J. Probab*, 12(44):1212–1237, 2007.

[2] Michael Ben-Or and Nathan Linial. Collective coin flipping. *Randomness and Computation*, 5:91–115, 1990.

[3] A. Ben-Tal, A. Nemirovski, and C. Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

[4] Aharon Ben-Tal and Arkadi Nemirovski. On safe tractable approximations of chance-constrained linear matrix inequalities. *Mathematics of Operations Research*, 34(1):1–25, 2009.

[5] V Bentkus. An inequality for tail probabilities of martingales with differences bounded from one side. *Journal of Theoretical Probability*, 16(1):161–173, 2003.

[6] V. Bentkus and D. Dzindzalieta. A tight Gaussian bound for weighted sums of Rademacher random variables. *Accepted to Bernoulli*, 2014.

[7] V Yu Bentkus. Large deviations in banach spaces. *Theory of Probability & Its Applications*, 31(4):627–632, 1987.

[8] Vidmantas Bentkus and Tomas Juškevičius. Bounds for tail probabilities of martingales using skewness and kurtosis. *Lithuanian Mathematical Journal*, 48(1):30–37, 2008.

[9] Vidmantas Yu Bentkus. On large deviations in Banach spaces. *Teoriya Veroyatnostei i ee Primeneniya*, 31(4):710–716, 1986.

[10] V.K. Bentkus. An inequality for large deviation probabilities of sums of bounded iid random variables. *Lithuanian Mathematical Journal*, 41(2):112–119, 2001.

[11] V.K. Bentkus. On Hoeffding's inequalities. *Annals of probability*, 32(2):1650–1673, 2004.

[12] V.K. Bentkus. On measure concentration for separately Lipschitz functions in product spaces. *Israel Journal of Mathematics*, 158(1):1–17, 2007.

[13] V.K. Bentkus, N. Kalosha, and M. Van Zuijlen. On domination of tail probabilities of (super) martingales: explicit bounds. *Lithuanian Mathematical Journal*, 46(1):1–43, 2006.

[14] Y. Benyamini. Two-point symmetrization, the isoperimetric inequality on the sphere and some applications, in âĂIJTexas Functional Analysis Seminar 1983/84, Austin, TexasâĂİ, Longhorn Notes, 1984.

[15] S.L. Bezrukov, M. Rius, and O. Serra. The vertex isoperimetric problem for the powers of the Diamond graph. *Discrete Mathematics*, 308(11):2067–2074, 2008.

[16] S.L. Bezrukov and O. Serra. A local-global principle for vertex-isoperimetric problems. *Discrete Mathematics*, 257(2):285–310, 2002.

[17] H.D. Block. The perceptron: A model for brain functioning. I. *Reviews of Modern Physics*, 34(1):123, 1962.

[18] Sergey G Bobkov, Friedrich Götze, and Christian Houdré. On Gaussian and Bernoulli covariance representations. *Bernoulli*, 7(3):439–451, 2001.

[19] S.G. Bobkov. Isoperimetric problem for uniform enlargement. *Studia Math*, 123(1):81–95, 1997.

[20] S.G. Bobkov. Localization Proof of the Bakry–Ledoux Isoperimetric Inequality and Some Applications. *Theory of Probability and its Applications*, 47:308, 2003.

[21] S.G. Bobkov and C. Houdré. Isoperimetric constants for product probability measures. *The Annals of Probability*, 25(1):184–205, 1997.

[22] B. Bollobás and I. Leader. Isoperimetric inequalities and fractional set systems. *Journal of Combinatorial Theory, Series A*, 56(1):63–74, 1991.

[23] Béla Bollobás. *Combinatorics: set systems, hypergraphs, families of vectors, and combinatorial probability.* Cambridge University Press, 1986.

[24] C. Borell. The Brunn–Minkowski inequality in Gauss space. *Inventiones Mathematicae*, 30(2):207–216, 1975.

[25] D. L. Burkholder. Independent sequences with the Stein property. *Ann. Math. Statist.*, 39:1282–1288, 1968.

[26] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[27] Kürşad Derinkuyu and Mustafa Ç Pınar. On the s-procedure and some variants. *Mathematical Methods of Operations Research*, 64(1):55–77, 2006.

[28] Kürşad Derinkuyu, Mustafa Ç Pınar, and Ahmet Camcı. An improved probability bound for the approximate S–lemma. *Operations Research Letters*, 35(6):743–746, 2007.

[29] D. Dzindzalieta. On random signs. *Submitted*, 2013.

