GEOMETRIC AND NONLINEAR LIMIT THEOREMS IN PROBABILITY THEORY

A Dissertation presented to

the Faculty of the Graduate School

University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

DANIEL JOHN FRESEN

Dissertation Supervisors:

Professor Nigel Kalton,
Professor Alexander Koldobsky,
Professor Mark Rudelson,

May 2012

©Copyright by Daniel Fresen 2012
All Rights Reserved

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

GEOMETRIC AND NONLINEAR LIMIT THEOREMS
IN PROBABILITY THEORY

presented by Daniel John Fresen,

a candidate for the degree of Doctor of Philosophy and hereby certify that in their opinion it is worthy of acceptance.

Professor William Banks

Professor Steven Hofmann

Professor Nigel Kalton (deceased)

Professor Alexander Koldobsky

Professor David Retzloff

Professor Mark Rudelson

Sun and moon, bless the Lord;
praise and exalt him above all forever.

Stars of heaven, bless the Lord;
praise and exalt him above all forever.

Fire and heat, bless the Lord;
praise and exalt him above all forever.

Dew and rain, bless the Lord;
praise and exalt him above all forever.

Frost and chill, bless the Lord;
praise and exalt him above all forever.

Mountains and hills, bless the Lord;
praise and exalt him above all forever.

Seas and rivers, bless the Lord;
praise and exalt him above all forever.

You sea monsters and all water creatures, bless the Lord;
praise and exalt him above all forever.

All you birds of the air, bless the Lord;
praise and exalt him above all forever.

All you beasts, wild and tame, bless the Lord;
praise and exalt him above all forever.

The book of Daniel, Chapter 3

So whether you eat or drink, or whatever you do, do everything for the glory of God.

1 Corinthians 10:31

ACKNOWLEDGMENTS

I am deeply grateful to all of my teachers. I am also grateful to those who wrote letters of recommendation for me, those willing to answer questions and discuss mathematics, and those who hosted me during several academic visits. I am thankful for friends and family that walked this long journey with me.

Nigel Kalton had a powerful influence on me. His classes on functional analysis were the best I ever took. For about three and a half years, until he passed away, he was my main adviser. In addition to mathematics, we played racquetball on a weekly basis and he would usually win despite my best efforts to beat him. I will always have fond memories of Nigel, as well as a deep respect and admiration for who he was.

It was while taking classes from Mark Rudelson and reading papers by Bo'az Klartag [42] [44], that I became interested in convex geometry. I found Vitali Milman's approach to Dvoretzky's theorem to be the most beautiful and interesting aspect of the theory.

Both Alexander Koldobsky and Mark Rudelson were very kind in taking me on as their student after Nigel passed away. Their generosity will not be forgotten and I hope to have interaction with both of them throughout my career.

My parents, Jill Fresen and John Fresen, had a profound impact on my education and mathematical development; equally so, and to a greater extent than anyone else. As an example, I first understood the concept of a function after an illuminating discussion with my mother. My father has had a long standing interest in the law of large numbers and the central limit theorem, which are the two main themes of this dissertation. I am grateful to them for many things; far too many to fully list here.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

ABSTRACT v

EXTENDED ABSTRACT vi

1. A MULTIVARIATE GNEDENKO LAW OF LARGE NUMBERS 1
    - 1.1. Main results 3
    - 1.2. Background 8
    - 1.3. The one dimensional case 13
    - 1.4. Main proofs 16
    - 1.5. Technical lemmas 22
    - 1.5. Proof of Theorem 3 31
    - 1.6. Optimality 36
    - 1.7. Proof of Theorem 4 37
2. SIMULTANEOUS CONCENTRATION OF ORDER STATISTICS 42
    - 2.1. The uniform distribution 48
    - 2.2. The general case 52
    - 2.3. Log-concave distributions 55
3. CONCENTRATION INEQUALITIES FOR LIPSCHITZ FUNCTIONS INTO AN ARBITRARY METRIC SPACE 58
    - 3.1. Main results 59
    - 3.2. Proofs 61
4. A NON-ASYMPTOTIC CENTRAL LIMIT THEOREM 65
    - 4.1. Main results 67

4.2. A simple example 69
4.3 Further discussion 70
4.4 Proofs 71
BIBLIOGRAPHY 75
VITA 83

ABSTRACT

The concentration of measure phenomenon is a nonlinear equivalent of the law of large numbers that deals with real valued Lipschitz functions and includes linear functionals such as the sample mean. In the first part of this dissertation we study functions that take values in more general metric-like spaces and have the property that they are invariant under coordinate permutations. In Chapter 1 we study functions that take values in the space of convex bodies, in Chapter 2 we study order statistics and in Chapter 3 we prove abstract concentration inequalities for functions taking values in an arbitrary metric space.

In the second part of the dissertation we study the central limit theorem. We show that if one conditions on certain tail events then convergence to the normal distribution can be achived without having to take a large number of summands. In fact 2 summands is enough.

The results presented here are taken from the author's papers [25], [27], [28] and [29]. In order to streamline the exposition, we have not included all of the results. We urge the reader to consult the published versions when they become available.

Concentration of measure

Let μ denote a probability measure on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} |x| d\mu(x) < \infty$$

The centroid of μ is defined as

$$\int_{\mathbb{R}} x d\mu(x)$$

The law of large numbers is the general phenomenon that if $X = (X_i)_1^n$ is a large random sample from μ then the sample mean

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{0.0.1}$$

typically approximates the centroid of μ. Although $\overline{x}$ is a random quantity, it has a distribution that is heavily *concentrated* around a single non-random quantity. There are various precise formulations of this principle, and many other linear functionals of the sample are similarly concentrated.

The concentration of measure phenomenon is a beautiful nonlinear extension of the law of large numbers that has various manifestations. Note that the sample mean in equation (0.0.1) is of the form $\overline{x} = f(X)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with Lipschitz constant $Lip(f) = n^{-1/2}$. A result going back to Lévy (see [57] section 2 and appendix V) is that if $f : \mathbb{R}^n \to \mathbb{R}$ is any Lipschitz function (not necessarily the sample mean) and $X = (X_i)_1^n$ is an i.i.d. sequence of random variables each with the standard normal distribution, then with probability at least $1 - \exp(-c\lambda^2)$,

$$|f(X) - \mathbb{E} f(X)| \leq \lambda Lip(f) \tag{0.0.2}$$

The unit vector $Y = ||X||_2^{-1}X$ is uniformly distributed on the unit sphere S^{n-1} and Lévy's inequality can be cast in the setting of a Lipschitz function $f : S^{n-1} \to \mathbb{R}$, in which case inequality (0.0.2) becomes

$$|f(Y) - \mathbb{E}f(Y)| \leq \frac{\lambda Lip(f)}{\sqrt{n}}$$

A similar inequality, due to Talagrand [69], guarantees concentration of convex Lipschitz functions with respect to product measures on the unit cube $[0, 1]^n$.

The concentration of measure phenomenon is not only a pleasing generalization of the law of large numbers, but plays a significant role in functional analysis and convex geometry. Lévy's inequality was the basis of Vitali Milman's proof of Dvoretzky's theorem on the ubiquity of ellipsoidal sections of convex bodies, or in the language of functional analysis, Hilbertian subspaces of Banach spaces. Let $K \subset \mathbb{R}^n$ be a convex body that contains the origin as an interior point. The Minkowski functional of K is defined as

$$||x||_K = \inf\{t > 0 : x \in tK\}$$

When K is symmetric, $|| \cdot ||_K$ is a norm. The associated Banach space $(\mathbb{R}^n, || \cdot ||_K)$ is a Hilbert space precisely when K is an ellipsoid. By appropriately positioning the body K, one may control the Lipschitz constant of $|| \cdot ||_K$ and conclude that on most of the sphere it oscillates very little. Where $|| \cdot ||_K$ has such limited oscillation, it is roughly proportional to the Euclidean norm, and after some careful footwork, one can actually produce a linear subspace of dimension $c \log n$ on which this happens. In fact *most* sections of K of dimension no larger than $c \log n$ are roughly ellipsoidal. Geometric aspects of the concentration of measure phenomenon are central to this dissertation.

