CONCENTRATION ESTIMATES AND THE CENTRAL LIMIT PROBLEM FOR CONVEX BODIES

Milla Anttila

Ph.D Thesis

2000

Department of Mathematics

University College London

Supervisor: Prof. Keith M. Ball

1

ProQuest Number: U643573

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.
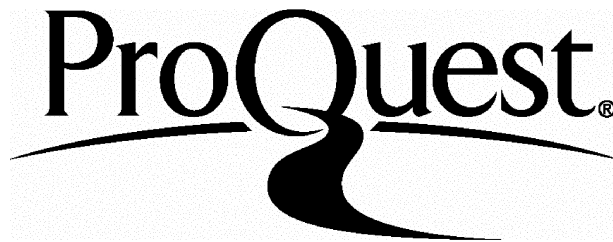
Abstract

It is shown that a large class of convex bodies satisfy a central limit property. In particular we show that if an isotropic, symmetric, convex body, K, has the property that most of its mass concentrates near its average radius, then its marginal distribution in direction θ, (whose density is given by scanning across K with hyperplanes perpendicular to θ), is close to a Gaussian in most directions. This closeness is shown in terms of the distribution function and the density function.

We also show how the transportation method for obtaining concentration results works for the cube and, more importantly, we find the best constant possible using this method. This constant turns out to be better than those obtained by traditional methods and cannot be far from that which is best possible.

Acknowledgements

Contents

Chapter 2

A Local Central Limit Property   42

Chapter 3

A Concentration Estimate for the Cube   58

Bibliography   77

List of Figures

Introduction

The principal question that we ask in this piece of work is whether symmetric, convex bodies in $\mathbf{R}^n$ exhibit a central limit property.

Let K be a symmetric, convex body of volume 1. Then we can regard K as a probability space with probability measure the Lebesgue measure, P, in K. For each unit direction $\theta \in S^{n-1}$, we consider the random variable on K given by $X_\theta : x \mapsto \langle x, \theta \rangle$. It can be seen that the density of X_θ is obtained by scanning across K with hyperplanes perpendicular to θ, if we observe that the probability that X_θ is less than a value t is the volume of K to one side of the slice $K \cap (\langle \theta \rangle^\perp + t\theta)$.

$$P(\langle x, \theta \rangle \leq t) = \int_{-\infty}^{t} \mathrm{vol}_{n-1}\left(K \cap \left(\langle \theta \rangle^\perp + s\theta\right)\right) \, ds.$$

So if we denote the density of X_θ, the "marginal" density of K in direction

θ, by g_θ, we have

$$g_\theta(s) = \mathrm{vol}_{n-1}\left(K \cap \left(\langle\theta\rangle^\perp + s\theta\right)\right).$$

The central limit problem asks whether the random variables X_θ are approximately Gaussian in an appropriate sense. It has been widely believed for some time that such marginal densities typically have Gaussian decay. We are concerned with whether they are really like Gaussians.

The starting point for this work was a result for a particular class of symmetric, convex bodies, namely the l_p^n balls. Recall that the unit l_p^n ball is the set $\{x \in \mathbf{R}^n : \sum_{i=1}^n |x_i|^p \leq 1\}$. If we take K to be an l_p^n ball, rescaled so as to have volume 1 , then the variance of X_θ can be shown to be the same for each direction θ. Moreover, as $n \to \infty$, this common variance, denoted by ρ_n^2, tends to a fixed number, ρ^2, depending only upon p. Perissinaki shows in [15] that for a given p, the average of the marginal densities over the sphere tends to a Gaussian density with the same variance, ρ^2. In other words as $n \to \infty$

$$\int_{S^{n-1}} g_\theta(t)\, d\sigma(\theta) \to \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{t^2}{2\rho^2}\right), \tag{1}$$

for all $t \in \mathbf{R}$. (We take σ to denote the rotation invariant probability measure on the sphere, $S^{n-1} = \{x \in \mathbf{R}^n : \sum x_i^2 = 1\}$).

8

The convergence in the above result relies on each X_θ having the same variance and on this variance, ρ_n^2, having a fixed bound for all n. Clearly this is not the case for all symmetric convex bodies - the variance of X_θ is certainly not the same for each θ when K is a long, thin body, for example.

We restrict our attention to isotropic bodies, i.e. ones for which there is some fixed ρ such that the variance of X_θ is equal to ρ^2 for all θ:

$$\int_K \langle x, \theta \rangle^2 \, dx = \rho^2 \quad \text{for all } \theta.$$

(We remark that each K has an affine image which is isotropic). From this point forward we refer to the Gaussian r.v. with variance ρ^2 as γ and denote its density by g so that

$$g(s) = \frac{1}{\rho\sqrt{2\pi}} \exp\left(-\frac{s^2}{2\rho^2}\right).$$

In the first two chapters of the current work we extend Perissinaki's result. Rather than considering the spherical average of the marginal densities, we show that *in most directions* X_θ is close to the Gaussian, γ. (It is clear that g_θ is not like a Gaussian density for every direction, θ, if we consider the example in which K is a cube and θ is perpendicular to a facet). The result will be shown to hold not only for the l_p^n balls, but for a large class of bodies which we believe includes all symmetric, convex bodies.

9

The first main result of this work says that provided most of K lies near its average radius, (a property stated as the Concentration Hypothesis below), then in most directions and for every positive t, the probability $P(|X_\theta| < t)$ differs from $P(|\gamma| < t)$ by very little:

Concentration Hypothesis *For a given $\varepsilon < \frac{1}{2}$ we say that K satisfies the ε-concentration hypothesis if:*

$$P\left(\left|\frac{|x|}{\sqrt{n}} - \rho\right| > \varepsilon\rho\right) \leq \varepsilon \quad .$$

Theorem 1.1 *Under the ε-concentration hypothesis, for $\delta > 0$*

$$\sigma\left(\left\{\theta : \left|\int_{-t}^{t} g_\theta(s)\, ds - \int_{-t}^{t} g(s)\, ds\right| \leq \delta + 4\varepsilon + \frac{c}{\sqrt{n}} \quad \text{for all } t\right\}\right) \geq 1 - 4n\, e^{-\frac{n\delta^2}{50}}$$

A surprising and attractive feature of this result is that it does not require a bound on ρ. Strangely, the larger is ρ, the weaker the concentration hypothesis becomes and the value of ρ mysteriously disappears altogether from the final estimate.

The origin and purpose of the concentration hypothesis become clear once an outline of the proof of Theorem 1.1 is given. The main issue is to show that $\int_{-t}^{t} g_\theta(s)\, ds$ is the reciprocal of a norm restricted to the unit sphere. Once we know this we can estimate its size to give a Lipschitz estimate on this function and then use a standard concentration of measure result on the

10

sphere to show that for a given t, in most directions, $\int_{-t}^{t} g_\theta(s)\,ds$ is close to its spherical average. (So far the proof holds for all symmetric, convex bodies).

To show that this spherical average, $\int_{S^{n-1}} \int_{-t}^{t} g_\theta(s)\,ds\,d\sigma(\theta)$, is close to the integral of g, we identify in Perissinaki's argument the property of l_p^n balls which ensures that the spherical average of the densities is close to g. This property is precisely the one stated in the concentration hypothesis. We remark that although the hypothesis says that most of K concentrates in a thin spherical shell of radius $\sqrt{n}\rho$, it does not tell us that K looks like a spherical shell. But since our initial aim is to estimate an *average* over the sphere, it does not matter how mass is distributed within the shell.

We discuss the concentration hypothesis in more detail and show that it holds for a large class of convex bodies in Chapter 1.

Having obtained a weak-type central limit property it is natural to ask whether the densities are locally close for most θ, i.e. whether for most θ, $g_\theta(x)$ is close to $g(x)$ for all $x \in \mathbf{R}$. We cannot use the same method as above to prove such a result since, unlike its integral, $g_\theta(x)$ is not Lipschitz. However in Chapter 2 we prove such a local central limit property using Theorem 1.1 and the fact that g_θ is log-concave:

11

Theorem 2.1 *There are constants, c, for which, under the ε-concentration hypothesis, for $\delta > 0$ and $\alpha = \sqrt{2\pi}\rho\left(\delta + \frac{c}{\sqrt{n}} + c\varepsilon\right)$, assuming α is less than $\frac{1}{10}$,*

$$\sigma\left(\left\{\theta : |g_\theta(x) - g(x)| \leq c\left(\alpha \log \frac{1}{\alpha}\right)^{\frac{1}{4}} \quad \text{for all } x\right\}\right) \geq 1 - 6n\, e^{-\frac{n\delta^2}{c}}.$$

Unfortunately, in order to obtain an estimate, here it is necessary to assume a bound on the variance, ρ^2.

In Chapter 3 we return to the problem of enlarging the class of bodies for which we know the concentration hypothesis to hold. In an attempt to find additional bodies, various existing methods of obtaining concentration of measure estimates were studied. Recall that a normalised K is said to possess a concentration of measure property if the mass in K concentrates around subsets of volume $\frac{1}{2}$. Such a concentration of measure result would give the concentration hypothesis since it implies that mass concentrates near the median radius and that this median value of $|x|$ is close to the mean radius, $\sqrt{n}\rho = \left(\int_K |x|^2\, dx\right)^{\frac{1}{2}}$.

One particular method of obtaining concentration of measure results, the "Transportation Method", was looked at in detail. However, attempts to

12

extend the ideas therein to non-product spaces did not work in our situation. We show in Chapter 3 how the method works for the cube and, more importantly, we find the best constant possible using this method; a constant which turns out to be better than those obtained by traditional methods and which cannot be far from that which is best possible.

In the first two chapters, for the sake of clarity, we often replace constants appearing in expressions by c. Therefore, unless otherwise stated, c will vary throughout these chapters to denote the appropriate constant where it appears.

Chapter 1

A Weak-Type Central Limit

Property for Convex Bodies

Let K be a symmetric, convex body of volume 1. We regard K as a probability space with probability measure the Lebesgue measure, P, in K. For each unit direction θ, we define the random variable $X_\theta : x \mapsto \langle x, \theta \rangle$: so the density of X_θ is obtained by scanning across K with hyperplanes perpendicular to θ. Now suppose that K is *isotropic*, i.e. that for some fixed ρ

$$\int_K \langle x, \theta \rangle^2 \, dx = \rho^2 \quad \text{for all } \theta.$$

(We remark that each K has an affine image which is isotropic.) Then each of the random variables, X_θ, has variance ρ^2.

Our aim is to show that most of these $r.v.$s are very close to a Gaussian $r.v.$, γ, with variance ρ^2. We shall prove this under the following hypothesis which states that the Euclidean norm concentrates near the value $\sqrt{n}\rho$, as a function on K.

Concentration Hypothesis *For a given $\varepsilon < \frac{1}{2}$ we say that K satisfies the ε-concentration hypothesis if:*

$$P\left(\left|\frac{|x|}{\sqrt{n}} - \rho\right| > \varepsilon\rho\right) \le \varepsilon \quad . \tag{1.1}$$

Under the above hypothesis, we shall show that if $\delta > 0$, then except for a set of directions of small spherical measure, for every positive t, the probability $P(|X_\theta| < t)$ differs from $P(|\gamma| < t)$ by at most

$$\delta + 4\varepsilon + \frac{c}{\sqrt{n}}.$$

More precisely, we prove

Theorem 1.1 *Under the ε-Concentration Hypothesis, for $\delta > 0$*

$$\sigma\left(\left\{\theta : \left|\int_{-t}^{t} g_\theta(s)\, ds - \int_{-t}^{t} g(s)\, ds\right| \le \delta + 4\varepsilon + \frac{c}{\sqrt{n}} \quad \text{for all } t\right\}\right) \ge 1 - 4n\, e^{-\frac{n\delta^2}{50}}.$$

We begin with a brief overview of this chapter and then continue to prove Theorem 1.1.

At first sight, the concentration hypothesis (1.1) looks, at once, too strong to be true, except in trivial cases, and too weak to be useful.

On the one hand, an estimate like (1.1), with a small value of ε, is considerably stronger than the estimate which follows from Borell's inequality [6]. On the other hand, although (1.1) states that most of K lies in a thin spherical shell, K will almost always occupy only a miniscule fraction of this shell. So the condition cannot automatically guarantee that K "looks like" a spherical shell.

We deal with the second point by using standard concentration methods on the sphere together with a Lipschitz estimate that depends ultimately upon a version of the Brunn-Minkowski inequality.

Concerning the first point, we will describe how (1.1) holds with a small ε for two classes of bodies. These classes exclude pathological bodies but together include a significant portion of "nice" convex bodies. The first is the class of l_p^n balls for which Perissinaki showed (1.1) to hold with $\varepsilon \approx \frac{1}{n^{\frac{1}{3}}}$. In Section 1.2 we outline her proof which uses the subindependence of complements of coordinate slabs in the l_p^n ball. A concentration result of Gromov and Milman can be used to show that (1.1) holds for a second class of bodies and an argument is detailed in Section 1.3. This class is somewhat ad

16

hoc, consisting of uniformly convex bodies which have the additional property of being contained in a Euclidean ball of appropriate radius, the necessary radius being dependent upon the modulus of convexity of K. However, this class encompasses all l_p^n balls for $1 < p < \infty$, so it is fairly broad.