[30] D. Dzindzalieta and F. Götze. Halfspaces with influential variable. *Submitted*, 2013.

[31] D. Dzindzalieta, T. Juškevičius, and M. Šileikis. Optimal probability inequalities for random walks related to problems in extremal combinatorics. *SIAM Journal on Discrete Mathematics*, 26(2):828–837, 2012.

[32] Dainius Dzindzalieta. Extremal lipschitz functions in the deviation inequalities from the mean. *Electron. Commun. Probab.*, 18:no. 66, 1–5, 2013.

[33] Dainius Dzindzalieta. Random walks maximizing the probability to visit an interval. *arXiv preprint arXiv:1305.6735*, 2013.

[34] M. L. Eaton and Bradley Efron. Hotelling's T^2 test under symmetry conditions. *J. Amer. Statist. Assoc.*, 65:702–711, 1970.

[35] P Erdös. On a lemma of littlewood and offord. *Bull. Amer. Math. Soc*, 51(12):898–902, 1945.

[36] A. Fiat and D. Pechyony. Decision trees: More theoretical justification for practical algorithms. In *Algorithmic Learning Theory*, pages 156–170. Springer, 2004.

[37] T. Figiel, J. Lindenstrauss, and VD Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1):53–94, 1977.

[38] O. Friedland and O. Guédon. Random embedding of l_p^n into l_r^N. *Mathematische Annalen*, 350(4):953–972, 2011.

[39] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.

[40] M. Gromov. Paul Levy's isoperimetric inequality. *Preprint I.H.E.S*, 1980.

[41] R.K. Guy. Any answers anent these analytical enigmas? *The American Mathematical Monthly*, 93(4):279–281, 1986.

[42] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129–154, 1993.

[43] L.H. Harper. Optimal assignments of numbers to vertices. *Journal of the Society for Industrial and Applied Mathematics*, 12(1):131–135, 1964.

[44] Simai He, Zhi-Quan Luo, Jiawang Nie, and Shuzhong Zhang. Semidefinite relaxation bounds for indefinite homogeneous quadratic optimization. *SIAM Journal on Optimization*, 19(2):503–523, 2008.

[45] J.B. Hiriart-Urruty. A new series of conjectures and open questions in optimization and matrix analysis. *Esaim. Cocv*, 15(2):454, 2008.

[46] Paweł Hitczenko and Stanisław Kwapień. On the Rademacher series. In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, volume 35 of *Progr. Probab.*, pages 31–36. Birkhäuser Boston, Boston, MA, 1994.

[47] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[48] Ron Holzman and Daniel J Kleitman. On the product of sign vectors and unit vectors. *Combinatorica*, 12(3):303–316, 1992.

[49] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *Foundations of Computer Science, 1988., 29th Annual Symposium on*, pages 68–80. IEEE, 1988.

[50] V.M. Karachanjan. A discrete isoperimetric problem on multidimensional torus. In *Russian), Doklady AN Arm. SSR, vol. LXXIV*, volume 2, pages 61–65, 1982.

[51] Gy Katona. Intersection theorems for systems of finite sets. *Acta Mathematica Hungarica*, 15(3):329–337, 1964.

[52] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[53] Daniel J Kleitman. On a combinatorial conjecture of erdös. *Journal of Combinatorial Theory*, 1(2):209–214, 1966.

[54] H. König and S. Kwapień. Best Khintchine type inequalities for sums of independent, rotationally invariant random vectors. *Positivity*, 5(2):115–152, 2001.

[55] Thomas William Körner. *Fourier analysis*. Cambridge University Press, 1989.

[56] Rafał Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, 34(6):2315–2331, 2006.

[57] Rafał Latała and Krzysztof Oleszkiewicz. On the best constant in the Khinchin-Kahane inequality. *Studia Math.*, 109(1):101–104, 1994.

[58] M. Ledoux. *The concentration of measure phenomenon*. Amer Mathematical Society, 2001.

[59] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer, 1991.

[60] Paul Lévy. *Problèmes concrets d'analyse fonctionnelle. Avec un complément sur les fonctionnelles analytiques par F. Pellegrino*. Gauthier-Villars, Paris, 1951. 2d ed.

[61] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2):491–523, 2005.

[62] K. Matulef, R. O'Donnell, R. Rubinfeld, and R.A. Servedio. Testing halfspaces. *SIAM Journal on Computing*, 39(5):2004–2047, 2010.