Random polytopes

In Chapter 1 we study the convex hull $P_n = conv\{x_i\}_1^n$, where $(x_i)_1^n$ is a large i.i.d. sample from a probability measure μ on $\mathbb{R}^d$ that decays rapidly. We assume that μ has a non-vanishing density function of the form $f(x) = \exp(-g(x))$, where g is a convex function. Such functions are referred to as log-concave. We show that P_n approximates a deterministic body $F_{1/n}$ called the floating body which serves as a multivariate quantile. Both of these bodies can also be approximated using the contours of the density f. Our main result is that with probability at least $1 - c(\log n)^{-1000}$,

$$d_{\mathcal{L}}(P_n, F_{1/n}) \leq 1 + c\frac{\log\log n}{\log n}$$

where $c > 0$ does not depend on n and $d_{\mathcal{L}}(K, L)$ is the logarithmic Hausdorff distance between convex bodies K and L defined as

$$d_{\mathcal{L}}(K, L) = \inf\{\lambda > 1 : \exists x \in int(K \cap L), \lambda^{-1}(K - x) + x \subset L \subset \lambda(K - x) + x\}$$

If $int(K \cap L) = \emptyset$ we define $d_{\mathcal{L}}(K, L) = \infty$. When μ decays super-exponentialy, one can obtain bounds in terms of the Hausdorff distance between P_n and $F_{1/n}$. In the case of the standard multivariate normal distribution, P_n is eventually an enormous convex body, say the size of the sun, that varies by less than a millimeter from the Euclidean ball of radius $\sqrt{2\log n}$. In the one dimensional case, our results reduce to the Gnedenko law of large numbers regarding the maximum and minimum of a random sample.

Order statistics

In Chapter 2 we return to the one dimensional case of a random sample $(x_i)_1^n$ from a probability measure on $\mathbb{R}$ with cumulative distribution function F. The random quantity of interest is the sequence of order statistics $(x_{(i)})_1^n$, which is the non-decreasing rearrangement of the coordinates of x. We view this as a random element of the space ℓ_∞^n which is Euclidean space endowed with the ℓ_∞ norm $||x||_\infty = \max\{|x_i|\}_1^n$. Provided that the support of μ is connected and the tails of μ decay rapidly, one can predict the entire sequence of order statistics (simultaneously) with high accuracy. This too is an extension of the Gnedenko law of large numbers.

An abstract setting

In Chapter 3 we prove abstract concentration inequalities for Lipschitz functions into an arbitrary metric space. Although they seem more general, they represent another interpretation of the results presented in Chapter 2, and our work is simply to translate into a new mathematical language. Let (Ω, ρ) denote an arbitrary metric space. We show that for a Lipschitz function $f : S^{n-1} \to \Omega$ that is invariant under coordinate permutations, if x and y are independent vectors uniformly distributed on S^{n-1} then with probability at least $1 - c(\log n)^{-1000}$,

$$\rho(f(x), f(y)) \leq c\frac{\log\log n}{\log n}Lip(f)$$

This result is probably not sharp. Using a different technique, we expect to improve it to a bound of the form $\rho(f(x), f(y)) \leq cn^{-\alpha}Lip(f)$ for some fixed $\alpha > 0$, possibly $\alpha = 1/4$.

The central limit theorem

In the final chapter of the dissertation we change gear and focus on another phenomenon; the central limit theorem. The classical central limit theorem is, at a very fundamental level, an asymptotic result. One considers a large i.i.d. sample $(X_i)_1^n$ of random variables with mean zero and unit variance and studies the normalized sum

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

One can also consider other linear combinations with coefficients taken from a unit vector $\nu \in S^{n-1}$ with small ℓ_∞ norm. Our main contribution is to remove the asymptotic nature of the theorem by conditioning on certain tail events. In this setting $n = 2$ is sufficient, although one can consider any value of n.

We condition on events of the form

$$E_{\theta,T} = \left\{ \sum \theta_i X_i = T \right\}$$

where $T > 0$ is large, $\theta \in S^{n-1}$ and $\min\{|\theta_i|\}_1^n$ is not too small. The event $E_{\theta,T}$ has measure zero, although we provide a sensible definition for such conditioning. Indeed, one could condition on the event

$$E_{\theta,T,\varepsilon} = \left\{ \sum \theta_i X_i \in [T-\varepsilon, T+\varepsilon] \right\}$$

which has positive measure and then let $\varepsilon \to 0$. We also assume that the tails of the distribution of each X_i have a certain log-concave property. Under these assumptions, the distribution of any other linear combination, say

$$\sum \nu_i X_i$$

is approximately normal. As an example of such linear combinations, one could take $\theta_i = (-1)^i/\sqrt{n}$ and $\nu_i = 1/\sqrt{n}$ for all $1 \leq i \leq n$.

Even though we deal with linear combinations, this phenomenon is inherently nonlinear. For example, the mean of $\sum \nu_i X_i$ (conditioned on $E_{\theta,T}$) is a nonlinear function of the parameters involved. It is the solution to an optimization problem and, in certain cases, can be found using the method of Lagrange multipliers.

We present the results in a slightly more general setting and consider the restriction of functions with various regularity properties to affine subspaces. Our theory mirrors, and also contrasts, an extensive existing theory for projections.

Chapter 1

A Multivariate Gnedenko Law of Large Numbers

The Gnedenko law of large numbers [34] states that if F is the cumulative distribution of a probability measure μ on $\mathbb{R}$ such that for all $\varepsilon > 0$

$$\lim_{t\to\infty} \frac{F(t+\varepsilon) - F(t)}{1 - F(t+\varepsilon)} = \infty \tag{1.0.1}$$

then there are functions δ, T and $\mathcal{P}$ defined on $\mathbb{N}$ with

$$\lim_{n\to\infty} \delta_n = 0 \tag{1.0.2}$$

$$\lim_{n\to\infty} \mathcal{P}_n = 1 \tag{1.0.3}$$

such that for any $n \in \mathbb{N}$ and any i.i.d. sample $(\gamma_i)_1^n$ from μ, with probability $\mathcal{P}_n$ we have

$$|\max\{\gamma_i\}_1^n - T_n| < \delta_n$$

We define $0/0 = \infty$ to allow for the trivial case when μ has bounded support. The condition (1.0.1) implies super-exponential decay of the tail probabilities $1 - F(t)$, i.e. for all $c > 0$,

$$\lim_{t\to\infty} e^{ct}(1 - F(t)) = 0$$

The converse is almost true and can be achieved if we impose some sort of regularity on F. One such regularity condition is log-concavity (see Section 1.2). Of course all of this can be re-worded in multiplicative form. Provided $1 - F(t)$ is regular enough and decays super-polynomially, i.e. for any $m \in \mathbb{N}$,

$$\lim_{t \to \infty} t^m (1 - F(t)) = 0$$

then (1.0.2) and (1.0.3) hold, and with probability $\mathcal{P}_n$,

$$\left| \frac{\max\{\gamma_i\}_1^n}{T_n} - 1 \right| \le \delta_n$$

Note that rapid decay of the left hand tail provides concentration of $\min\{\gamma_i\}_1^n$, and that $[\min\{\gamma_i\}_1^n, \max\{\gamma_i\}_1^n] = conv\{\gamma_i\}_1^n$.

In this chapter we extend the Gnedenko law of large numbers to the multivariate setting. We consider a large collection of i.i.d. random vectors $\{x_i\}_1^n$ in $\mathbb{R}^d$ that follow a log-concave distribution μ with density function f. The object of interest is the convex hull $P_n = conv\{x_i\}_1^n$, which is called a random polytope. It is shown that with high probability, P_n approximates a deterministic body $F_{1/n}$ called the floating body, which is what remains of $\mathbb{R}^d$ after deleting all open half-spaces $\mathfrak{H}$ such that $\mu(\mathfrak{H}) < 1/n$. As in the one dimensional case, the way in which P_n approximates $F_{1/n}$ depends on how rapidly μ decays. Of primary interest is a quantitative analysis in terms of the number of points, and in this regard our results are essentially optimal (see Section 1.7).