In the ensuing discussion we have chosen, for the sake of clarity, to separate the abstract part, which holds for all convex, symmetric bodies under the concentration hypothesis (1.1), from the proofs of the hypothesis for specific bodies.

We should remark that the above results are very much in the same spirit as results of Diaconis and Freedman [10], Sudakov [18] and von Weizsäcker [23]. These show, in a general probabilistic setting, that a kind of weak law of large numbers implies that most marginals are approximately Gaussian under conditions which correspond in our setting to a restriction to isotropic bodies with ρ bounded. We do not assume their moment condition. Indeed ρ "magically" cancels out in our exposition. However, perhaps more importantly, in the case of convex bodies we achieve much finer probability estimates than for the general case.

Independent work in this area has been done by Brehm, Vogt and Voigt who also recognised that the concentration hypothesis is the correct property

17

to consider when aiming for a central limit theorem. Since the concentration property can be derived from bounds on the second and fourth moments, they study such inequalities and find them to be persistent under the operations of forming isotropic normed cones, cartesian products, joins and p-products and hence they prove the concentration hypothesis for a wide class of bodies, [22], [7]. They also obtain fine estimates for the central limit theorem for the cube and regular simplex in [8].

1.1 Proof of Theorem 1.1

The proof of Theorem 1.1 is composed of three main steps. We begin by considering not the individual X_θ and g_θ, but an average of the integral of g_θ over all θ :

$$A(t) = \int_{S^{n-1}} \int_{-t}^{t} g_\theta(s) \, ds \, d\sigma(\theta) \quad .$$

This averaging enables us to ignore how the volume of K is distributed within the relevant spherical shell. We approximate $A(t)$ by an integral over K, of an integral of densities of Gaussian random variables with different variances. The concentration hypothesis is then used to show that most of the Gaussians

have about the same variance. This will ensure that

$$\left| A(t) - \int_{-t}^{t} g(s)\, ds \right| \le 4\varepsilon + \frac{c_1}{\sqrt{n}} \quad .$$

The second step does not use the concentration hypothesis at all. The most important issue is to show that $\int_{-t}^{t} g_\theta(s)\, ds$ is the reciprocal of a norm restricted to the sphere. Once we know this we need only estimate its size to get a Lipschitz estimate on this function. Then applying a standard concentration of measure result, we get that for each t, in most directions, the distribution functions, $P(|X_\theta| \le t)$, are essentially the same regardless of whether the body satisfies hypothesis (1.1):

$$\sigma\left(\left\{ \theta : \left| \int_{-t}^{t} g_\theta(s)\, ds - A(t) \right| > \delta + \frac{c_2}{\sqrt{n}} \right\} \right) \le 2e^{-\frac{n\delta^2}{50}}.$$

Combining this with step 1 which told us that $A(t)$ is close to the integral of g gives

$$\sigma\left(\left\{ \theta : \left| \int_{-t}^{t} g_\theta(s)\, ds - \int_{-t}^{t} g(s)\, ds \right| > \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}} \right\} \right) \le 2e^{-\frac{n\delta^2}{50}}$$

for each positive δ.

Finally, by dividing the real line into intervals of appropriate width and using the Lipschitz property of the function

$$H(t) = \left| \int_{-t}^{t} g_\theta(s)\, ds - \int_{-t}^{t} g(s)\, ds \right|$$

we obtain a result for every t simultaneously.

19

1.1.1 The spherical average of $P(|X_\theta| \leq t)$

In this section we estimate the spherical average, $A(t)$, of the probability $P(|X_\theta| \leq t)$. Indeed we show that for any fixed t, $A(t)$ is close to the integral of the Gaussian density, g, provided K satisfies the concentration hypothesis. The method here is similar to that used by Perissinaki to show that for l_p^n balls the spherical average of the marginal densities tends to g.

We begin with a simple geometric lemma which approximates the average $A(t)$ by an integral over K.

Lemma 1.1

$$\left| A(t) - \frac{2}{\sqrt{2\pi}} \int_K \int_0^{\frac{t\sqrt{n}}{|x|}} e^{-\frac{v^2}{2}} \, dv \, dx \right| \leq \frac{c_1}{\sqrt{n}}$$

Proof: If v is a unit vector in $\mathbf{R}^n$, then

$$\sigma\left(\{\theta : |\langle \theta, v \rangle| \leq t\}\right) = \frac{\int_0^t (1-u^2)^{\frac{n-3}{2}} du}{\int_0^1 (1-u^2)^{\frac{n-3}{2}} du} \qquad .$$

We then have:

$$
\begin{aligned}
\int_{S^{n-1}} \int_{-t}^t g_\theta(s) ds d\sigma(\theta) &= \int_{S^{n-1}} \int_{-t}^t vol_{n-1}\left(K \cap \left(\langle \theta \rangle^\perp + s\theta\right)\right) ds d\sigma(\theta) \\
&= \int_{S_{n-1}} \int_K \mathbf{1}_{\{-t \leq \langle x, \theta \rangle \leq t\}} dx d\sigma(\theta) \\
&= \int_K \sigma\left(\left\{\theta : -\frac{t}{|x|} \leq \langle \frac{x}{|x|}, \theta \rangle \leq \frac{t}{|x|}\right\}\right) dx \\
&= \int_K \frac{\int_0^{\min\left\{1, \frac{t}{|x|}\right\}} (1-u^2)^{\frac{n-3}{2}} du}{\int_0^1 (1-u^2)^{\frac{n-3}{2}} du} dx \qquad (1.2)
\end{aligned}
$$

20

It is not difficult to obtain the following estimates for the denominator in the integrand. (The upper bound is obtained by estimating the integrand by the exponential function and the lower bound by estimating the square of the integral and changing to polar coordinates).

$$\frac{\sqrt{2\pi}}{2\sqrt{n-1}} \leq \int_0^1 \left(1 - u^2\right)^{\frac{n-3}{2}} du \leq \frac{\sqrt{2\pi}}{2\sqrt{n-3}}.$$

These estimates can then be used first to show that (1.2) differs from

$$\int_K \frac{2\sqrt{n}}{\sqrt{2\pi}} \int_0^{\min\left\{1, \frac{t}{|x|}\right\}} \left(1 - u^2\right)^{\frac{n-3}{2}} du\, dx \tag{1.3}$$

by at most $\frac{c}{\sqrt{n}}$ and then to show that (1.3) differs from the required integral over K by at most $\frac{c}{\sqrt{n}}$. □

From now on, let

$$F(s) = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{t}{s}} e^{-\frac{v^2}{2}}\, dv$$

denote the integral of the Gaussian density with variance s^2. To show that $A(t)$ is close to $\int_{-t}^t g(s)\, ds$, clearly we need to show that the average $\int_K F\left(\frac{|x|}{\sqrt{n}}\right) dx$ in Lemma 1.1 is close to $F(\rho)$. Here we invoke the concentration hypothesis.

We divide K into two subsets: K_1, where $\frac{|x|}{\sqrt{n}}$ is within $\varepsilon\rho$ of ρ, and its complement in K, K_2. Since we find $F(s)$ to be Lipschitz with constant $\frac{2}{\rho}$

near ρ. $F\left(\frac{|x|}{\sqrt{n}}\right)$ is within 2ε of $F(\rho)$ in K_1. The volume of K_2 is sufficiently

small for it not to matter how far apart the functions are here.

Let

$$
\begin{aligned}
K_1 &= K \cap \left\{ \left| \frac{|x|}{\sqrt{n}} - \rho \right| \leq \varepsilon\rho \right\} \\
K_2 &= K \cap \left\{ \left| \frac{|x|}{\sqrt{n}} - \rho \right| > \varepsilon\rho \right\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
&\left| \int_K F\left(\frac{|x|}{\sqrt{n}}\right) \, dx - F(\rho) \right| \\
&\leq \int_{K_1} \left| F\left(\frac{|x|}{\sqrt{n}}\right) - F(\rho) \right| \, dx + \int_{K_2} \left| F\left(\frac{|x|}{\sqrt{n}}\right) - F(\rho) \right| \, dx.
\end{aligned}
$$

To estimate the second integral, we need only recall that $F\left(\frac{|x|}{\sqrt{n}}\right)$ and

$F(\rho)$ are at most one. Therefore, by the concentration hypothesis,

$$
\int_{K_2} \left| F\left(\frac{|x|}{\sqrt{n}}\right) - F(\rho) \right| \, dx \leq 2|K_2| \leq 2\varepsilon.
$$

For the first integral we shall use a Lipschitz estimate for F. The derivative, $|F'(s)|$, is bounded by $\frac{1}{s}$ so, provided $s > \frac{\rho}{2}$, we have a bound of order ρ^{-1}. In K_1, $\frac{|x|}{\sqrt{n}} > \frac{\rho}{2}$ since $\varepsilon < \frac{1}{2}$ and therefore

$$
\begin{aligned}
\int_{K_1} \left| F\left(\frac{|x|}{\sqrt{n}}\right) - F(\rho) \right| \, dx &\leq \int_{K_1} \frac{2}{\rho} \left| \frac{|x|}{\sqrt{n}} - \rho \right| \, dx \\
&\leq \int_{K_1} 2\varepsilon \, dx \\
&\leq 2\varepsilon.
\end{aligned}
$$

22

So we have

$$\left| \int_K F\left(\frac{|x|}{\sqrt{n}} \right) dx - F(\rho) \right| \leq 4\varepsilon.$$

Combining this with Lemma 1.1 we get

$$\left| A(t) - \int_{-t}^{t} g(s)ds \right| \leq 4\varepsilon + \frac{c_1}{\sqrt{n}} \quad \square$$

1.1.2 The Lipschitz property of $P(|X_\theta| \leq t)$

The problem is now to pass from an estimate for the average, $A(t)$, to an estimate for specific directions. We show that in most directions $\int_{-t}^{t} g_\theta(s)ds$ is close to $A(t)$ regardless of whether K satisfies the concentration hypothesis. Then we combine this with the result of the previous section to get that under the concentration hypothesis, for each positive t and δ,

$$\sigma\left(\left\{ \theta : \left| \int_{-t}^{t} g_\theta(s)ds - \int_{-t}^{t} g(s)ds \right| > \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}} \right\} \right) \leq 2e^{-\frac{n\delta^2}{50}}.$$

Central to the proof is a standard concentration of measure result of the type developed by Milman and others, based upon Lévy's isoperimetric inequality on the sphere. A simple exposition of this kind of result can be found in [2].

Lemma 1.2 *If* $f : S^{n-1} \to \mathbf{R}$ *is 1-Lipschitz and* M *is its median, then*

$$\sigma\left(\{ \theta : |f(\theta) - M| > \delta \} \right) \leq 2e^{-\frac{n\delta^2}{2}}$$

23

The above tells us that the median is close to the mean, m, by the following simple argument:

$$\begin{aligned}
|m - M| &= \left| \int_{S^{n-1}} f(\theta)\, d\sigma(\theta) - M \right| \\
&\leq \int_{S^{n-1}} |f(\theta) - M|\, d\sigma(\theta) \\
&= \int_0^\infty \sigma(\{\theta : |f(\theta) - M| > t\})\, dt \\
&\leq 2 \int_0^\infty e^{-\frac{nt^2}{2}}\, dt = \sqrt{\frac{2\pi}{n}}.
\end{aligned}$$

Hence we obtain a similar result for the mean:

Corollary 1.1 *If $f : S^{n-1} \to \mathbf{R}$ is 1-Lipschitz and m is its mean, then*

$$\sigma \left(\left\{ \theta : |f(\theta) - m| > \delta + \frac{c_2}{\sqrt{n}} \right\} \right) \leq 2 e^{-\frac{n\delta^2}{2}}$$

Obviously we shall take $\int_{-t}^t g_\theta(s)ds$ to be our Lipschitz function of θ and so our m will be $A(t)$, which we already know from section 1.1.1 to be close to the required $\int_{-t}^t g(s)\, ds$. To obtain the Lipschitz estimate we shall show that $\int_{-t}^t g_\theta(s)ds$ is the reciprocal of a norm (restricted to S^{n-1}). To do this, we apply the following Theorem of Busemann to a certain convex body in $\mathbf{R}^{n+1}$ constructed from K.

Theorem 1.2 (Busemann's Theorem) *Let C be a symmetric convex body in $\mathbf{R}^n$, and for each unit vector u let $r(u)$ be the volume of the slice of C by*

24

the subspace orthogonal to u. *Then the body whose radius in each direction*

u *is* $r(u)$, *is itself convex.*

Lemma 1.3 below is closely related to the so-called "convexity of the floating

body" which was proved simultaneously by Meyer and Reisner [14] and by

Ball. As in the latter's argument we apply Busemann's theorem to $K \times [-1, 1]$

but the difference here is that the earlier proofs involved slabs of fixed volume,

whereas here we fix the slab width.