[63] K.M. Matulef. *Testing and learning Boolean functions.* PhD thesis, Massachusetts Institute of Technology, 2009.

[64] Colin McDiarmid. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

[65] Vitali D Milman. *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds*, volume 1200. Springer Verlag, 1986.

[66] EC Milner. A combinatorial theorem on systems of sets. *Journal of the London Mathematical Society*, 1(1):204–206, 1968.

[67] M. Minsky and S. Papert. Perceptron: an introduction to computational geometry. *The MIT Press, Cambridge, expanded edition*, 19:88, 1969.

[68] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Ann. of Math. (2)*, 171(1):295–341, 2010.

[69] Arkadi Nemirovski. Sums of random symmetric matrices and quadratic optimization under orthogonality constraints. *Mathematical programming*, 109(2-3):283–317, 2007.

[70] A.B.J. Novikoff. On convergence proofs for perceptrons. Technical report, DTIC Document, 1963.

[71] Ryan O'Donnell. Some topics in analysis of Boolean functions. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 569–578. ACM, 2008.

[72] Krzysztof Oleszkiewicz. On the Stein property of Rademacher sequences. *Probab. Math. Statist.*, 16(1):127–130, 1996.

[73] Krzysztof Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 157–168. Birkhäuser, Basel, 2003.

[74] G. Peshkir and A. N. Shiryaev. Khinchin inequalities and a martingale extension of the sphere of their action. *Uspekhi Mat. Nauk*, 50(5(305)):3–62, 1995.

[75] Valentin Vladimirovič Petrov. *Sums of independent random variables.* Berlin, 1975.

[76] I. Pinelis. Optimal tail comparison based on comparison of moments. *Progress in probability*, pages 297–314, 1998.

[77] Iosif Pinelis. Extremal probabilistic problems and Hotelling's T2 test under a symmetry condition. *The Annals of Statistics*, 22(1):357–368, 1994.

[78] Iosif Pinelis. Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities. In *Advances in stochastic inequalities (Atlanta, GA, 1997)*, volume 234 of *Contemp. Math.*, pages 149–168. Amer. Math. Soc., Providence, RI, 1999.

[79] Iosif Pinelis. Toward the best constant factor for the Rademacher-Gaussian tail comparison. *ESAIM: Probability and Statistics*, 11(1):412–426, 2007.

[80] O. Riordan. An ordering on the even discrete torus. *SIAM Journal on Discrete Mathematics*, 11:110, 1998.

[81] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[82] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.

[83] E. Schmidt. Die Brunn–Minkowskische Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. I. *Mathematische Nachrichten*, 1(2-3):81–157, 1948.

[84] I.G. Shevtsova. An improvement of convergence rate estimates in the Lyapunov theorem. In *Doklady Mathematics*, volume 82-3, pages 862–864. Springer, 2010.

[85] Galen R Shorack and Jon A Wellner. *Empirical processes with applications to statistics*, volume 59. SIAM, 2009.

[86] Anthony Man-Cho So. Approximation bound for quadratic optimization problems with orthogonality constraints. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1201–1209. Society for Industrial and Applied Mathematics, 2009.

[87] V.N. Sudakov and B.S. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *Journal of Mathematical Sciences*, 9(1):9–18, 1978.

[88] Rafal Sztencel. On boundedness and convergence of some banach space valued random series. *Probab. Math. Statist*, 2(1):83–88, 1981.

[89] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathematiques de l'IHES*, 81(1):73–205, 1995.

[90] M. Talagrand. The missing factor in Hoeffding's inequalities. *Ann. Inst. H. Poincaré Probab. Statist*, 31(4):689–702, 1995.

[91] Michel Talagrand. A new look at independence. *The Annals of probability*, 24(1):1–34, 1996.

[92] I. Tyurin. New estimates of the convergence rate in the Lyapunov theorem. *arXiv preprint arXiv:0912.0726*, 2009.

[93] Mark Veraar. A note on optimal probability lower bounds for centered random variables. *Colloq. Math.*, 113(2):231–240, 2008.

[94] Mark Veraar. On Khintchine inequalities with a weight. *Proc. Amer. Math. Soc.*, 138(11):4119–4121, 2010.

[95] AC-C Yao. On ACC and threshold circuits. In *Foundations of Computer Science, 1990. Proceedings., 31st Annual Symposium on*, pages 619–627. IEEE, 1990.

[96] AV Zhubr. On one extremal problem for N-cube. *Izvestija Komi NC UrO RA*, 10:4–11, 2012.