The fact that the floating body can be used in order to model random polytopes is well known in the setting where μ is the uniform distribution on a convex body (see for example [8] and [10]). Our main contribution is to study this approximation in the more general setting of log-concave measures. Unlike the former case, the objects that we study can have many different shapes as $n \to \infty$ and are not limited to lie within a bounded region of space.

The notion of a multivariate Gnedenko law of large numbers has also been considered by Goodman [35] in the setting of Gaussian measures on seperable Banach spaces. In his paper he shows that with probability 1, the Hausdorff distance between the sample $\{x_i\}_1^n$ and the ellipsoid $\sqrt{2\log n}\,\mathcal{E}$ converges to zero as $n \to \infty$, where $\mathcal{E}$ is the unit ball of the reproducing kernel Hilbert space associated to μ.

1.1 Main Results

Let $d \geq 1$, $n \geq d+1$ and let μ be a log-concave probability measure on $\mathbb{R}^d$ with a density function $f = d\mu/dx$. This means that f is of the form $f(x) = \exp(-g(x))$ where g is convex. Let $(x_i)_1^n$ denote a sequence of i.i.d. random vectors in $\mathbb{R}^d$ with distribution μ, and consider the random polytope $P_n = conv\{x_i\}_1^n$. For any $x \in \mathbb{R}^d$, define

$$\widetilde{f}(x) = \inf_{\mathfrak{H}} \mu(\mathfrak{H})$$

where $\mathfrak{H}$ runs through the collection of all open half-spaces that contain x. For any $\delta > 0$, the *floating body* is defined as

$$F_\delta = \{x \in \mathbb{R}^d : \widetilde{f}(x) \geq \delta\} \tag{1.1.1}$$

Note that F_δ is the intersection of all closed half-spaces $\mathfrak{H}$ such that $\mu(\mathfrak{H}) \geq 1/n$, and is therefore convex. If $\mathfrak{H}$ is any open half-space that contains the centroid of μ, then $\mu(\mathfrak{H}) \geq e^{-1}$ (see Lemma 5.12 in [51] or Lemma 3.3 in [12]) hence F_δ is non-empty provided that $\delta \leq e^{-1}$. Such a floating body was defined by Schütt and Werner [65] in the case where μ is the uniform distribution on a convex body. We define the *logarithmic Hausdorff distance* between convex bodies $K, L \subset \mathbb{R}^d$ as,

$$d_{\mathfrak{L}}(K, L) = \inf\{\lambda \geq 1 : \exists x \in int(K \cap L),\, \lambda^{-1}(L - x) + x \subset K \subset \lambda(L - x) + x\}$$

where we use the convention that $\inf(\emptyset) = \infty$. The main result of the chapter is as follows:

Theorem 1. *There exist universal constants $c, c', \widetilde{c} > 0$ with the following property. Let $q \geq 1$, $d \in \mathbb{N}$ and $n \geq c \exp\exp(5d) + c'q^3$. Let μ be a probability measure on $\mathbb{R}^d$ with a log-concave density function, $(x_i)_1^n$ an i.i.d. sample from μ, $P_n = conv\{x_i\}_1^n$ and $F_{1/n}$ the floating body as in (1.1.1). With probability at least $1 - 3^{d+3}(\log n)^{-q}$,*

$$d_{\mathfrak{L}}(P_n, F_{1/n}) \leq 1 + \widetilde{c}d(d+q)\frac{\log\log n}{\log n} \tag{1.1.2}$$

The strategy of the proof is to use quantitative bounds in the one dimensional case to analyze the Minkowski functional of P_n in different directions. The idea is simple, however there are some subtle complications. The lack of symmetry is a complicating factor, and the fact that the half-spaces of mass $1/n$ do not necessarily touch $F_{1/n}$ adds to the intricacy of the proof.

We define f to be p-log-concave if it is of the form $f(x) = c\exp(-g(x)^p)$ where g is a non-negative convex function and $c > 0$.

Theorem 2. *For all $q > 0$, $p > 1$ and $d \in \mathbb{N}$, and any probability measure μ on $\mathbb{R}^d$ with a non-vanishing p-log-concave density function, there exist $c, \widetilde{c} > 0$ such that for all $n \in \mathbb{N}$ with $n \geq d + 2$, if $(x_i)_1^n$ is an i.i.d. sample from μ, $P_n = conv\{x_i\}_1^n$ and*

$F_{1/n}$ *is the floating body as in (1.1.1), then with probability at least* $1 - \widetilde{c}(\log n)^{-q}$ *we have*

$$d_{\mathcal{H}}(P_n, F_{1/n}) \le c\frac{\log\log n}{(\log n)^{1-\frac{1}{p}}} \tag{1.1.3}$$

Theorem 2 can easily be extended to a much larger class of log-concave distributions. Using Theorem 1, any bound on the growth rate of $diam(F_{1/n})$ automatically transfers to a bound on $d_{\mathcal{H}}(P_n, F_{1/n})$.

Our prototypical example is the class of distributions introduced by Schechtman and Zinn [64] of the form $f(x) = c_p^d \exp(-||x||_p^p)$, where $1 \le p < \infty$ and $c_p = p/(2\Gamma(p^{-1}))$. For these distributions, $P_n \approx (\log(c_p^d n))^{1/p} B_p^d$. Of particular interest is the Gaussian distribution, where $p = 2$. In this case (actually for the standard Gaussian distribution), Bárány and Vu [9] obtained a similar approximation (see Remark 9.6 in their paper) and showed that there exist two radii R and r, both functions of n and d, such that for any fixed $d \ge 2$ both $r, R = (2\log n)^{1/2}(1 + o(1))$ as $n \to \infty$, and with 'high probability' $rB_2^d \subset P_n \subset RB_2^d$. Their sandwiching result served as a key step in their proof of the central limit theorem for Gaussian polytopes (asymptotic normality of various functionals such as the volume and the number of faces).

In the setting where μ is the uniform distribution on a convex body, the floating body is usually denoted by K_δ. In this context it is trivial that $\lim_{n\to\infty} d_{\mathcal{H}}(P_n, K) = 0$ (almost surely) and the phenomenon of interest is the rate at which P_n approached the boundary of K. Bárány and Larman [8] proved that for $n \ge n_0(d)$,

$$c'\mathrm{vol}_d(K\backslash K_{1/n}) \le \mathbb{E}\mathrm{vol}_d(K\backslash P_n) \le c''(d)\mathrm{vol}_d(K\backslash K_{1/n})$$

The reader may be interested to contrast our results with the results in [14]. The results presented here require a very large sample size and guarantee a precise approximation, somewhat in the spirit of the 'almost-isometric' theory of convex bodies. On the other hand, the results presented in [14] describe a type of approximation in the spirit of the 'isomorphic' theory, and are most interesting specifically in high dimensional spaces.

We also study two other deterministic bodies that serve as approximants to the random body. Define

$$f^{\sharp}(x) = \inf_{\mathcal{H}} \int_{\mathcal{H}} f(y) d_{\mathcal{H}}(y)$$

where $\mathcal{H}$ runs through the collection of all hyperplanes that contain x, and $d_{\mathcal{H}}$ stands for Lebesgue measure on $\mathcal{H}$. For any $\delta > 0$, define the bodies

$$D_{\delta} = Cl\{x \in \mathbb{R}^d : f(x) \geq \delta\}$$

$$R_{\delta} = Cl\{x \in \mathbb{R}^d : f^{\sharp}(x) \geq \delta\}$$

where $Cl(E)$ denotes the closure of a set E. By log-concavity of f, both D_{δ} and R_{δ} are convex.