Lemma 1.3 *For all positive* t,

$$\|x\| = \frac{|x|}{\int_{-t}^{t} g_{\frac{x}{|x|}}(s)ds}$$

defines a norm on $\mathbf{R}^n$.

Proof: Let us first recall that $g_{\frac{x}{|x|}}(s) = \left| K \cap \left(\frac{x}{|x|}^{\perp} + s\frac{x}{|x|} \right) \right|$. We will denote

the volume of the slab of K perpendicular to x and of width $2t$,

$$v(x, t) = \int_{-t}^{t} g_{\frac{x}{|x|}}(s)ds.$$

Our aim is thus to prove the following triangle inequality for all $x, y \ \epsilon \ \mathbf{R}^n$:

$$\frac{1}{2} \left(\frac{|x|}{v(x,t)} + \frac{|y|}{v(y,t)} \right) \geq \frac{\left| \frac{x+y}{2} \right|}{v \left(\frac{x+y}{2}, t \right)} \quad .$$

Notice that the result is obvious when the angle between x and y is zero or

π.

We consider the convex body

$$K' = K \times [-1, 1] \quad \subset \quad \mathbf{R}^n \times \mathbf{R} = \mathbf{R}^{n+1},$$

shown in Figure 1.1. Busemann's Theorem tells us that, since K' is a symmetric, convex body in $\mathbf{R}^{n+1}$, then $\frac{|\theta|}{|K' \cap \theta^\perp|}$ defines a norm on $\mathbf{R}^{n+1}$. Hence

$$\frac{1}{2} \left(\frac{|\theta|}{|K' \cap \theta^\perp|} + \frac{|\phi|}{|K' \cap \phi^\perp|} \right) \geq \frac{\left| \frac{\theta+\phi}{2} \right|}{\left| K' \cap \left(\frac{\theta+\phi}{2} \right)^\perp \right|} \quad (1.4)$$

for all $\theta, \phi \in \mathbf{R}^{n+1}$.

Given $x \in \mathbf{R}^n$ and $t = \frac{\sqrt{1-r^2}}{r}$ where $r > 0$, if we choose $u \in \mathbf{R}^{n+1}$ as

$$u = \left(r \frac{x}{|x|}, \sqrt{1 - r^2} \right)$$

then the projection of $K' \cap u^\perp$ onto the first n coordinates is precisely the slab of K perpendicular to x and of width $2t$. The ratio of $v(x, t)$ to the volume $|K' \cap u^\perp|$ is then just $\sqrt{1 - r^2}$. (Observe that r determines the angle between $u^\perp$ and $\mathbf{R}^n$).

Now take $\theta = |x| u$ so that $|K' \cap u^\perp| = |K' \cap \theta^\perp|$ and $|\theta| = |x|$. Then inequality (1.4) simplifies to

$$\frac{1}{2} \left(\frac{|x|}{v(x, t)} + \frac{|y|}{v(y, t)} \right) \geq \frac{1}{\sqrt{1 - r^2}} \frac{\left| \frac{\theta+\phi}{2} \right|}{\left| K' \cap \left(\frac{\theta+\phi}{2} \right)^\perp \right|}$$

where we have repeated the above for $y \in \mathbf{R}^n$ with ϕ.

Figure 1.1: $K' = K \times [-1, 1]$
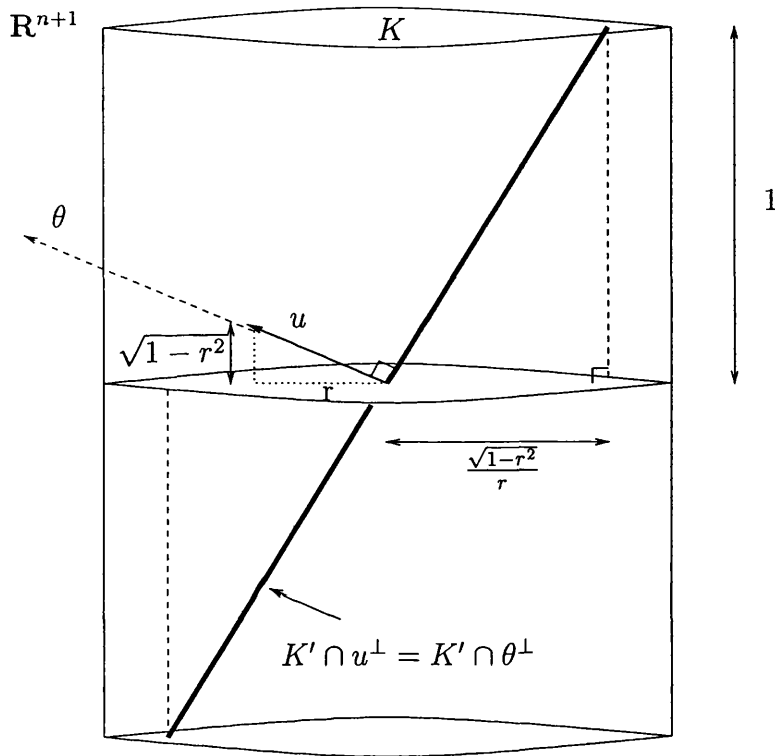
Now

$$\frac{\theta + \phi}{2} = \left(r \, \frac{(x + y)}{2}, \sqrt{1 - r^2} \, \frac{(|x| + |y|)}{2} \right).$$

Hence the projection of $K' \cap \left(\frac{\theta+\phi}{2}\right)^{\perp}$ is a slab perpendicular to $\frac{x+y}{2}$ whose width now depends on r, $\left|\frac{x+y}{2}\right|$ and $\frac{|x|+|y|}{2}$. In fact the width is $2s$ where

$$s = \frac{|x| + |y|}{|x + y|} t$$

and we get

$$\frac{v\left(\frac{x+y}{2}, s\right)}{\left|K' \cap \left(\frac{\theta+\phi}{2}\right)^{\perp}\right|} = \frac{\sqrt{1-r^2}}{2} \frac{(|x|+|y|)}{\left|\frac{\theta+\phi}{2}\right|} \quad .$$

Inequality (1.4) thus simplifies further to

$$\frac{1}{2}\left(\frac{|x|}{v(x,t)} + \frac{|y|}{v(y,t)}\right) \geq \frac{1}{2}\frac{(|x|+|y|)}{v\left(\frac{x+y}{2}, s\right)} \quad .$$

Now we need only notice that for any $a \geq 1$ and $x \in R^n$, $v(x, at) \leq a\,v(x,t)$

since $g_{\frac{x}{|x|}}(as) \leq g_{\frac{x}{|x|}}(s)$ for all s. So

$$\frac{1}{2}\frac{(|x|+|y|)}{v\left(\frac{x+y}{2}, s\right)} \geq \frac{\left|\frac{x+y}{2}\right|}{v\left(\frac{x+y}{2}, t\right)} \quad \square$$

Since $\int_{-t}^{t} g_\theta(s)\,ds$ is the reciprocal of a norm, we can get a Lipschitz

estimate for this function, just by estimating its size. This we do in the

following lemmas. The first follows from the fact that g_θ is a log-concave

function, (a well-known consequence of the Brunn-Minkowski inequality). A

proof of Lemma 1.4 can be found, for example, in [3].

Lemma 1.4 *For all positive t,*

$$\int_{t}^{\infty} g_\theta(s)ds \leq \frac{1}{2}e^{-2g_\theta(0)t}$$

Lemma 1.5 *For each t there are constants a and b so that if $||x|| = \dfrac{|x|}{\int_{-t}^{t} g_{\frac{x}{|x|}}(s)\,ds}$,*

then

$$a|x| \leq ||x|| \leq b|x|$$

where $\frac{b}{a} \leq 5$ and $\frac{1}{a} \leq 1$.

Proof: We shall use the two following well-known facts about log-concave, even functions which are decreasing on $[0, \infty)$ to relate the value $g_\theta(0)$ to the variance, ρ^2:

$$f(0)^2 \int_0^\infty f(x)x^2\,dx \leq 2\left(\int_0^\infty f(x)\,dx\right)^3$$

and

$$f(0)\left(\int_{-\infty}^\infty x^2 f(x)\,dx\right)^{\frac{1}{2}} \geq \frac{1}{\sqrt{12}} \quad \text{provided} \quad \int_{-\infty}^\infty f(x)\,dx = 1.$$

Substituting g_θ for f, it quickly follows that

$$\frac{1}{\sqrt{12}\rho} \leq g_\theta(0) \leq \frac{1}{\sqrt{2}\rho}.$$

Using the right hand estimate and the fact that an even, log-concave function has its maximum at zero we get

$$\int_{-t}^{t} g_\theta(s)\,ds \leq \min\{2t\,g_\theta(0), 1\} \leq \min\left\{\frac{\sqrt{2}t}{\rho}, 1\right\}$$

29

and using the left hand estimate alongside Lemma 1.4 we have

$$\int_{-t}^{t} g_\theta(s)\,ds \;=\; 1 - 2\int_{t}^{\infty} g_\theta(s)\,ds$$

$$\geq \; 1 - e^{-2g_\theta(0)t}$$

$$\geq \; 1 - e^{-\frac{t}{\sqrt{3}\rho}}$$

$$\geq \; \min\left\{\frac{t}{3\rho}, \frac{1}{3\sqrt{2}}\right\}.$$

Therefore, taking

$$a \;=\; \frac{1}{\min\left\{\frac{\sqrt{2}t}{\rho}, 1\right\}}$$

$$b \;=\; \frac{1}{\min\left\{\frac{t}{3\rho}, \frac{1}{3\sqrt{2}}\right\}}$$

we obtain the required bounds, whatever the value of $\frac{t}{\rho}$. (This is desirable as it stops t and ρ from entering our estimate at this stage). $\square$

We are now ready to obtain the estimate for specific directions. As was explained earlier, we need to determine a Lipschitz estimate for $\int_{-t}^{t} g_\theta(s)\,ds$, i.e. to find a constant d such that

$$\left|\int_{-t}^{t} g_\theta(s)\,ds - \int_{-t}^{t} g_\phi(s)\,ds\right| \leq d|\theta - \phi| \qquad \text{for } \theta, \phi \,\epsilon\, S^{n-1}.$$

Now, with the norm notation used above,

$$\left|\int_{-t}^{t} g_\theta(s)\,ds - \int_{-t}^{t} g_\phi(s)\,ds\right| \;=\; \left|\frac{1}{||\theta||} - \frac{1}{||\phi||}\right|$$

$$\leq \frac{\|\theta - \phi\|}{\|\theta\| \, \|\phi\|}$$

$$\leq \frac{b}{a^2} |\theta - \phi|$$

$$\leq 5|\theta - \phi|$$

by Lemma 1.5.

Coupled with this Lipschitz estimate, Corollary 1.1 immediately gives

$$\sigma\left(\left\{\theta : \left|\int_{-t}^{t} g_\theta(s)\, ds - A(t)\right| > \delta + \frac{c_2}{\sqrt{n}}\right\}\right) \leq 2\exp\left(-\frac{n\delta^2}{50}\right).$$

Combining this with the result of section 1.1.1, which estimates $A(t)$, we get

$$\sigma\left(\left\{\theta : \left|\int_{-t}^{t} g_\theta(s)ds - \int_{-t}^{t} g(s)ds\right| > \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}}\right\}\right) \leq 2\exp\left(-\frac{n\delta^2}{50}\right)$$

as required. □

1.1.3   A result for every t simultaneously

To finish this section, we pass to a statement which holds for every t simultaneously. This then tells us that most X_θ are close to a Gaussian r.v. with variance ρ^2.

For a given θ, let

$$H(t) = \left|\int_{-t}^{t} g_\theta(s)\, ds - \int_{-t}^{t} g(s)\, ds\right|$$

be the error at position t.

We saw in the proof of Lemma 1.5 that $g_\theta(s)$ is bounded above by $\frac{1}{\sqrt{2}\rho}$ and we recall that $g(s)$ is the Gaussian density $\frac{1}{\rho\sqrt{2\pi}}e^{-\frac{s^2}{2\rho^2}}$. Hence

$$|H'(t)| \le \frac{\sqrt{2}}{\rho} + \frac{2}{\rho\sqrt{2\pi}}.$$

So H is Lipschitz with a constant like $\frac{1}{\rho}$.

By pinning H down at appropriate points we can use the Lipschitz property to pin it down elsewhere as long as we allow an additional error. We need only consider points in $[0, 2\rho\log n]$ since we see, with the aid of Lemma 1.4, that H is sufficiently small for all t beyond this interval. Dividing this interval into $2\sqrt{n}\log n$ smaller intervals of length $\frac{\rho}{\sqrt{n}}$ gives an additional error of $\left(\sqrt{2} + \frac{2}{\sqrt{2\pi}}\right)\frac{1}{\sqrt{n}}$ which is absorbed by the earlier error terms. We thus give up a factor of $2\sqrt{n}\log n$ in the probability:

Let us denote the points at which we pin H down by t_i. So $H(t_i) \le \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}}$, $t_i \le 2\rho\log n$ and $(t_{i+1} - t_i) = \frac{\rho}{\sqrt{n}}$ for all i. Since H is Lipschitz, for any $t \in (t_i, t_{i+1})$,

$$H(t)$$
$$= \left| \int_{-t}^{t} g_\theta(s)ds - \int_{-t}^{t} g(s)ds + \left(\int_{-t_i}^{t_i} g_\theta(s)ds - \int_{-t_i}^{t_i} g_\theta(s)ds \right) + \left(\int_{-t_i}^{t_i} g(s)ds - \int_{-t_i}^{t_i} g(s)ds \right) \right.$$
$$\le \left(\frac{\sqrt{2}}{\rho} + \frac{2}{\rho\sqrt{2\pi}} \right)(t_{i+1} - t_i) + \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}}$$

$$= \left(\sqrt{2} + \frac{2}{\sqrt{2\pi}} \right) \frac{1}{\sqrt{n}} + \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}}.$$

Therefore

$$\sigma \left(\left\{ \theta : H(t) \leq \delta + 4\varepsilon + \frac{c}{\sqrt{n}} \quad \text{for all } t \right\} \right)$$

$$\geq \sigma \left(\left\{ \theta : H(t_i) \leq \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}} \quad \text{for all } i \right\} \right)$$

$$= 1 - \sigma \left(\left\{ \theta : H(t_i) > \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}} \quad \text{for some } i \right\} \right)$$

$$\geq 1 - \sum_i \sigma \left(\left\{ \theta : H(t_i) > \delta + 4\varepsilon + \frac{c_3}{\sqrt{n}} \right\} \right)$$

$$\geq 1 - (\text{number of intervals}) 2 e^{-\frac{n\delta^2}{50}}$$

$$= 1 - 4\sqrt{n} \log n \, e^{-\frac{n\delta^2}{50}}$$

Clearly the aim here was to choose interval widths to shrink the upper bound on $H(t)$ whilst keeping the number of intervals small. So

Under the concentration hypothesis (1.1), for $\delta > 0$ we have,

$$\sigma \left(\left\{ \theta : H(t) \leq \delta + 4\varepsilon + \frac{c}{\sqrt{n}} \quad \text{for all } t \right\} \right) \geq 1 - 4\sqrt{n} \log n \, e^{-\frac{n\delta^2}{50}}$$

$$\geq 1 - 4n \, e^{-\frac{n\delta^2}{50}}$$

1.2   The concentration property for l_p^n balls

In this section we show that l_p^n balls satisfy the concentration hypothesis (1.1) with $\varepsilon \approx \frac{1}{n^{\frac{1}{3}}}$. The proof here is a sketch of that given by Perissinaki in

[15].

The precise statement is the following, in which ρ_n^2 is the variance for the appropriate ball.

Theorem 1.3 *If P is the Lebesgue measure on the normalized l_p^n ball, K, then for all positive numbers r,*

$$P\left(\left|\frac{|x|^2}{n} - {\rho_n}^2\right| \geq r\right) \leq \frac{35{\rho_n}^4}{nr^2}$$

The estimate depends upon a subindependence property for the complements of coordinate slabs in the l_p^n ball proved by Ball and Perissinaki in [5]. Here we offer a more succinct proof which was alluded to in the aforementioned article. However, as with the earlier proof, the argument depends on the property of the l_p^n ball that each slice perpendicular to a coordinate direction is also an l_p ball of dimension $n - 1$.

Lemma 1.6 (Subindependence of complements of coordinate slabs)

If P is the Lebesgue measure on the normalised l_p^n ball, K, then for any sequence $t_1, \ldots, t_n$ of positive numbers,

$$P\left(\bigcap_{i=1}^{n}\{|x_i| \geq t_i\}\right) \leq \prod_{i=1}^{n} P(\{|x_i| \geq t_i\})$$

34

Proof: Plainly it is enough to prove that

$$P\left(\bigcap_{i=1}^{n}\{|x_i| \geq t_i\}\right) \leq P\left(|x_1| \geq t_1\right) P\left(\bigcap_{i=2}^{n}\{|x_i| \geq t_i\}\right).$$

We first rewrite the above with $S = \bigcap_{i=2}^{n}\{|x_i| \geq t_i\}$:

$$\frac{|K \cap S \cap \{|x_1| \geq t_1\}|}{|K|} \leq \frac{|K \cap \{|x_1| \geq t_1\}|}{|K|} \frac{|K \cap S|}{|K|}.$$

Now we use the obvious fact that if $f : [0,1] \to \mathbf{R}$ is increasing and satisfies

$$\int_0^1 f d\mu = \alpha \, \mu[0,1]$$

for some positive measure μ and constant $\alpha > 0$, then

$$\int_0^s f d\mu \leq \alpha \, \mu[0,s] \qquad \text{for all } s \, \epsilon \, [0,1] \quad . \tag{1.5}$$

We take

$$f(u) = \frac{|K \cap S \cap \{|x_1| = 1 - u\}|}{|K \cap \{|x_1| = 1 - u\}|},$$

which is clearly increasing since each slice $K \cap \{|x_1| = 1 - u\}$ is an l_p^{n-1} ball,

and let

$$g(u) = \frac{|K \cap \{|x_1| = 1 - u\}|}{|K|}$$

be the density of our measure μ. Then

$$\begin{aligned}
\int_0^1 f(u) d\mu &= \int_0^1 \frac{|K \cap S \cap \{|x_1| = 1 - u\}|}{|K \cap \{|x_1| = 1 - u\}|} \frac{|K \cap \{|x_1| = 1 - u\}|}{|K|} du \\
&= \frac{|K \cap S|}{|K|}
\end{aligned}$$

35

and

$$\int_0^{1-t_1} f(u)d\mu = \frac{|K \cap S \cap \{|x_1| \ge t_1\}|}{|K|}.$$

Taking $\alpha = \frac{|K \cap S|}{|K|}$ and $s = 1 - t_1$ in (1.5) completes the proof. $\square$

The corollary below follows directly from the subindependence result,

Lemma 1.6, if we notice that

$$\int_{K \cap \{x_1 \ge 0, x_2 \ge 0\}} x_1^2 x_2^2 \quad \text{can be expressed as} \quad 4 \int_{\mathbf{R}_+^2} uv P(x_1 \ge u, x_2 \ge v) du dv .$$

Corollary 1.2

$$\int_K x_1^2 x_2^2 \le \int_K x_1^2 \int_K x_2^2$$

The proof of Theorem 1.3 relies on the fact that $\int_K |x|^4$ can be written in

terms of $\sum_1^n \int_K x_i^4$ and $\sum_{i \ne j} \int_K x_i^2 x_j^2$. The second term is dealt with using

subindependence via Corollary 1.2. For the first we use a cruder estimate

derived from standard results concerning log-concave functions:

$$\int_K x_i^4 \le 36 \left(\int_K x_i^2 \right)^2 .$$

The crudity doesn't matter since there are so few contributions to this term.

Proof of Theorem 1.3: We first prove that $\frac{1}{n^2} \int_K |x|^4$ is close to ρ_n^4; namely

that,

$$\rho_n^4 \le \frac{1}{n^2} \int_K |x|^4 \le \left(1 + \frac{35}{n} \right) \rho_n^4.$$

The first inequality is obvious by Cauchy-Schwarz.

For the second one we have:

$$
\begin{aligned}
\int_K |x|^4 &= \int_K \left(\sum_{i=1}^n x_i^2 \right)^2 \\
&= \sum_{i=1}^n \int_K x_i^4 + \sum_{i \neq j} \int_K x_i^2 x_j^2 \\
&\leq n \int_K x_i^4 + \sum_{i \neq j} \int_K x_i^2 \int_K x_j^2 \\
&\leq 36 n \rho_n^4 + n(n-1) \rho_n^4 \\
&= n^2 \left(1 + \frac{35}{n} \right) \rho_n^4.
\end{aligned}
$$

From this we can conclude that the integral $\int_K \left(\frac{|x|^2}{n} - \rho_n^2 \right)^2$ is small and

therefore that in general $\frac{|x|^2}{n}$ is close to ρ_n^2. Indeed,

$$
\begin{aligned}
0 \leq \int_K \left(\frac{|x|^2}{n} - \rho_n^2 \right)^2 &= \frac{1}{n^2} \int_K |x|^4 - \frac{2\rho_n^2}{n} \int_K |x|^2 + \rho_n^4 \\
&= \frac{1}{n^2} \int_K |x|^4 - \frac{2}{n} \rho_n^2 n \rho_n^2 + \rho_n^4 \\
&= \frac{1}{n^2} \int_K |x|^4 - \rho_n^4 \\
&\leq \frac{35}{n} \rho_n^4.
\end{aligned}
$$

So by Chebychev's inequality we have:

$$
P \left(\left| \frac{|x|^2}{n} - \rho_n^2 \right| \geq r \right) r^2 = P \left(\left(\frac{|x|^2}{n} - \rho_n^2 \right)^2 \geq r^2 \right) r^2
$$

37

$$\leq \int_K \left(\frac{|x|^2}{n} - \rho_n^2 \right)^2$$

$$\leq \frac{35}{n} \rho_n^4. \qquad \Box$$

By factorising the difference of two squares, we can estimate the deviation of $\frac{|x|}{\sqrt{n}}$ instead of $\frac{|x|^2}{n}$.

Corollary 1.3 *For all positive numbers u,*

$$P\left(\left| \frac{|x|}{\sqrt{n}} - \rho_n \right| \geq u \right) \leq \frac{35\rho_n^2}{nu^2}$$

By taking u to be of the order of $\frac{1}{n^{\frac{1}{3}}}$ we obtain ε-concentration with $\varepsilon \approx \frac{1}{n^{\frac{1}{3}}}$.

Remark: It can be easily seen using the coordinate symmetry of l_p^n balls that for fixed n and p, each random variable X_θ on the normalised l_p^n has the same variance, ρ_n^2. To see that ρ_n tends to a value bounded in p as $n \to \infty$, we notice that since we can express $\int_K x_1^2$ as the integral of the volumes of l_p^{n-1} balls, we have

$$\rho_n \to c_p \int_0^\infty y^2 e^{-\frac{y^p}{p}} \, dy.$$

Now c_p is a constant which using Stirling's formula can be seen to be bounded in p, a fact that is also true of the integral.

1.3 The concentration property for uniformly convex bodies contained in small Euclidean balls

K is said to be uniformly convex if for every $\gamma > 0$,

$$\inf \left\{ 1 - \frac{\|x+y\|}{2} \; : \; \|x\|, \|y\| \leq 1, \; \|x - y\| \geq \gamma \right\} = \delta(\gamma) > 0$$

where $\|.\|$ is the norm whose unit ball is K. We shall consider bodies for which there is a constant, c, such that

$$\delta(\gamma) \geq c\gamma^q \qquad \text{for some } 2 \leq q < \infty \tag{1.6}$$

and assume that

$$K \subset R\sqrt{n}\, B_2^n \quad . \tag{1.7}$$

Then we shall show that

Theorem 1.4 *For K satisfying (1.6) and (1.7), there is a constant, C, such that*

$$P\left(\left| \frac{|x|}{\sqrt{n}} - \rho \right| > \frac{CR\sqrt{\log n}}{n^{\frac{1}{q}}} \right) \leq \frac{4}{n}$$

and hence we have concentration as long as $\mu = \frac{CR\sqrt{\log n}}{n^{\frac{1}{q}}}$ is small. Notice that in order for μ to be small for a body with a good modulus of convexity

39

(i.e. with q small in (1.6)), the Euclidean ball containing it may be quite large. As K becomes less uniformly convex, it needs to be contained in a smaller ball.

Conditions (1.6) and (1.7) are satisfied in an appropriate way by all l_p^n balls, for $1 < p < \infty$, see for example [4]. For $1 < p \leq 2$, $\delta(\gamma) \geq \frac{(p-1)}{8}\gamma^2$. Here the bodies exhibit good uniform convexity since the power of γ remains 2. This compensates for the fact that $R \approx n^{\frac{1}{p}-\frac{1}{2}}$. On the other hand, for $2 < p < \infty$, the uniform convexity deteriorates rapidly since here, $\delta(\gamma) \approx \gamma^p$. However, in this case R is at most a constant.

Theorem 1.4 is a simple corollary of the following result of Gromov and Milman [11] which guarantees that uniformly convex bodies exhibit concentration with respect to the distance given by the norm. (A short new proof of this was recently found by Arias de Reyna, Ball and Villa [1] and arose from a sequence of ideas of Talagrand [19], Maurey [13] and Schmuckensläger [17].) We use hypothesis (1.7) to transfer the estimate in Theorem 1.5 to the Euclidean distance.

Theorem 1.5 (Gromov-Milman) *If $A \subset K$ has positive measure, and $d(x, A)$ is the distance from x to A (measured in the norm whose unit ball is*

40

K), then for any $\varepsilon > 0$

$$P\left(d(x, A) > \varepsilon\right) \leq \frac{e^{-2n\delta(\varepsilon)}}{P(A)}$$

Proof of Theorem 1.4: Let λ be the median of $\frac{|x|}{\sqrt{n}}$ on K and $A = \left(\frac{|x|}{\sqrt{n}} \leq \lambda\right)$.