Theorem 3. *Let $d \in \mathbb{N}$ and let μ be a probability measure on $\mathbb{R}^d$ with a non-vanishing log-concave density function. Then we have*

$$\lim_{\delta \to 0} d_{\mathfrak{L}}(F_{\delta}, D_{\delta}) = 1 \tag{1.1.4}$$

$$\lim_{\delta \to 0} d_{\mathfrak{L}}(F_{\delta}, R_{\delta}) = 1 \tag{1.1.5}$$

Similar results hold in the Hausdorff distance for log-concave distributions that decay super-exponentially.

Let $X \in \mathbb{R}^d$ be a random vector with distribution μ. The random variable $-\log f(X)$ is a sort of differential information content (see [13]). The differential entropy of μ is defined as

$$\begin{aligned} h(\mu) &= -\mathbb{E}\log f(X) \\ &= -\int_{\mathbb{R}^d} f(x)\log f(x)dx \end{aligned}$$

and the entropy power defined as $N(\mu) = \exp(2d^{-1}h(\mu))$. Note that the distribution of $-\log f(X)$ can be expressed in terms of the function $\delta \mapsto \mu(D_\delta)$,

$$\mathbb{P}\{-\log f(X) \le t\} = \mu(D_\delta) \quad : \quad \delta = e^{-t}$$

Because of the rapid decay of f, the body D_δ acts as an essential support for the measure μ. For $\delta = e^{-d}$, this was studied by Klartag and Milman [43] (see Lemma 2.2 and Corollary 2.4 in their paper). Bobkov and Madiman later provided a more precise description. In [13] they show that the variance of $-\log f(X)$ is at most Cd, where $C > 0$ is a universal constant, and that in high dimensional spaces, $f(X)^{2/d}$ is strongly concentrated around $N(\mu)$. Theorem 1.1 in their paper can be written as

$$\mu\{x \in \mathbb{R}^d : N(\mu)^{d/2}\delta \le f(x) \le N(\mu)^{d/2}\delta^{-1}\} \ge 1 - 2\delta^{p(d)}$$

provided $\delta \in (0,1)$, where $p(d) = 16^{-1}d^{-1/2}$. In Lemma 16 we show that if μ is isotropic and has a continuous density function, then for all $\delta < \exp(-11d\log d - 7)$,

$$\mu\{x \in \mathbb{R}^d : f(x) \ge \delta\} \ge 1 - \alpha_d\delta(\log\delta^{-1})^d \tag{1.1.6}$$

where $\alpha_d = \exp(d^2(2\log d + c_1))$. In a fixed dimension, inequality (1.1.6) displays the natural quantitative behaviour of $\mu(D_\delta)$ as $\delta \to 0$ and is sharp up to a factor of $\log\delta^{-1}$.

Let $\mathcal{K}_d$ denote the collection of all convex bodies in $\mathbb{R}^d$. For all $K, L \in \mathcal{K}_d$, define

$$d_{BM}(K, L) = \inf\{\lambda \geq 1 : \exists x \in \mathbb{R}^d,\, \exists T,\, K \subset TL \subset \lambda(K - x) + x\} \tag{1.1.7}$$

where T represents any affine transformation of $\mathbb{R}^d$. This is a modification of the classical Banach-Mazur distance between normed spaces (origin symmetric bodies).

Theorem 4. *For all $d \in \mathbb{N}$, there exists a probability measure μ on $\mathbb{R}^d$ with the following universality property. Let $(x_i)_1^\infty$ be an i.i.d. sample from μ, and for each $n \in \mathbb{N}$ with $n \geq d+1$ let $P_n = conv\{x_i\}_1^n$. Then with probability 1, the sequence $(P_n)_{d+1}^\infty$ is dense in $\mathcal{K}_d$ with respect to d_{BM}.*

Throughout the chapter we will make use of variables c, $\widetilde{c}$, c_1, c_2, n_0, m etc. At times they represent universal constants and at other times they depend on parameters such as the dimension d or the measure μ. Such dependence will always be clear from the context, and will either be indicated explicitly as c_d, $c(d)$, $n_0(d)$ etc., or implicitly as in Theorem 2, where c and $\widetilde{c}$ depend on q, p, d and μ. Half-spaces shall be indexed as $\mathfrak{H}_{\theta,t} = \{x \in \mathbb{R}^d : \langle x, \theta \rangle \geq t\}$ and hyperplanes as $\mathcal{H}_{\theta,t} = \{x \in \mathbb{R}^d : \langle x, \theta \rangle = t\}$, where $\theta \in S^{d-1}$ and $t \in \mathbb{R}$.

1.2 Background

Most of the material in this section is discussed in [6], [7], [54] and [57]. We denote the standard Euclidean norm on $\mathbb{R}^d$ by $||\cdot||_2$. For any $\varepsilon > 0$, an ε-net in S^{d-1} is a subset $\mathcal{N}$ such that for any distinct $\omega_1, \omega_2 \subset \mathcal{N}$, $||\omega_1 - \omega_2||_2 > \varepsilon$, and for all $\theta \in S^{d-1}$ there exists $\omega \in \mathcal{N}$ such that $||\theta - \omega||_2 \leq \varepsilon$. Such a subset can easily be constructed using induction. By a standard volumetric argument we have

$$|\mathcal{N}| \leq \left(\frac{3}{\varepsilon}\right)^d \tag{1.2.1}$$

By induction, any $\theta \in S^{d-1}$ can be expressed as a series

$$\theta = \omega_0 + \sum_{i=1}^{\infty} \varepsilon_i \omega_i \tag{1.2.2}$$

where each $\omega_i \in \mathcal{N}$ and $0 \leq \varepsilon_i \leq \varepsilon^i$. To see this, express $\theta = \omega_0 + r_0$, where $\omega_0 \in \mathcal{N}$ and $||r_0||_2 \leq \varepsilon$. Then express $||r_0||^{-1} r_0 \in S^{d-1}$ in a similar fashion and iterate this procedure.

Define the functional

$$||x||_{\mathcal{N}} = \max\{\langle x, \omega \rangle : \omega \in \mathcal{N}\}$$

As an easy consequence of the Cauchy-Schwarz inequality, provided $\varepsilon \in (0, 1)$ we have

$$(1 - \varepsilon)||x||_2 \leq ||x||_{\mathcal{N}} \leq ||x||_2 \tag{1.2.3}$$

which implies that

$$B_2^d \subset B_{\mathcal{N}} \subset (1 - \varepsilon)^{-1} B_2^d \tag{1.2.4}$$

where $B_{\mathcal{N}} = \{x : ||x||_{\mathcal{N}} \leq 1\}$. The body $B_{\mathcal{N}}$ is what remains if one deletes all half-spaces that are tangent to B_2^d at points in $\mathcal{N}$.

A *convex body* is a compact convex subset of Euclidean space with nonempty interior. For a convex body $K \subset \mathbb{R}^d$ that contains the origin as an interior point, its *Minkowski functional* is defined as

$$||x||_K = \inf\{\lambda > 0 : x \in \lambda K\}$$

for all $x \in \mathbb{R}^d$. By convexity of K, one can easily show that $||\cdot||_K$ obeys the triangle inequality. The dual Minkowski functional is defined as

$$||y||_{K^\circ} = \sup\{\langle x, y \rangle : x \in K\}$$

for all $y \in \mathbb{R}^d$, and the *polar* of K is

$$K^\circ = \{y \in \mathbb{R}^d : ||y||_{K^\circ} \leq 1\}$$

By the Hahn-Banach theorem, $K^{\circ\circ} = K$.

The *Hausdorff distance* $d_{\mathcal{H}}$ between K and L is defined as

$$d_{\mathcal{H}}(K, L) = \max\{\max_{k \in K} d(k, L); \max_{l \in L} d(K, l)\}$$

By convexity this reduces to

$$\begin{aligned} d_{\mathcal{H}}(K, L) &= \sup_{\theta \in S^{d-1}} |\sup_{k \in K} \langle k, \theta \rangle - \sup_{l \in L} \langle l, \theta \rangle\,| \\ &= \sup_{\theta \in S^{d-1}} |(||\theta||_{K^\circ} - ||\theta||_{L^\circ})| \end{aligned}$$

We define the *logarithmic Hausdorff distance* between K and L about a point $x \in int(K \cap L)$ as

$$d_{\mathfrak{L}}(K, L, x) = \inf\{\lambda \geq 1 : \lambda^{-1}(L - x) + x \subset K \subset \lambda(L - x) + x\}$$

provided $int(K \cap L) \neq \emptyset$, and

$$d_{\mathfrak{L}}(K, L) = \inf\{d_{\mathfrak{L}}(K, L, x) : x \in int(K \cap L)\}$$

Note that

$$\log d_{\mathfrak{L}}(K, L, 0) = \sup_{\theta \in S^{d-1}} |\log ||\theta||_K - \log ||\theta||_L|$$

The following relations follow from the definitions above,

$$\begin{aligned} d_{\mathfrak{L}}(K, L, 0) &= d_{\mathfrak{L}}(K^\circ, L^\circ, 0) \\ d_{\mathfrak{L}}(TK, TL) &= d_{\mathfrak{L}}(K, L) \end{aligned} \tag{1.2.5}$$

where T is any invertible affine transformation. In addition one can check that,

$$\begin{aligned} d_{BM}(K,L) &\leq d_{\mathfrak{L}}(K,L)^2 \\ d_{\mathcal{H}}(K,L) &\leq diam(K)(d_{\mathfrak{L}}(K,L)-1) \end{aligned} \tag{1.2.6}$$

hence all of our bounds in terms of $d_{\mathfrak{L}}$ apply equally well to d_{BM}. For large bodies, $d_{\mathcal{H}}$ is more sensative than $d_{\mathfrak{L}}$. More precisely, if $r > 1$ and $rB_2^d + x \subset K$ for some $x \in \mathbb{R}^d$, and if $d_{\mathcal{H}}(K,L) \leq 1/2$, then

$$d_{\mathfrak{L}}(K,L) \leq 1 + 2r^{-1}d_{\mathcal{H}}(K,L) \tag{1.2.7}$$

By a simple compactness argument, there is an ellipsoid of maximal volume $\mathcal{E}_K \subset K$. This ellipsoid is called the *John ellipsoid* [7] associated to K. It can be shown that $\mathcal{E}_k$ is unique and has the property that $K \subset d(\mathcal{E}_k - x) + x$, where x is the center of $\mathcal{E}_k$. In particular, $d_{\mathfrak{L}}(\mathcal{E}_k, K) \leq d$.

In [26] it is shown that provided $\lambda < 8^{-d}$, we have

$$d_{\mathfrak{L}}(K, K_\lambda, x) \leq 1 + 8\lambda^{1/d} \tag{1.2.8}$$

where x is the centroid of K and K_δ is the floating body inside K.

The *cone measure* on ∂K is defined as

$$\mu_K(E) = \text{vol}_d(\{r\theta : \theta \in E,\ r \in [0,1]\})$$

for all measurable $E \subset \partial K$. The significance of the cone measure is that it leads to a natural polar integration formula (see [58]); for all $f \in L_1(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} f(x)dx = d\int_0^\infty \int_{\partial K} r^{d-1} f(r\theta) d\mu_K(\theta) dr \tag{1.2.9}$$

A probability measure μ is called *isotropic* if its centroid lies at the origin and its covariance matrix is the $d \times d$ identity matrix.

A function $f : \mathbb{R}^d \to [0, \infty)$ is called *log-concave* (see [43]) if

$$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda}$$

for all $x, y \in \mathbb{R}^d$ and all $\lambda \in (0,1)$. Any such function can be written in the form $f(x) = e^{-g(x)}$ where $g : \mathbb{R}^d \to (-\infty, \infty]$ is convex. If f is the density of a probability measure μ, then it must decay exponentially to zero. In this case g lies above a cone, i.e.

$$g(x) \geq m||x||_2 - c \tag{1.2.10}$$

with $m, c > 0$. As a consequence of the Prékopa-Leindler inequality [6], if x is a random vector with log-concave density and y is any fixed vector, then $\langle x, y \rangle$ has a log-concave density in $\mathbb{R}$. Log-concave functions are very rigid. One such example of this rigidity (see Lemma 5.12 in [51]) is the fact that if $\mathfrak{H}$ is any half-space containing the centroid of μ, then $\mu(\mathfrak{H}) \geq e^{-1}$. Another example (see Theorem 5.14 in [51]) is that if μ is isotropic, then

$$2^{-7d} \leq f(0) \leq d(20d)^{d/2} \tag{1.2.11}$$

$$(4\pi e)^{-d/2} \leq ||f||_{\infty} \leq 2^{8d} d^{d/2} \tag{1.2.12}$$

and if $||x||_2 \leq 1/9$ then,

$$2^{-8d} \leq f(x) \leq d2^d (20d)^{d/2} \tag{1.2.13}$$

Let $1 \leq p < \infty$. If $g : \mathbb{R}^d \to [0, \infty]$ is convex and $\lim_{x\to\infty} g(x) = \infty$, then the probability measure with density given by

$$f(x) = ce^{-g(x)^p}$$

will be called p-log-concave. This is a natural generalization of the normal distribution. If f is p-log-concave, then it is also p'-log-concave for all $1 \leq p' \leq p$.

Let $\mathbf{H}_d$ denote the collection of all $d-1$ dimensional affine subspaces (hyperplanes) of $\mathbb{R}^d$. The Radon transform of a log-concave function $f : \mathbb{R}^d \to [0, \infty)$ is the function $Rf : \mathbf{H}_d \to [0, \infty)$ defined by

$$Rf(\mathcal{H}) = \int_{\mathcal{H}} f(y) d_{\mathcal{H}}(y)$$

where $d_{\mathcal{H}}$ is Lebesgue measure on $\mathcal{H}$. The Radon transform is closely related to the Fourier transform. See [48] for a discussion of these operators and their connections to convex geometry.

1.3 The One Dimensional Case

Let f be a non-vanishing log-concave probability density function on $\mathbb{R}$ associated to a probability measure μ. In particular, $f(t) = e^{-g(t)}$ where $g : \mathbb{R} \to \mathbb{R}$ is convex. For $t \in \mathbb{R}$, define

$$\begin{aligned} J(t) &= \int_{-\infty}^{t} f(s) ds \\ u(t) &= -\log(1 - J(t)) \end{aligned}$$

The cumulative distribution function J is a strictly increasing bijection between $\mathbb{R}$ and $(0, 1)$. The following lemma is a standard result (see e.g. Theorem 5.1 in [51] for the statement, and the references given there). However we include a short proof here for completeness.

Lemma 5. *u is convex*

Proof. Assume momentarily that $g \in C^2(\mathbb{R})$. For $t \in (0,1)$ define

$$\psi(t) = f(J^{-1}(1-t))$$

Note that

$$\psi''(t) = \frac{-g''(J^{-1}(1-t))}{\psi(t)} \leq 0$$

Hence ψ is concave. In addition, $\lim_{t\to 0}\psi(t) = \lim_{t\to 1}\psi(t) = 0$. Hence, the function $\kappa(t) = \psi(t)/t$ is non-increasing on $(0,1)$ and the function $f(t)/(1-J(t)) = \kappa(1-J(t))$ is non-decreasing on $\mathbb{R}$. Since $u'(t) = f(t)/(1-J(t))$, u is convex.

If $g \notin C^2(\mathbb{R})$, then the result follows by approximation (convolve μ with a Gaussian). $\square$

Lemma 6. *If* $T \geq 1$ *and* $x > 2T\log T$, *then* $(\log x)/x < T^{-1}$.

Proof. Since the function $y = e^{-1}x$ is tangent to the strictly concave function $y = \log x$, the function $y = (\log x)/x$ has a global maximum of e^{-1} and is decreasing on $[e,\infty)$. We now consider two cases. In case 1, $T < e$ and therefore $(\log x)/x \leq e^{-1} < T^{-1}$. In case 2, $T \geq e$. Since $(\log T)/T < 2^{-1}$, it follows that $\log(2\log T) < \log T$. For $x' = 2T\log T$,

$$\frac{\log x'}{x'} = \frac{\log T + \log(2\log T)}{2T\log T} < \frac{1}{T}$$

Since $x > x' > e$, $(\log x)/x < (\log x')/x'$ and the result follows. $\square$

The following lemma is a quantitative version of the Gnedenko law of large numbers for log-concave probability measures on $\mathbb{R}$.