Then by Theorem 1.5

$$P(d(x, A) > \gamma) \leq 2\, e^{-2n\delta(\gamma)}$$

and

$$P(d(x, A^c) > \gamma) \leq 2e^{-2n\delta(\gamma)} \quad .$$

Now if $y \in A$ and $d(x, y) \leq \gamma$ then $|x - y| \leq R\sqrt{n}\gamma$ since $K \subset R\sqrt{n}B_2^n$.

Hence

$$P\left(\left||x| - \lambda\sqrt{n}\right| > R\sqrt{n}\gamma\right) \leq 4e^{-2n\delta(\gamma)}$$

$$\leq 4e^{-2nc\gamma^q}.$$

This implies that for some constant c', the mean, ρ, differs from the median,

λ, by at most $c'Rn^{-\frac{1}{q}}$. Hence

$$P\left(\left|\frac{|x|}{\sqrt{n}} - \rho\right| > R\gamma + c'Rn^{-\frac{1}{q}}\right) \leq 4e^{-2cn\gamma^q}.$$

and letting γ be of the order $\left(\frac{\log n}{n}\right)^{\frac{1}{q}}$ we get the theorem. $\square$

Chapter 2

A Local Central Limit Property

In this section we use the main result of Chapter 1 and a property of the marginal density, g_θ, to show that, except on a set of directions of small spherical measure, for all $x \in \mathbb{R}$, $g_\theta(x)$ differs from the Gaussian density at x by very little.

To recap, in Chapter 1 we used a standard concentration of measure result on the sphere and the Lipschitz property of $\int_{-t}^{t} g_\theta(x)\, dx$ on the sphere to show that, for most directions, the integrals of the marginal densities are roughly the same. It was then shown that this common integral is approximately that of a Gaussian. So Theorem 1.1 says that under the concentration hypothesis, for $\delta > 0$, the integrals of the densities differ by at most $\delta + 4\varepsilon + \frac{c}{\sqrt{n}}$ in most

directions.

Unfortunately, a similar method cannot be used to show that the densities, g_θ and g, are locally close for most θ since $g_\theta(x)$ is not Lipschitz on the sphere for all x. (Consider, for example, the case when θ is near a coordinate direction and K is the cube.).

We can, however, derive the local central limit property from Theorem 1.1. The idea is that in addition to knowing that most g_θ have integrals close to Gaussian, we also know that each g_θ is rather smooth. To be precise, g_θ is log-concave. So in our argument we consider the convex function $-\log g_\theta$ and, for each spherical direction in Theorem 1.1, aim to show that if this is far from $-\log g$ at any point, then the integrals of the densities on an appropriate interval are far apart. The only drawback to this method is that, unlike in the weak-type result of Chapter 1, here we need to assume a bound on the variance, ρ^2. It is not entirely clear why such control is required when moving from a result for integrals to a local estimate.

Before stating the main result of this section, we remark that in order to simplify calculations we enforce some extra control at zero and further restrict the spherical directions we are considering to those for which $g_\theta(0)$ is very close to $g(0)$.

The main result of this section is the following:

Theorem 2.1 *There are constants, c, for which, under the ε-concentration hypothesis, for $\delta > 0$ and $\alpha = \sqrt{2\pi}\rho\left(\delta + \frac{c}{\sqrt{n}} + c\varepsilon\right)$, assuming α is less than $\frac{1}{10}$,*

$$\sigma\left(\left\{\theta : |g_\theta(x) - g(x)| \le c\left(\alpha\log\frac{1}{\alpha}\right)^{\frac{1}{4}} \quad \text{for all } x\right\}\right) \ge 1 - 6n\,e^{-\frac{n\delta^2}{c}}$$

2.1   Method

For the spherical directions in Theorem 1.1 we wish to show that if at any point $-\log g_\theta$ is far from $-\log g$ then the integrals of the densities on some interval are far apart, giving a contradiction. A brief outline of the argument is as follows.

We know that $-\log g(x)$ is equal to $\log(\sqrt{2\pi}\rho) + \frac{x^2}{2\rho^2}$ and we assume that at a particular point, z, $-\log g_\theta$ and $-\log g$ are a given distance apart. If the logarithms are close at zero, then , since $-\log g_\theta$ is convex, we can estimate from below the distance between the two functions, on some interval, by comparing $-\log g_\theta$ with a linear function. The distance between the functions at z then ensures that the integral of this calculated estimate on the aforementioned interval is sufficiently large.

44

It is our first task, therefore, to estimate the distance between the logarithms at zero. We can get bounds on $g_\theta(0)$ directly from results on logconcave functions, as in Lemma 1.5:

$$\frac{1}{\sqrt{12}\rho} \le g_\theta(0) \le \frac{1}{\sqrt{2}\rho}. \tag{2.1}$$

However this method does not give very good bounds. We could also use Theorem 1.1 to show that the functions are close at zero by noticing that if this were not the case, the integrals of the densities on an interval close to zero would be large. However, we choose a simpler argument which gives that for most θ, $|g_\theta(0) - g(0)|$ is small. This is described before we give the main part of the proof of Theorem 2.1.

Lemma 2.1 *Under the ε-concentration hypothesis, for $\delta > 0$,*

$$\sigma\left(\left\{\theta : |g_\theta(0) - g(0)| \le \delta + \frac{c}{\sqrt{n}} + c\varepsilon\right\}\right) \ge 1 - 2e^{-\frac{n\delta^2\rho^2}{6}}$$

Proof: The proof here is very similar to, although somewhat simpler than, the first two steps in the proof of Theorem 1.1 so we simply highlight the main issues involved. The argument splits into two parts, namely showing that $g_\theta(0)$ differs from it's mean over the sphere in very few directions and showing that this mean is close to $g(0)$.

For the first part, as we are dealing with central slices, we apply Busemann's theorem directly to show that $g_\theta(0)$ is the reciprocal of a norm restricted to the unit sphere and hence, using (2.1), Lipschitz with constant $\frac{\sqrt{3}}{\rho}$. Applying the concentration of measure result, Corollary 1.1, we get:

$$\sigma\left(\left\{\theta : \left|g_\theta(0) - \int_{S^{n-1}} g_\theta(0)\, d\sigma(\theta)\right| > \delta + \frac{c}{\sqrt{n}}\right\}\right) \leq 2e^{-\frac{n\delta^2\rho^2}{6}}. \qquad (2.2)$$

For the second part we estimate the mean over the sphere by an integral over K of Gaussian densities at zero with different variances:

$$\left|\int_{S^{n-1}} g_\theta(0)\, d\sigma(\theta) - \int_K \frac{\sqrt{n}}{|x|} \frac{1}{\sqrt{2\pi}}\, dx\right| \leq \frac{c}{n} \int_K \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n}}{|x|}\, dx$$

and, as before, we invoke the concentration hypothesis to show that most of these Gaussians have about the same variance, namely ρ^2. We remark that, unlike in the proof of Theorem 1.1, where integrating the densities made things nicer, here we must divide K into four sections to deal with the case when $|x|$ is small:

As before we denote by K_1 the region $K \cap \left\{\left|\frac{|x|}{\sqrt{n}} - \rho\right| \leq \varepsilon\rho\right\}$ and use a Lipschitz property, this time of the function $\frac{1}{y}$, to show that $\frac{\sqrt{n}}{|x|}$ is close to $\frac{1}{\rho}$ in this region. To get that $\frac{\sqrt{n}}{|x|}$ is like zero elsewhere we divide the rest of K into the subsets given by

$$K_{2,1} = K \cap \left\{\frac{|x|}{\sqrt{n}} \leq \frac{r}{\sqrt{n}}\right\}$$

46

$$K_{2,2} = K \cap \left\{ \frac{r}{\sqrt{n}} < \frac{|x|}{\sqrt{n}} < \rho - \varepsilon\rho \right\}$$

$$K_{2,3} = K \cap \left\{ \frac{|x|}{\sqrt{n}} > \rho + \varepsilon\rho \right\},$$

where r is chosen so that $r^n v_n \sqrt{n} = 1$. $K_{2,2}$ and $K_{2,3}$ are dealt with quickly using the concentration hypothesis which tells us that their volumes are small enough for it not to matter how big $\frac{\sqrt{n}}{|x|}$ is there. For $K_{2,1}$ (where $|x|$ is small) we notice that

$$\left| \int_{K_{2,1}} \frac{\sqrt{n}}{|x|} \, dx \right| \leq \int_{B^n(r)} \frac{\sqrt{n}}{|x|} \, dx$$

$$= n\sqrt{n}v_n \int_0^r \frac{u^{n-1}}{u} \, du,$$

where $B^n(r)$ is the n-dimensional Euclidean ball of radius r and v_n is the volume of $B^n(1)$.

So, from the second step we get

$$\left| \int_{S^{n-1}} g_\theta(0) \, d\sigma(\theta) - g(0) \right| \leq \frac{c}{\sqrt{n}} + c\varepsilon$$

and combining this with (2.2) we obtain the result. $\square$

2.2 Proof of Theorem 2.1

Assume that g_θ is such that its integral is close to that of g for all $t \subset \mathbf{R}$:

$$\left| \int_{-t}^t g_\theta(x) \, dx - \int_{-t}^t g(x) \, dx \right| \leq \delta + 4\varepsilon + \frac{c}{\sqrt{n}}.$$

Assume also the bound on $g_\theta(0)$ given by Lemma 2.1 so that

$$1 - \alpha \leq \frac{g_\theta(0)}{g(0)} \leq 1 + \alpha,$$

where

$$\alpha = \sqrt{2\pi}\rho \left(\delta + \frac{c}{\sqrt{n}} + c\varepsilon \right).$$

The idea is to show that $g_\theta(x)$ is close to $g(x)$ for all x by examining the difference between the convex functions $-\log g_\theta(x)$ and $-\log g(x) = \log(\sqrt{2\pi}\rho) + \frac{x^2}{2\rho^2}$.

To simplify the calculations, we rescale g_θ by the amount required to make it equal to g at zero and call this rescaled function h_θ,

$$h_\theta(x) = \frac{g(0)}{g_\theta(0)} g_\theta(x).$$

Notice that provided α is not too big, we have only scaled g_θ a little. So if we restrict to $\alpha < \frac{1}{10}$,

$$1 - \alpha \leq \frac{g(0)}{g_\theta(0)} \leq 1 + \frac{10}{9}\alpha.$$

To show that g_θ is close to g it then suffices to show that h_θ is close to g. Since the integral of g_θ is less than one and is close to the integral of g, it is clear that the integrals of g and the new function h_θ are also close. So since ρ cannot be too small, (we know that ρ^2 is larger than the variance of

the normalised Euclidean ball, $\frac{1}{6\pi e}$), there is a constant, c', for which

$$\left| \int_{-t}^{t} h_\theta(x) - g(x)\,dx \right| \leq \left| \frac{g(0)}{g_\theta(0)} - 1 \right| + \delta + 4\varepsilon + \frac{c}{\sqrt{n}} \leq c'\alpha \quad \text{for any } t. \quad (2.3)$$

(This constant, c', will reappear throughout the argument). So the idea is to show that if at some point $-\log h_\theta$ differs significantly from $-\log g$ then the difference in their integrals over an appropriate interval will be big. In other words we will show that if for some $z \in \mathbf{R}$, $-\log h_\theta(z)$ and $-\log g(z)$ differ significantly then

$$\left| \int_{s}^{t} h_\theta(x) - g(x)\,dx \right| > c'\alpha,$$

for some interval $[s, t]$, so that

$$\frac{1}{2} \left(\left| \int_{-t}^{t} h_\theta(x) - g(x)\,dx \right| + \left| \int_{-s}^{s} h_\theta(x) - g(x)\,dx \right| \right)$$

$$\geq \frac{1}{2} \left| \int_{-t}^{t} h_\theta(x) - g(x)\,dx - \int_{-s}^{s} h_\theta(x) - g(x)\,dx \right|$$

$$= \left| \int_{s}^{t} h_\theta(x) - g(x)\,dx \right|$$

$$> c'\alpha, \quad \text{contradicting (2.3).}$$

We first remark that we do not need to worry about the case when the two functions are far apart for $x \geq \rho\sqrt{\log \frac{1}{\alpha}}$, as long as they are close for smaller x. This is because $g(x)$ is small for large x and $g_\theta(x)$ is decreasing so the functions are automatically sufficiently close.

49
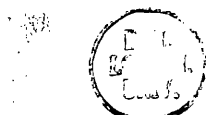
We consider two cases: when $-\log g(x)$ is assumed to be significantly

larger than $-\log h_\theta(x)$ at some point and, vice versa.

Case 1: $-\log g(z)$ *significantly bigger than* $-\log h_\theta(z)$.

We shall show that the logarithms differ by no more than γ, where

$$\gamma = \alpha^{\frac{1}{4}} \left(\frac{16c'\sqrt{2\pi}}{3} \sqrt{\log \frac{1}{\alpha}} \right)^{\frac{1}{2}}.$$

We consider separately the case when z is close to zero and otherwise.

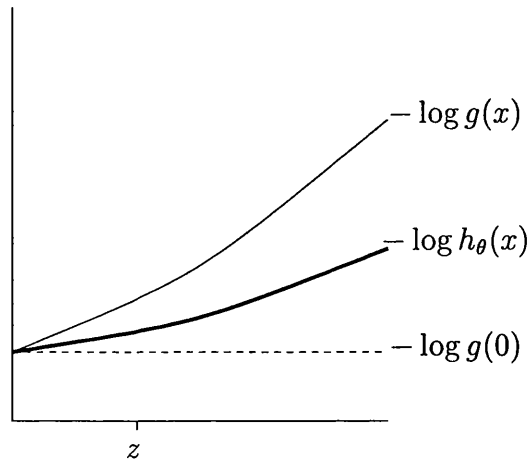

Figure 2.1: $-\log g(z)$ bigger than $-\log h_\theta(z)$, (z small).

If z is small, $-\log g(z)$ cannot be too much bigger than $-\log h_\theta(z)$ simply

because $-\log h_\theta$ is symmetric about zero and so is bigger than $-\log g(0)$,

see Figure 2.1. So we automatically have

$$-\log g(z) + \log h_\theta(z) \leq -\log g(z) + \log g(0)$$

$$= \frac{z^2}{2\rho^2}$$

for all z. Therefore for $z \leq \sqrt{\frac{3}{2}\rho^2\gamma}$,

$$-\log g(z) + \log h_\theta(z) \leq \frac{3}{4}\gamma. \tag{2.4}$$

Suppose then that $-\log g$ is significantly bigger than $-\log h_\theta$ at some point $z > \sqrt{\frac{3}{2}\rho^2\gamma}$:

$$-\log g(z) + \log h_\theta(z) > \gamma. \tag{2.5}$$

We wish to show that this implies that the integral of $h_\theta - g$ is bigger than $c'\,\alpha$ on some interval to give a contradiction to (2.3).

By the convexity of $-\log h_\theta$, we see using Figure 2.2 that for all $x \leq z$

$$-\log g(x) + \log h_\theta(x) > -\log g(x) - L_1(x)$$

$$= -\log g(x) - \left(-\log g(0) + \frac{x}{z}(-\log g(z) - \gamma + \log g(0))\right)$$

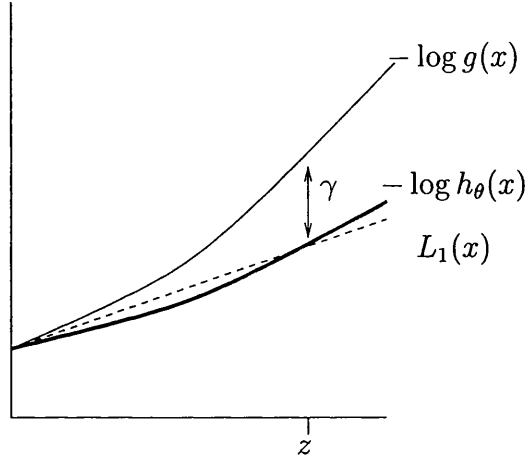$$= \frac{x^2}{2\rho^2} - \frac{xz}{2\rho^2} + \frac{\gamma x}{z}. \tag{2.6}$$

Figure 2.2: $-\log g(z)$ significantly bigger than $-\log h_\theta(z)$, (z big).

It is not difficult to check that

$$\frac{x^2}{2\rho^2} - \frac{xz}{2\rho^2} + \frac{\gamma x}{z} \geq \frac{\gamma}{2} \quad \text{for} \quad x \geq z - \frac{3\rho^2\gamma}{8z} \quad (= y \quad \text{say}), \qquad (2.7)$$

so taking exponentials in (2.6) and integrating over a region where (2.7)

holds, we get

$$\int_y^z h_\theta(x)\, dx \ > \ \int_y^z g(x) e^{\frac{x^2}{2\rho^2} - \frac{xz}{2\rho^2} + \frac{\gamma x}{z}}\, dx$$

$$\geq \ \int_y^z g(x) e^{\frac{\gamma}{2}}\, dx.$$

Now taking the integral of $g(x)$ from each side and estimating from below in

the natural way we have that the integrals of h_θ and g differ by at least a

52

certain value:

$$\int_y^z h_\theta(x) - g(x)\, dx > \int_y^z g(x)\left(e^{\frac{\gamma}{2}} - 1\right) dx$$

$$\geq \frac{3\rho^2\gamma}{8z}\frac{1}{\sqrt{2\pi}\rho}e^{-\frac{z^2}{2\rho^2}}\left(e^{\frac{\gamma}{2}} - 1\right)$$

$$\geq \frac{3\rho^2\gamma}{8z}\frac{1}{\sqrt{2\pi}\rho}e^{-\frac{z^2}{2\rho^2}}\frac{\gamma}{2}.$$

To see that this value is big we recall that we are only interested in $z < \rho\sqrt{\log\frac{1}{\alpha}}$. Therefore $e^{-\frac{z^2}{2\rho^2}} > \sqrt{\alpha}$ and

$$\int_y^z h_\theta(x) - g(x)\, dx > \frac{3}{16\sqrt{2\pi}}\gamma^2\frac{\sqrt{\alpha}}{\sqrt{\log\frac{1}{\alpha}}}$$

$$= c'\,\alpha,$$

substituting in the value of γ. So we have a contradiction to (2.3).

We conclude then that for $z > \sqrt{\frac{3}{2}\rho^2\gamma}$,

$$-\log g(z) + \log h_\theta(z) \leq \gamma \tag{2.8}$$

and, combining this with the result for small z, we have that (2.8) holds for all $z \in \mathbf{R}$. So for all $z \in \mathbf{R}$

$$h_\theta(z) - g(z) \leq \frac{1}{\sqrt{2\pi}\rho}e^{-\frac{z^2}{2\rho^2}}\left(e^\gamma - 1\right). \tag{2.9}$$

Now we need only notice that since γ is less than some constant, we can find a c such that $(e^\gamma - 1) \leq c\gamma$ to give

$$h_\theta(z) - g(z) \leq c\gamma. \tag{2.10}$$

53

Case 2: $-\log h_\theta(z)$ *significantly bigger than* $-\log g(z)$.

The argument for Case 2 is very similar to that for the previous case except in that here we do not need to consider small and large z separately. Nevertheless we include the proof for completeness.

We will show that the logarithms differ by no more than

$$\beta = \alpha^{\frac{1}{4}} \left(\frac{4\sqrt{2\pi}\, c'\, e^{\frac{c^2}{2}}}{c} \sqrt{\log \frac{1}{\alpha}} \right)^{\frac{1}{2}}.$$
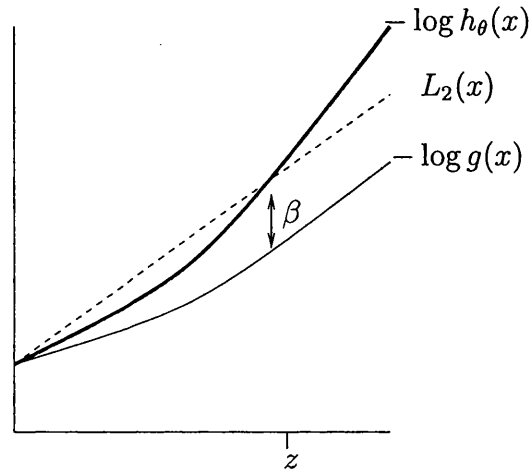

Figure 2.3: $-\log h_\theta(z)$ significantly bigger than $-\log g(z)$.

Suppose that the log functions at some point, z, are far apart:

$$-\log h_\theta(z) + \log g(z) > \beta. \tag{2.11}$$

Then we can observe from Figure 2.3 that by the convexity of $-\log h_\theta$, for

54

all $x \geq z$:

$$-\log h_\theta(x) + \log g(x) > L_2(x) + \log g(x)$$

$$= -\log g(0) + \frac{x}{z}\left(-\log g(z) + \beta + \log g(0)\right) + \log g(x)$$

$$= -\frac{x^2}{2\rho^2} + \frac{xz}{2\rho^2} + \frac{\beta x}{z}. \tag{2.12}$$

As before, we wish to simplify this bound and find that

$$e^{\frac{x^2}{2\rho^2} - \frac{xz}{2\rho^2} - \frac{\beta x}{z}} \leq 1 - \left(\frac{1 - e^{-\beta}}{\beta}\right)\left(-\frac{x^2}{2\rho^2} + \frac{xz}{2\rho^2} + \frac{\beta x}{z}\right)$$

provided $x \leq z + \frac{2\beta\rho^2}{z}$. (To show this we notice that the left hand side is less than or equal to one provided $x \leq z + \frac{2\beta\rho^2}{z}$ and, since we are only interested in $x \geq z$, we use that for $0 \leq t \leq \beta$, $e^{-t} \leq 1 - \left(\frac{1-e^{-\beta}}{\beta}\right)t$). The above estimate can be simplified if we restrict x further to $x \leq z + \frac{\beta\rho^2}{2z}$ by noticing that, for such x, the quadratic on the right hand side is greater than or equal to $\frac{\beta}{2}$. So taking exponentials in (2.12), applying the above estimates and rearranging we get

$$g(x) - h_\theta(x) > (1 - e^{-\beta})\frac{g(x)}{2}$$

for $z \leq x \leq z + \frac{\beta\rho^2}{2z}$.

At this stage in Case 1 (large z) we simply integrated over the region where the estimates were shown to hold. However, here we have not restricted ourselves to large z and so if z is small the region over which we would be

integrating would be large. So we choose instead $z + \min\left\{c\rho, \frac{\beta\rho^2}{2z}\right\}$ as the

upper limit of integration, where c is chosen so that the minimum is always

greater than $\frac{\beta\rho^2}{2\rho\sqrt{\log\frac{1}{\alpha}}}$. Once again, we bound the integral from below in the

natural way and observe that we are only interested in $z < \rho\sqrt{\log\frac{1}{\alpha}}$.

$$\int_z^{z+\min\left\{c\rho, \frac{\beta\rho^2}{2z}\right\}} g(x) - h_\theta(x)\, dx > \int_z^{z+\min\left\{c\rho, \frac{\beta\rho^2}{2z}\right\}} (1 - e^{-\beta})\frac{g(x)}{2}\, dx$$

$$\geq \min\left\{c\rho, \frac{\beta\rho^2}{2z}\right\}\frac{(1 - e^{-\beta})}{2\sqrt{2\pi}\rho}e^{-\frac{z^2}{2\rho^2}-\frac{\beta}{2}-\frac{c^2}{2}}$$

$$> \frac{\beta}{2\sqrt{\log\frac{1}{\alpha}}}\frac{(1 - e^{-\beta})}{2\sqrt{2\pi}}\sqrt{\alpha}\,e^{-\frac{\beta}{2}}e^{-\frac{c^2}{2}}.$$

It is clear that since α is bounded above by a constant (and hence also

β) that $\beta(1 - e^{-\beta})e^{-\frac{\beta}{2}}$ can be bounded below by a constant times β^2. So the

right hand side is greater than or equal to

$$\frac{c\,e^{-\frac{c^2}{2}}}{4\sqrt{2\pi}}\beta^2\frac{\sqrt{\alpha}}{\sqrt{\log\frac{1}{\alpha}}}$$

which is equal to $c'\,\alpha$.

So again we have a contradiction to (2.3) and

$$-\log h_\theta(z) + \log g(z) \leq \beta$$

for all $z \in \mathbf{R}$. Hence after reversing the above inequality and using the fact

that $e^{-\beta} \geq 1 - \beta$ we have

$$g(z) - h_\theta(z) \leq \beta\frac{1}{\sqrt{2\pi}\rho}e^{-\frac{z^2}{2\rho^2}} \leq c\beta.$$

56

So combining the above with (2.10) gives a local estimate for $|h_\theta - g|$. $\quad\square$

Chapter 3

A Concentration Estimate for

the Cube

In this chapter we prove a deviation inequality for the cube using a method developed by Marton to show similar results for Markov chains. Talagrand named this method the transportation method when simplifying Marton's arguments for certain product spaces.

Let us consider the cube, $[0,1]^n \subset \mathbf{R}^n$, and denote by P the n-dimensional Lebesgue measure on it. If B is any measurable subset of the cube, let B_t be its expansion,

$$B_t = \{x \, \epsilon \, [0,1]^n : d(x,B) \leq t\},$$

where $d(x, B)$ denotes the Euclidean distance from x to B. We shall prove a deviation inequality of the form

$$1 - P(B_t) \leq e^{-c't^2}, \tag{3.1}$$

provided B does not have too small probability, where c' is a constant dependent on $P(B)$.