Lemma 7. *Let* $q \geq 1$ *and* $n \geq 120q^2(2+\log q)^2$. *Let* μ *be a probability measure on* $\mathbb{R}$ *with a non-vanishing log-concave density function and cumulative distribution* *function J, and let $(\gamma_i)_1^n$ be an i.i.d. sample from μ. With probability at least $1 - 2(\log n)^{-q}$,*

$$\frac{\left|\gamma_{(n)} - J^{-1}(1-1/n)\right|}{J^{-1}(1-1/n) - \mathbb{E}\mu} \leq 6q\frac{\log\log n}{\log n} \tag{1.3.1}$$

where $\gamma_{(n)} = \max\{\gamma_i\}_1^n$ and $\mathbb{E}\mu$ denotes the centroid of μ.

Proof. We shall implicitly make use of Lemma 6 several times throughout the proof. Let $a = (\log n)^{-q}$ and $b = q\log n$. It follows that $0 < a < b < ne^{-1}$. Set $s = J^{-1}(1 - b/n)$ and $t = J^{-1}(1-a/n)$. As mentioned in the preliminaries (see also Lemma 3.3 in [12]), $1 - J(\mathbb{E}\mu) \geq e^{-1}$, hence $u(\mathbb{E}\mu) \leq 1$. Since $b/n < e^{-1}$, we have $\mathbb{E}\mu < s < t$. By convexity of u we have the inequality $(s-\mathbb{E}\mu)^{-1}(u(s)-u(\mathbb{E}\mu)) \leq (t-s)^{-1}(u(t)-u(s))$ which can be rewritten as

$$\frac{J^{-1}(1-a/n) - J^{-1}(1-b/n)}{J^{-1}(1-b/n) - \mathbb{E}\mu} \leq \frac{\log b - \log a}{\log n - \log b - 1} \tag{1.3.2}$$

Since $2qe\log\sqrt{n} \leq \sqrt{n}$, it follows that $\log(qe\log n) \leq \frac{1}{2}\log n$ which implies that

$$\frac{\log b - \log a}{\log n - \log b - 1} \leq \frac{3q\log\log n}{\log n - \log(qe\log n)} \leq 6q\frac{\log\log n}{\log n}$$

By independence,

$$\begin{aligned}
\mathbb{P}\{J^{-1}(1-b/n) &\leq \gamma_{(n)} \leq J^{-1}(1-a/n)\} \\
&= \left(1-\frac{a}{n}\right)^n - \left(1-\frac{b}{n}\right)^n \\
&\geq 1 - a - e^{-b} \\
&\geq 1 - 2(\log n)^{-q}
\end{aligned}$$

If the event $\{J^{-1}(1-b/n) \leq \gamma_{(n)} \leq J^{-1}(1-a/n)\}$ occurs, then the event defined by inequality (1.3.1) also occurs. □

Although Lemma 7 applies to the multiplicative version of the Gnedenko law of large numbers, it also recovers the additive version as long as

$$J^{-1}(1-1/n) = o\left(\frac{\log n}{\log\log n}\right) \tag{1.3.3}$$

If, in the proof, we take $a^{-1} = b = \log_{(m)} n$ (the mth iterate of the logarithm), then the probability bound becomes $1 - 2(\log_{(m)} n)^{-1}$, and the right hand side of (1.3.1) becomes

$$\frac{4\log_{(m+1)} n}{\log n}$$

provided $n > n_0(m)$.

1.4 Main Proofs

Since Lebesgue measure depends on the underlying Euclidean structure of $\mathbb{R}^d$, so does the definition of $f = d\mu/dx$, and therefore also the definition of $D_\delta = Cl\{x : f(x) \geq \delta\}$. A natural variation of the body D_δ which does not depend on Euclidean structure is the body

$$D_\delta^\natural = Cl\{x \in \mathbb{R}^d : f(x) \geq \tau_d^{-1} 9^d |\det cov(\mu)|^{-1/2}\delta\}$$

where the quantity

$$\tau_d = \mathrm{vol}_{d-1}(B_2^{d-1}) \int_{1/2}^{1} (1-t^2)^{(d-1)/2} dt \tag{1.4.1}$$

represents the volume of the set $\{x \in \mathbb{R}^d : ||x||_2 \leq 1,\, x_1 \geq 1/2\}$. Associated to $D_\delta^\natural$ are three ellipsoids that play a central role in our proof. The John ellipsoid of $D_\delta^\natural$ is denoted $\mathcal{E}_{D_\delta^\natural}$ and the centroid of $\mathcal{E}_{D_\delta^\natural}$ will be denoted $\mathcal{O}_\delta$. We also consider

$$\mathcal{E}_\delta^\sharp = 3d(\mathcal{E}_{D_\delta^\natural} - \mathcal{O}_\delta) + \mathcal{O}_\delta \tag{1.4.2}$$

and

$$\mathcal{E}^\flat_\delta = \frac{1}{2}(\mathcal{E}_{D^\natural_\delta} - \mathcal{O}_\delta) + \mathcal{O}_\delta \tag{1.4.3}$$

The advantage of using $D^\natural_\delta$ is that we may place μ in different positions at various stages of our analysis. We first position μ to be isotropic and then position it so that $\mathcal{E}_{D^\natural_\delta} = B_2^d$. We include the proofs of Lemma 8 and Lemma 9 in Section 1.5.

Lemma 8. *There exists a universal constant $c > 0$ with the following property. Let $d \in \mathbb{N}$, let μ be a log-concave probability measure with a continuous density function f, and let $\delta < c\exp(-5d^2 \log d)$. Let $\mathfrak{H}$ be a half-space (either open or closed) with $\mu(\mathfrak{H}) = \delta$ and let $\mathcal{E}^\sharp_\delta$ and $\mathcal{E}^\flat_\delta$ be defined by (1.4.2) and (1.4.3) respectively. Then*

$$\begin{aligned} \mathfrak{H} \cap \mathcal{E}^\sharp_\delta &\neq \emptyset \qquad (1.4.4) \\ \mathfrak{H} \cap \mathcal{E}^\flat_\delta &= \emptyset \qquad (1.4.5) \end{aligned}$$

Consequently,

$$\mathcal{E}^\flat_\delta \subset F_\delta \subset \mathcal{E}^\sharp_\delta$$

We shall use the Euclidean structure corresponding to $\mathcal{E}_{D^\natural_\delta}$ in order to compare $F_{1/n}$ and P_n. The following lemma together with Lemma 8 allows us to do so.

Lemma 9. *Let $d \in \mathbb{N}$ and let K and L be convex bodies in $\mathbb{R}^d$ such that $rB_2^d \subset K \subset RB_2^d$ for some $r, R > 0$. Let $0 < \rho < 1/2$ and $0 < \varepsilon < (16R/r)^{-1}$, and let $\mathcal{N}$ be an ε-net in S^{d-1}. Suppose that for each $\omega \in \mathcal{N}$,*

$$(1-\rho)||\omega||_L \leq ||\omega||_K \leq (1+\rho)||\omega||_L \tag{1.4.6}$$

Then for all $x \in \mathbb{R}^d$ we have

$$(1 + 2\rho + 28Rr^{-1}\varepsilon)^{-1}||x||_L \leq ||x||_K \leq (1 + 2\rho + 28Rr^{-1}\varepsilon)||x||_L \tag{1.4.7}$$

In particular,

$$d_{\mathfrak{L}}(K, L) \leq d_{\mathfrak{L}}(K, L, 0) \leq 1 + 2\rho + 28Rr^{-1}\varepsilon \tag{1.4.8}$$

Proof of Theorem 1 . By convolving μ with a Gaussian measure of the form

$$\phi_{\lambda,d}(x) = \lambda^{-d}\phi_d(\lambda^{-1}x)$$

where $\phi_d(x) = (2\pi)^{-d}\exp(-2^{-1}||x||_2^2)$ is the standard normal density function, and taking $\lambda \to 0$, we may assume that the density of μ is continuous and non-vanishing. This is possible because the bounds in the theorem do not depend on μ. The condition $n \geq c \exp\exp(5d) + c'q^3$ insures that the probability bound is non-trivial. It is also sufficiently large so that we may use Lemma 8 with $\delta = 1/n$ and Lemma 7 with $q' = d + q$. Let $\varepsilon = (\log n)^{-1}$. By applying a suitable affine transformation, we may assume that $\mathcal{E}_{D^{\natural}_{1/n}} = B_2^d$. By Lemma 8, if $\mathfrak{H}_{\theta,t}$ is a half-space with $\mu(\mathfrak{H}_{\theta,t}) = 1/n$, then

$$1/2 \leq t \leq 3d \tag{1.4.9}$$

This implies that $1/2B_2^d \subset F_{1/n} \subset 3dB_2^d$. For each $\theta \in S^{d-1}$, the function $f_\theta(t) = -\frac{d}{dt}\mu(\mathfrak{H}_{\theta,t})$ is the density of a log-concave probability measure μ_θ on $\mathbb{R}$ with cumulative distribution function $J_\theta(t) = 1 - \mu(\mathfrak{H}_{\theta,t})$. Furthermore, the sequence $(\langle\theta, x_i\rangle)_{i=1}^n$ is an i.i.d. sample from this distribution. Recalling the definition of the dual Minkowski functional, for any $y \in \mathbb{R}^d$

$$\begin{aligned} ||y||_{P_n^\circ} &= \sup\{\langle x, y\rangle : x \in P_n\} \\ &= \max_{i=1\ldots n} \langle x_i, y\rangle \end{aligned}$$

We use this notation even when $0 \notin P_n$. Let $\mathcal{N}$ denote a generic ε-net in S^{d-1} and consider the function

$$\widetilde{f}_{\mathcal{N}}(x) = \inf\{\mu(\mathfrak{H}_{\omega,t}) : \omega \in \mathcal{N},\ t = \langle\omega, x\rangle\}$$

For all $\delta > 0$, define the discrete floating body

$$F_\delta^{\mathcal{N}} = \{x \in \mathbb{R}^d : \widetilde{f}_{\mathcal{N}}(x) \geq \delta\}$$

Note that $\widetilde{f}(x) = \inf_{\mathcal{N}} \widetilde{f}_{\mathcal{N}}(x)$ and $F_\delta = \cap_{\mathcal{N}} F_\delta^{\mathcal{N}}$, where $\mathcal{N}$ runs through the collection of all ε-nets in S^{d-1}. By (1.4.9), $\frac{1}{2}B_2^d \subset F_{1/n}^{\mathcal{N}} \subset 3dB_{\mathcal{N}}$ and by (1.2.3) we have $1/2B_2^d \subset F_{1/n}^{\mathcal{N}} \subset 4dB_2^d$ which implies that $(4d)^{-1}B_2^d \subset (F_{1/n}^{\mathcal{N}})^\circ \subset 2B_2^d$. For each $\theta \in S^{d-1}$, we have

$$\begin{aligned} \mathbb{E}\mu_\theta &\geq J_\theta^{-1}(e^{-1}) \\ &\geq J_\theta^{-1}(1/n) \end{aligned}$$

Combining this and (1.3.1), with probability at least $1 - 2(\log n)^{-d-q}$ we have that,

$$\frac{\big| ||\theta||_{P_n^\circ} - J_\theta^{-1}(1-1/n) \big|}{J_\theta^{-1}(1-1/n) - J_\theta^{-1}(1/n)} \leq 6(d+q)\frac{\log\log n}{\log n}$$

Since both $-J_\theta^{-1}(1/n)$ and $J_\theta^{-1}(1-1/n)$ lie in the interval $[1/2, 3d]$, both have roughly the same order of magnitude and we have

$$(1-\rho)J_\theta^{-1}(1-1/n) \leq ||\theta||_{P_n^\circ} \leq (1+\rho)J_\theta^{-1}(1-1/n)$$

where

$$\rho = 42d(d+q)\frac{\log\log n}{\log n}$$

With probability at least $1 - \varepsilon^{-d}3^{d+1}(\log n)^{-d-q} = 1 - 3^{d+1}(\log n)^{-q}$, this happens for all $\omega \in \mathcal{N}$. Hence,

$$(1+\rho)^{-1}P_n \subset F_{1/n}^{\mathcal{N}}$$

which implies that

$$(1-\rho)||\theta||_{P_n^\circ} \leq ||\theta||_{(F_{1/n}^{\mathcal{N}})^\circ}$$

for all $\theta \in S^{d-1}$. On the other hand, for all $\omega \in \mathcal{N}$ we have

$$\begin{aligned} ||\omega||_{P_n^\circ} &\geq (1-\rho)J_\omega^{-1}(1-1/n) \\ &\geq (1-\rho)||\omega||_{(F_{1/n}^{\mathcal{N}})^\circ} \end{aligned}$$

By (1.4.7),

$$(1+4\rho+224d\varepsilon)^{-1}||x||_{P_n^\circ} \leq ||x||_{(F_{1/n}^{\mathcal{N}})^\circ} \leq (1+4\rho+224d\varepsilon)||x||_{P_n^\circ} \tag{1.4.10}$$

for all $x \in \mathbb{R}^d$. Let $\mathcal{M}$ be any other ε-net in S^{d-1}. By the calculations above, with probability at least $1-3^{d+1}(\log n)^{-q}$,

$$(1+4\rho+224d\varepsilon)^{-1}||x||_{P_n^\circ} \leq ||x||_{(F_{1/n}^{\mathcal{M}})^\circ} \leq (1+4\rho+224d\varepsilon)||x||_{P_n^\circ} \tag{1.4.11}$$

for all $x \in \mathbb{R}^d$. By the union bound, with probability at least $1-3^{d+2}(\log n)^{-q} > 0$, both (1.4.10) and (1.4.11) hold. Since both $F_{1/n}^{\mathcal{N}}$ and $F_{1/n}^{\mathcal{M}}$ are deterministic bodies, the only way that this can be true is if

$$(1+4\rho+224d\varepsilon)^{-2}F_{1/n}^{\mathcal{N}} \subset F_{1/n}^{\mathcal{M}} \subset (1+4\rho+224d\varepsilon)^{2}F_{1/n}^{\mathcal{N}}$$

Since $F_{1/n} = \cap_{\mathcal{M}} F_{1/n}^{\mathcal{M}}$, where the intersection is taken over all ε-nets in S^{d-1}, we have

$$(1+4\rho+224d\varepsilon)^{-2}F_{1/n}^{\mathcal{N}} \subset F_{1/n} \subset (1+4\rho+224d\varepsilon)^{2}F_{1/n}^{\mathcal{N}}$$

Combining this with the polar of (1.4.10) gives that with probability at least $1-3^{d+3}(\log n)^{-q}$ we have

$$(1+4\rho+224d\varepsilon)^{-3}P_n \subset F_{1/n} \subset (1+4\rho+224d\varepsilon)^{3}P_n$$

from which the result follows by the inequality $(1+\varepsilon')^3 \leq 1+12\varepsilon'$ valid if $0 \leq \varepsilon' \leq 1$. □

Lemma 10. *Let* $g : \mathbb{R}^d \to [0, \infty)$ *be convex with* $\lim_{x\to\infty} g(x) = \infty$*, let* $K \subset \mathbb{R}^d$ *be a convex body containing* 0 *in its interior, and let* $p > 1$*. Then there exist* $c_1, c_2 > 0$ *such that for all* $x \in \mathbb{R}^d$*,*

$$g(x)^p \geq c_1||x||_K^p - c_2 \tag{1.4.12}$$

Proof. We leave the easy proof of this to the reader. □

Lemma 11. *Let* $p > 1$*,* $d \in \mathbb{N}$ *and let* μ *be a* p*-log-concave probability measure on* $\mathbb{R}^d$*. Then there exist* $c_1, c_2, t_0 > 0$ *such that for all* $\theta \in S^{d-1}$ *and all* $t \geq t_0$*,*

$$\mu(\mathfrak{H}_{\theta,t}) \leq c_1 t^{1-p} e^{-c_2 t^p} \tag{1.4.13}$$

where $\mathfrak{H}_{\theta,t} = \{x \in \mathbb{R}^d : \langle x, \theta \rangle \geq t\}$*.*