Concentration results of this form have been known for the cube for some time. Indeed, it was pointed out by Tsirel'son in [21] and independently by Pisier (see e.g. [13]) that inequality (3.1), with bound $\frac{1}{P(B)}e^{-\frac{\pi}{2}t^2}$, can be obtained directly from concentration in Gauss space via a measure preserving Lipschitz map. Here our objective is to point out that the transportation method works directly for the cube and, more importantly, to ascertain the best constant that can be found using this method. This constant is better than those obtained by traditional methods and cannot be far from best possible. Finding this constant gives rise to a "text-book example" of a variational problem which has a surprisingly neat solution.

Marton's original method uses an inequality bounding the so-called $\bar{d}$-distance by informational divergence to prove a concentration of measure result for certain Markov chains (see [12] for definitions and a detailed ac-

count). The important thing in her method is that her one-dimensional inequality can be inducted on dimension and quickly implies a concentration of measure result. Marton's method certainly works for product spaces. However, Talagrand simplified it and strengthened the result for certain product spaces in [20], by considering l_2, rather than l_1, distance in the inequality. More precisely, Talagrand's inequality bounds the so-called transportation cost, with the square of the l_2-distance as the cost function. The definition of transportation cost is as follows.

Suppose we have two probability measures μ_1 and μ_2 on measurable spaces Ω_1 and Ω_2 respectively. The basic idea is to look at all bijections $b : \Omega_1 \to \Omega_2$ which transport μ_1 to μ_2, i.e. for which

$$\mu_1(A) = \mu_2(b(A)) \quad \text{whenever } A \subset \Omega_1.$$

For a given function $C : \Omega_1 \times \Omega_2 \to \mathbf{R}^+ \cup \{\infty\}$, where $C(x,y)$ measures the cost of moving a unit mass from x to y, we seek to minimise

$$\int_{\Omega_1} C(x, b(x)) \, d\mu_1(x).$$

If μ_1 or μ_2 has atoms then there may be no such function b. So, formally, the transportation cost is defined in terms of an integral over the product space $\Omega_1 \times \Omega_2$ with respect to a probability measure with marginals μ_1 and

μ_2. However, in our case $\Omega_1 = \Omega_2 = [0,1]^n$ and our measures will be the Lebesgue measure on the cube itself and a weighted Lebesgue measure on one of its subsets, so we can work with a transport function, b.

As already mentioned, we shall use the square of the Euclidean distance as our "cost function", C, just as Talagrand did for Gaussian measure. So now we can define the transportation cost, $\tau(\mu_1, \mu_2)$, to be the minimum, over all functions b as above, of

$$\int_{\Omega_1} |x - b(x)|_2^2 \, d\mu_1(x).$$

The main result of this chapter is the following:

Theorem 3.1 *(Bound on Transportation Cost) If A is a subset of $[0,1]^n$ and μ is the normalised restriction of the Lebesgue measure, P, to A (i.e. has density $1_A/P(A)$ with respect to P), then*

$$\tau(\mu, P) \leq \frac{2}{\pi^2} \log \frac{1}{P(A)}.$$

From this it is easy to get a concentration estimate using the following short argument. Let $B \subset [0,1]^n$. The cost of transporting $[0,1]^n$ to the complement of the expanded B, B_t^c, is clearly greater than that of transporting B (a subset

of $[0,1]^n$) to B_t^c. The Theorem gives an upper bound on the former and the latter is greater than $P(B)t^2$. So

$$P(B)t^2 \leq \frac{2}{\pi^2} \log \frac{1}{P(B_t^c)}.$$

Rearranging this we have

$$P(B_t^c) \leq e^{\frac{-\pi^2}{2} P(B)t^2}.$$

However, this bound can be improved by applying the Theorem to B as well as to B_t^c as in [12] and [20]. This gives the slightly better estimate

$$P(B_t^c) \leq \exp \left\{ \frac{-\pi^2}{2} \left(t - \sqrt{\frac{2}{\pi^2} \log \frac{1}{P(B)}} \right)^2 \right\}.$$

As already mentioned, we will see, in the proof of Theorem 3.1, that $\frac{\pi^2}{2}$ is the best constant that we can find using this method. Before we begin the proof, however, we observe that our constant is not far from best possible.

The following tells us that c' in (3.1) cannot be greater than 6. Let K be the cube of volume 1, now centered at zero. We regard K as a probability space and define on it the random variable $X_\theta : x \mapsto \langle x, \theta \rangle$, where $\theta = \left(\frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}} \right)$, so that the density of X_θ is obtained by scanning across K

with hyperplanes perpendicular to θ. Since

$$X_\theta(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(x),$$

where $X_i : x \mapsto x_i \in [-\frac{1}{2}, \frac{1}{2}]$ are random variables with zero mean and variance

$\rho^2 = \frac{1}{12}$, the Central Limit Theorem tells us that as $n \to \infty$

$$\text{Prob}\,(X_\theta > t) \quad \to \quad \frac{1}{\sqrt{2\pi}\rho} \int_t^\infty e^{-\frac{y^2}{2\rho^2}}\,dy$$

$$\geq \quad \left(\frac{t}{(12t)^2 + 1}\right) e^{-6t^2}.$$

Now we need only notice that the left hand side is precisely $P(H_t^c)$ for $H \subset$

$[0,1]^n$ given by the intersection of the cube with the halfspace through zero

perpendicular to θ:

$$H = \{x \in [0,1]^n : \langle x, \theta \rangle \leq 0\}. \qquad \square$$

3.1 Proof of Theorem 3.1, (Bound on Transportation Cost)

We wish to show by induction that

$$\tau(\mu, P) \leq \frac{2}{\pi^2} \log \frac{1}{P(A)} \quad \text{for } A \subset [0,1]^n,$$

63

where the left hand side is the minimum cost of transporting A to $[0, 1]^n$.

We shall begin by showing in section 3.1.1 that a result of the form

$$\tau(\mu, P) \leq c \log \frac{1}{P(A)} \quad \text{for } A \subset [0, 1]^n, \tag{3.2}$$

can be obtained from the following one-dimensional inequality for absolutely continuous $f : [0, 1] \to [0, 1]$,

$$\int_0^1 (f(t) - t)^2 f'(t) \, dt \leq c \int_0^1 f'(t) \log f'(t) \, dt, \tag{3.3}$$

where c is the same constant in both inequalities.

We continue in the ensuing sections to use the Calculus of Variations to show that (3.3) holds for all appropriate f if (and only if) $c \geq \frac{2}{\pi^2}$. More precisely we show that there is an optimising function satisfying an appropriate Euler-Lagrange equation and then we analyse the solutions of this equation.

3.1.1   The inductive step

Let us choose one of the n coordinate directions, e_1 say. We denote by $g(t)$ the $(n - 1)$-dimensional volume of A intersected with the "slice of the cube at $t \in [0, 1]$":

$$\{x \in [0, 1]^n : \langle x, e_1 \rangle = t\}.$$

The idea is that we first spread out A in the e_1 direction. We do so via an increasing, absolutely continuous function, f, which transports the $(n-1)$-dimensional slices in the e_1 direction so that the proportion of A between the slices at t and $t+\delta$ is equal to the proportion of $[0,1]^n$ between $f(t)$ and $f(t+\delta)$:

$$\frac{\delta g(t)}{P(A)} = f(t+\delta) - f(t). \tag{3.4}$$

The weighted cost of transporting in this way in one dimension is clearly

$$\int_0^1 (f(t) - t)^2 \frac{g(t)}{P(A)} \, dt. \tag{3.5}$$

We then use the inductive hypothesis, (3.2), to transport in each $(n-1)$-dimensional slice. The total of the transportation costs in all of the slices is at most

$$\int_0^1 c \log \frac{1}{g(f^{-1}(s))} \, ds.$$

After substituting $s = f(t)$, this is

$$-\int_0^1 cf'(t) \log g(t) \, dt. \tag{3.6}$$

We see from (3.4) that $f'(t) = \frac{g(t)}{P(A)}$. So to complete the inductive step we combine (3.5) and (3.6) and ask whether

$$\int_0^1 (f(t) - t)^2 f'(t) \, dt - \int_0^1 cf'(t) \log[f'(t)P(A)] \, dt \ \leq \ c \log \frac{1}{P(A)}.$$

When rearranged, this is

$$\int_0^1 (f(t) - t)^2 f'(t)\, dt - \int_0^1 cf'(t) \log f'(t)\, dt \;\leq\; c \log \frac{1}{P(A)} \int_0^1 (1 - f'(t))\, dt,$$

which simplifies to

$$\int_0^1 (f(t) - t)^2 f'(t)\, dt \;\leq\; c \int_0^1 f'(t) \log f'(t)\, dt, \qquad (3.7)$$

since $f(0) = 0$ and $f(1) = 1$.

The same inequality handles the one-dimensional case because we can transport in exactly the same way in dimension one, where $g(t) = 1_A(t)$ and where clearly we will not be required to transport further within $(n - 1)$-dimensional sheets. So the transportation cost is at most (3.5). Further, since $f'(t) = \frac{1_A(t)}{P(A)}$, we have

$$c \log \frac{1}{P(A)} = c \int_0^1 f'(t) \log f'(t)\, dt. \quad \square$$

It is not difficult to find some c for which (3.7) holds (and hence such that (3.2),

$$\tau(\mu, P) \leq c \log \frac{1}{P(A)},$$

is true). For example, if we rewrite the left hand side of (3.7) as below, we see that (3.7) holds with $c = 2$ by using the Csiszár-Kullback-Pinsker inequality. However we wish to find the smallest c.

66

We begin by rewriting (3.7). Notice that

$$\int_0^1 (f(t) - t)^2 (f'(t) - 1) \ dt = 0, \quad \text{since } f(0) = 0 \text{ and } f(1) = 1.$$

So we can rewrite the left hand side of (3.7) as

$$\int_0^1 (f(t) - t)^2 \ dt.$$

If we consider instead the deviation of f from t, $h(t) = f(t) - t$, then (3.7)

becomes

$$\int_0^1 h^2(t) \ dt \leq c \int_0^1 (1 + \dot{h}'(t)) \log(1 + h'(t)) \ dt. \tag{3.8}$$

Our problem is to find the smallest constant, c, such that the functional

in (3.8),

$$\mathcal{F}_c(h) = \int_0^1 \left[c(1 + h'(t)) \log(1 + h'(t)) - h^2(t) \right] \ dt, \tag{3.9}$$

is non-negative for all h in the admissible class of functions given by

$$\mathcal{C} = \{ h \text{ absolutely continuous}: \ h(0) = h(1) = 0, \ h' \geq -1 \}.$$

This variational problem is the subject of the following sections.

3.1.2   The variational problem

Recall that our aim is to find the smallest c such that the functional $\mathcal{F}_c$ is

non-negative for all functions in $\mathcal{C}$. First we will show that for all $c > 0$, a

minimiser of $\mathcal{F}_c$ exists and satisfies the Euler-Lagrange equation:

$$(1 + h'(t)) \; h(t) + \frac{c}{2} \; h''(t) = 0. \qquad (3.10)$$

Then we will find that if $c > \frac{2}{\pi^2}$, the only solution of (3.10) which satisfies the boundary conditions of $\mathcal{C}$, is the trivial one, $h = 0$. Hence $\mathcal{F}_c \geq 0$ for such c. To show that $c = \frac{2}{\pi^2}$ is the smallest constant for which $\mathcal{F}_c$ is non-negative, we will consider specific functions in our admissible class.

A classical theorem of Tonelli on the existence of minimisers of a one-dimensional variational integral,

$$\mathcal{F}(v) = \int_I F(x, v, v') \; dx,$$

can be found, for example, in [9]. The standard conditions are that the derivative of the Lagrangian with respect to p, $F(x, v, p)_p$, and the Lagrangian itself are continuous in (x, v, p) and that $F(x, v, p)$ is convex in p and has superlinear growth in p at ∞ (i.e. is such that there exists a function $\theta(p)$ such that

$$F(x, v, p) \; \geq \; \theta(p) \quad \text{for all } (x, v, p) \; \epsilon \; I \times \mathbf{R} \times \mathbf{R}$$

$$\text{and } \frac{\theta(p)}{|p|} \; \rightarrow \; \infty \quad \text{as } |p| \rightarrow \infty).$$

The superlinearity condition clearly does not hold for our Lagrangian,

$$F_c(x, v, p) = c(1 + p) \log(1 + p) - v^2,$$

because the "v" term could make F_c very small. However, it is not hard to see that the standard arguments can be adapted to demonstrate the existence of minimisers in our case. In fact, our Lagrangian has certain invariance properties which, if anything, make our problem easier than the general one. We include here a very rough explanation.

We wish to show that there exists a function $u \in C$ such that

$$\mathcal{F}_c(u) = \inf\{\mathcal{F}_c(v) : v \in C\}.$$

Call this infimum λ, say. From the boundary conditions on C, we have that $|v| \leq 1$ for $v \in C$, so $\mathcal{F}_c$ is bounded below. Hence we can find a minimising sequence, $\{u_k\} \subset C$, such that $\mathcal{F}_c(u_k) \to \lambda$.

The following properties of our Lagrangian allow us to take the functions in this minimising sequence to be positive and concave. Since F_c comprises only the square of the function, v, and its derivatives, rotating a negative section of the function by 180° leaves the functional unaltered. Further, if we approximate any positive function in the minimising sequence by a piecewise linear function and make this concave in steps, it is clear that in doing so

69

the functional, $\mathcal{F}_c$, decreases. This follows since v increases and since

$$(1+p)\log(1+p) \quad \text{is convex for } p \geq -1.$$

We can use the Ascoli-Arzelà Theorem to show that a subsequence of $\{u_k\}$ converges uniformly to a continuous function u, say. Equiboundedness is clear. To prove equicontinuity, we need to show that u_k' cannot get too large on $[0,1]$. But since we restricted u_k to being positive and concave, we need only show that u_k' is not too large near zero.

Notice that since $u_k(0) = u_k(1) = 0$, we can write $\mathcal{F}_c(x, u_k, u_k')$ as

$$\int_0^1 c[(1 + u_k')\log[1 + u_k'] - u_k'] - u_k^2 \, dx.$$

So if, for $\varepsilon > 0$, $u_k(\varepsilon) = L\varepsilon$, it is not hard to see, using the restriction $|u_k| \leq 1$ and that $(1+p)\log[1+p] - p \geq 0$ for $p \geq -1$, that

$$\mathcal{F}_c(x, u_k, u_k') \geq \varepsilon c[(1 + L)\log[1 + L] - L] - 1.$$

This in turn gives us an upper bound on $L\varepsilon$ which tends to zero as $\varepsilon \to 0$.

Finally, the concavity of the functions in the minimising sequence ensures that $u_k' \to u'$ a.e.. Then $\mathcal{F}_c(u_k) \to \mathcal{F}_c(u)$ dominatedly. $\square$

That any minimiser satisfies the Euler-Lagrange equation (3.10) is standard, see e.g. [9]. The only possible issue in our case is that the functional

must be defined for all functions in a neighbourhood of the minimiser, u.
But this does not pose a problem since our Lagrangian forces $u' > -1$. This
is because $(1 + p) \log(1 + p)$ has an infinite derivative at $p = -1$ and so a
minimiser will not have a derivative equal to -1, except possibly at 1.

3.1.3 Periodicity analysis

It remains to show that for $c > \frac{2}{\pi^2}$, the only solution of the Euler-Lagrange
equation is the trivial one (hence $\mathcal{F}_c \geq 0$ for such c) and that conversely
there are functions in our admissible class for which $\mathcal{F}_c < 0$ if $c < \frac{2}{\pi^2}$.

Recall that the Euler-Lagrange equation is given by

$$(1 + h'(t)) \ h(t) + \frac{c}{2} \ h''(t) = 0. \tag{3.11}$$

If we rearrange and multiply both sides by $h'(t)$, (3.11) becomes

$$h(t)h'(t) = -\frac{c \ h''(t)}{2(1 + h'(t))} \ h'(t)$$

and this integrates to

$$-h^2(t) + M^2 = c \left[h'(t) - \log(1 + h'(t)) \right], \quad \text{where } M = \sup \ |h|. \tag{3.12}$$

If we define the function $\Omega : (-1, \infty) \to [0, \infty)$ to be

$$\Omega(s) = s - \log(1 + s),$$

71

then (3.12) can be written in terms of Ω as
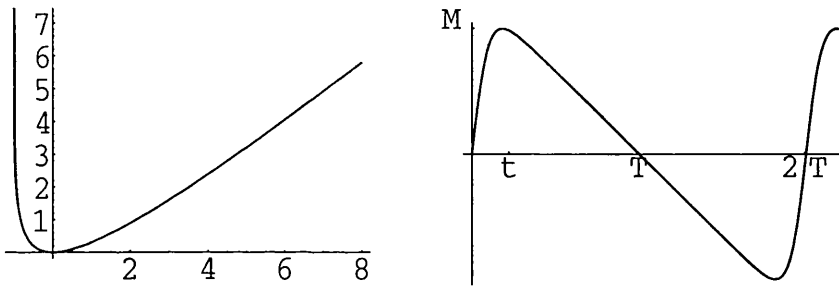
$$\frac{1}{c}(-h^2 + M^2) = \Omega(h').$$ (3.13)

Figure 3.1: The function Ω, (left), and a solution, h, of (3.12), (right).

It is not difficult to see that a solution of (3.12) which starts off positive, either increases to M or is periodic. Since we have the restriction that any function in our admissible class is zero at 1, we need only consider periodic solutions. So if for $c > \frac{2}{\pi^2}$ every non-trivial solution has period greater than 2, then we know that there is no non-trivial solution in our admissible class for such c. Hence we will have $\mathcal{F}_c \geq 0$ for $c > \frac{2}{\pi^2}$.

Let $2T$ denote the period of a solution, h, of (3.12). Suppose that h attains its maximum at the point $t \in (0, T)$. Then we can express t as an integral over h between 0 and M:

$$t = \int_0^t ds = \int_0^M \frac{1}{h'}\, dh.$$ (3.14)

72

Similarly, for the second section of the semiperiod, on which $h' \leq 0$ we have

$$T - t = \int_M^0 \frac{1}{h'}\, dh. \tag{3.15}$$

So if we denote the two branches of Ω^{-1}, Ω_+^{-1} and Ω_-^{-1}, using (3.13), we can express h' in terms of the inverses Ω_+^{-1} and Ω_-^{-1}, depending on the sign of h'.

Hence from (3.14) and (3.15), we know the semiperiod of a periodic solution of (3.12) to be

$$T = \int_0^M \frac{1}{\Omega_+^{-1}(\frac{1}{c}(M^2 - h^2))}\, dh - \int_0^M \frac{1}{\Omega_-^{-1}(\frac{1}{c}(M^2 - h^2))}\, dh. \tag{3.16}$$

We shall see below that

$$\frac{1}{\Omega_+^{-1}(x)} - \frac{1}{\Omega_-^{-1}(x)} \geq \frac{\sqrt{2}}{\sqrt{x}} \quad \text{for } x \geq 0. \tag{3.17}$$

Applying this to (3.16) we have

$$T \geq \int_0^M \frac{\sqrt{2}}{\sqrt{\frac{1}{c}(M^2 - h^2)}}\, dh = \frac{\pi}{\sqrt{2}}\sqrt{c}.$$

Hence for $c > \frac{2}{\pi^2}$ the return time, T, is strictly greater than 1 and we are done.

To prove (3.17) we fix $x \in [0, \infty)$ and define $s, t \geq 0$ by

$$\Omega_+^{-1}(x) = t \quad \text{and} \quad \Omega_-^{-1}(x) = -s.$$

Then

$$x = t - \log(1 + t) = -s - \log(1 - s) \qquad (3.18)$$

and we need to show that

$$\frac{1}{s} + \frac{1}{t} \geq \frac{\sqrt{2}}{\sqrt{x}},$$

i.e

$$\frac{1}{2}\left(\frac{1}{s} + \frac{1}{t}\right) \geq \frac{1}{\sqrt{2x}}.$$

By the AM/GM inequality, the left hand side is at least $\frac{1}{\sqrt{st}}$, so it suffices to show that under (3.18),

$$st \leq 2x.$$

By (3.18) this will follow if we show that for any $s, t \geq 0$,

$$st \leq t - \log(1 + t) - s - \log(1 - s),$$

i.e.

$$\log(1 + t) + \log(1 - s) \leq t - s - st.$$

But the left hand side is

$$
\begin{aligned}
\log((1 + t)(1 - s)) &= \log(1 + t - s - st) \\
&\leq t - s - st. \qquad \square
\end{aligned}
$$

Finally, to show that $\frac{2}{\pi^2}$ is the best constant, we find that there are specific admissible functions for which the inequality, (3.8),

$$\int_0^1 h^2(t)\ dt \le c \int_0^1 (1 + h'(t)) \log(1 + h'(t))\ dt,$$

does not hold, if $c < \frac{2}{\pi^2}$.

Let $j(t) = \delta \sin \pi t$ where, among other things, δ is sufficiently small to ensure that $j \in C$. Substituting this function into (3.8) we have

$$\int_0^1 \delta^2 (\sin \pi t)^2\ dt \le c \int_0^1 (1 + \delta \pi \cos \pi t) \log(1 + \delta \pi \cos \pi t)\ dt. \qquad (3.19)$$

For small δ, the right hand side is

$$c \int_0^1 \delta \pi \cos \pi t + \frac{(\delta \pi \cos \pi t)^2}{2}\ dt\ + O(\delta^3).$$

So we can rewrite (3.19) as

$$\int_0^1 \delta^2 (\sin \pi t)^2\ dt \le c \int_0^1 \frac{(\delta \pi \cos \pi t)^2}{2}\ dt\ + O(\delta^3).$$

Dividing both sides by δ^2 and letting $\delta \to 0$, we get

$$\int_0^1 (\sin \pi t)^2\ dt \le c \frac{\pi^2}{2} \int_0^1 (\cos \pi t)^2\ dt$$

which does not hold if $c < \frac{2}{\pi^2}$. $\square$

75

After the work in this chapter had already been completed, M. Ledoux communicated an alternative method of finding the same constant, c, in (3.2). His argument depends upon methods developed by himself and others which relate spectral methods to the log-Sobolev inequality. The log-Sobolev constant obtained can be transferred to a transportation constant using the methods of Otto and Villani [16].

Bibliography

[1] J.Arias de Reyna, K.Ball and R.Villa: Concentration of the distance in finite-dimensional normed spaces, *Mathematika 45*, 245-252, 1998.

[2] K.Ball: An Elementary Introduction to Modern Convex Geometry, *Flavors of Geometry, Mathematical Sciences Research Institute Publications 31*, Cambridge Univ. Press, 1997.

[3] K.Ball: Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$, *Stud. Math. 88, No.1*, 69-84, 1988.

[4] K. Ball, E. Carlen and E. Lieb: Sharp uniform convexity and smoothness inequalities for trace norms, *Inventiones Mathematicae 115*, 463-482, 1994

[5] K.Ball and I.Perissinaki: Subindependence of coordinate slabs in l_p, *Israel Journal 107*, 289-300, 1998 .

[6] C.Borell: Convex measures on locally convex spaces, *Ark. Math. 12*, 239-252, 1974.

[7] U. Brehm, H. Vogt and J. Voigt: Permanence of moment estimates for p-products of convex bodies, *Preprint TU Dresden*, MATH-AN-01-00.

[8] U. Brehm and J. Voigt: Asymptotics of cross sections for convex bodies, *Beiträge zur Algebra und Geometrie.*

[9] G.Buttazzo, M.Giaquinta and S.Hildebrandt: One-Dimensional Variational Problems, *Oxford Lecture Series in Mathematics and its Applications, 15*, Clarendon Press, Oxford, 1998.

[10] P.Diaconis and D.Freedman: Asymptotics of graphical projection pursuit, *Ann. of Stat. 12, 793-815*, 1984.

[11] M. Gromov and V. Milman: Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces, *Compos. Math. 62*, 263-282, 1987.

[12] K.Marton: Bounding $\bar{d}$-Distance by Informational Divergence: A Method to Prove Measure Concentration, *The Annals of Probability*, Vol. 24, *No.2*, $857 - 866$, 1996.

[13] B. Maurey: Some Deviation Inequalities, *Geometric and Functional Analysis*, Vol. 1, No. 2, 188-197, 1991.

[14] M.Meyer and S.Reisner: A geometric property of the boundary of symmetric convex bodies and convexity of flotation surfaces, *Geom Dedicata 37, No 3, 327-337*, 1991

[15] I.Perissinaki: Probabilistic Inequalities for Special Convex Bodies, *Ph.D. Thesis*, 1998.

[16] F.Otto and C.Villani: Generalization of an Inequality by Talagrand, and Links with the Logarithmic Sobolev Inequality, Preprint 1999.

[17] M.Schmuckensläger: A concentration of measure phenomenon on uniformly convex bodies, *Geometric Aspects of Functional Analysis*, (Israel 1992-1994), 275-287, 1995.

[18] V. Sudakov: Typical distributions of linear functionals on the spaces of high dimension, *Soviet Math. Dokl. 19 no.6*, 1578-1582, 1978.

[19] M.Talagrand: A new isoperimetric inequality and the concentration of measure phenomenon, *Geometric Aspects of Functional Analysis*, (Israel Seminar 1989-1990), *Lecture notes in Math. 1469*, 94-124, 1991.

[20] M.Talagrand: Transportation Cost for Gaussian and Other Product Measures, *Geometric and Functional Analysis*, Vol. 6, No. 3, 587 − 599, 1996.

[21] B.S.Tsirel'son, I.A.Ibragimov and V.N.Sudakov: Norms of Gaussian sample functions, *Proceedings of the Third Japan-USSR Symposium on Probability Theory*, Lecture Notes in Mathematics, Vol. 550, 1976.

[22] J. Voigt: A concentration of mass property for isotropic convex bodies in high dimensions, *Israel J. Math. 115*, 235-251, 2000.

[23] H.von Weizsäcker: Sudakov's typical marginals, random linear functionals and a conditional central limit theorem, *Probab. Theory Relat. Fields 107, 313-324*, 1997.