Proof. For all $t \geq 1$ we have

$$\begin{aligned} e^{-t^p} &\leq -\frac{d}{dt}\left(p^{-1}t^{1-p}e^{-t^p}\right) \\ &= p^{-1}(p-1)t^{-p}e^{-t^p} + e^{-t^p} \\ &\leq p^{-1}(2p-1)e^{-t^p} \end{aligned}$$

Hence, by the fundamental theorem of calculus,

$$(2p-1)^{-1}t^{1-p}e^{-t^p} \leq \int_t^\infty e^{-s^p} ds \leq p^{-1}t^{1-p}e^{-t^p} \tag{1.4.14}$$

Since the image of a p-log-concave probability measure under an orthogonal transformation is p-log-concave, we may assume without loss of generality that $\theta = e_1 = (1, 0, 0, 0 \ldots)$. By (1.4.12), there exist $c_1, c_2 > 0$ such that for all $x \in \mathbb{R}^d$,

$$f(x) \leq c_1 e^{-c_2||x||_p^p}$$

where $||x||_p^p = \sum_{i=1}^d |x_i|^p$. Hence,

$$\begin{aligned} \mu(\mathfrak{H}_{\theta,t}) &\leq \int_{\mathfrak{H}_{\theta,t}} c_1 e^{-c_2||x||_p^p} dx \\ &= \int_t^\infty c_3 e^{-c_2 s^p} ds \end{aligned} \tag{1.4.15}$$

The result now follows from a change of variables, (1.4.15) and (1.4.14). □

Proof of Theorem 2. Let c_1, c_2 and t_0 be the constants appearing in Lemma 11. Let $n_0 > c_1 + \exp(2^{-1}c_2 t_0^p)$. Without loss of generality, $t_0 > 1$ and $n > n_0$. Set $\alpha = (2c_2^{-1}\log n)^{1/p}$ and consider any $x \in \mathbb{R}^d$ with $||x||_2 > \alpha$. Let $\theta = ||x||_2^{-1}x$ and $t = (\alpha + ||x||_2)/2$. Since $t > \alpha > t_0$ and $n > c_1$, Lemma 11 implies that

$$\mu(\mathfrak{H}_{\theta,t}) < c_1 n^{-2} < n^{-1}$$

Since $||x||_2 > t$, $x \in int(\mathfrak{H}_{\theta,t})$. By definition of the floating body, $x \notin F_{1/n}$. Since this is true for all such x, $diam(F_{1/n}) \leq 2\alpha = c_4(\log n)^{1/p}$. The result now follows from Theorem 1 and the relation (1.2.6) between the Hausdorff and the logarithmic Hausdorff distances. □

1.5 Technical Lemmas

This section contains some technical results on the rigidity of log-concave functions that enable us to obtain a lower bound on the sample size.

Lemma 12. *There exist universal constants $c_1, c_2 > 0$ such that for all $d \in \mathbb{N}$,*

$$c_1^d d^{-d/2} \leq \mathrm{vol}_d(B_2^d) \leq c_2^d d^{-d/2}$$

Proof. This follows from Stirling's formula and the expression $\mathrm{vol}_d(B_2^d) = \pi^{d/2}(\Gamma(1 + d/2))^{-1}$ (see Corollary 2.20 in [48] or p.11 in [59]). □

Lemma 13. *There exists a universal constant $c > 0$ with the following property. Let $d \in \mathbb{N}$ and let μ be an isotropic log-concave probability measure on $\mathbb{R}^d$ with density function f. For all $x \in \mathbb{R}^d$,*

$$f(x) \leq e^{-\alpha_d ||x||_2 + \beta_d}$$

where $\alpha_d = c^d d^{-d/2}$ and $\beta_d = 11d\log(d) + 7$.

Proof. We first consider the case $d \geq 2$. The volume of a cone in $\mathbb{R}^d$ with height h and base radius r is $d^{-1}r^{d-1}h\mathrm{vol}_{d-1}(B_2^{d-1})$. For any $x \in \mathbb{R}^d$, let A_x be the cone with vertex x and base $(1/9)B_2^d \cap x^\perp$. Then $\mathrm{vol}_d(A_x) = d^{-1}9^{-d+1}||x||_2\mathrm{vol}_{d-1}(B_2^{d-1}) > e^{-4d+3}||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})$. By log-concavity of f and inequality (1.2.13), for all $y \in A_x$,

$$f(y) \geq \min\{f(x), 2^{-8d}\} \tag{1.5.1}$$

If $f(x) \geq 2^{-8d}$, then

$$1 \geq \int_{A_x} f(y)dy \geq 2^{-8d}\mathrm{vol}_d(A_x) > e^{-10d+3}||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})$$

and it follows that

$$||x||_2 < \frac{e^{10d-3}}{\mathrm{vol}_{d-1}(B_2^{d-1})}$$

Hence, if $||x||_2 \geq e^{10d-3}/\mathrm{vol}_{d-1}(B_2^{d-1})$ then $f(x) < 2^{-8d}$ and (1.5.1) becomes $f(y) \geq f(x)$. Then

$$1 \geq \int_{A_x} f(y)dy \geq f(x)\mathrm{vol}_d(A_x) > f(x)e^{-4d+3}||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})$$

which implies that

$$f(x) < \frac{e^{4d-3}}{\mathrm{vol}_{d-1}(B_2^{d-1})}||x||_2^{-1}$$

If $\widetilde{x} \in \mathbb{R}^d$ obeys $||\widetilde{x}||_2 = e^{10d-3}/\mathrm{vol}_{d-1}(B_2^{d-1})$, then

$$f(\widetilde{x}) < e^{-6d} \tag{1.5.2}$$

For any $x \in \mathbb{R}^d$ with $||x||_2 \geq e^{10d-3}/\mathrm{vol}_{d-1}(B_2^{d-1})$, we have the convex combination

$$\frac{e^{10d-3}}{\mathrm{vol}_{d-1}(B_2^{d-1})}\frac{x}{||x||_2} = \frac{e^{10d-3}}{||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})}x + \left(1 - \frac{e^{10d-3}}{||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})}\right)0$$

Set

$$\widetilde{x} = \frac{e^{10d-3}}{\mathrm{vol}_{d-1}(B_2^{d-1})}\frac{x}{||x||_2}$$

Using concavity of $\log f$ and inequality (1.5.2),

$$-6d \geq \left(\frac{e^{10d-3}}{||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})}\right)\log f(x) + \left(1 - \frac{e^{10d-3}}{||x||_2\mathrm{vol}_{d-1}(B_2^{d-1})}\right)\log f(0)$$

After some simplification, and using inequality (1.2.11), we get

$$f(x) \leq \exp\left(-de^{-10d+3}\mathrm{vol}_{d-1}(B_2^{d-1})||x||_2 - 7d\log 2\right)$$

If, on the other hand, $||x||_2 < e^{10d-3}/\mathrm{vol}_{d-1}(B_2^{d-1})$, then by (1.2.12)

$$\begin{aligned} f(x) &\leq ||f||_\infty \leq d^{d/2}2^{8d} \\ &\leq \exp(-de^{-10d+3}\mathrm{vol}_{d-1}(B_2^{d-1})||x||_2 + 11d\log d) \end{aligned}$$

The case $d = 1$ is simpler and we leave the details to the reader. First show that $f(2^8) \leq 2^{-8}$ and then proceed as in the case $d \geq 2$ to obtain $f(x) \leq \exp(-2^{-9}|x| + 7)$ for all $x \in \mathbb{R}$. The result now follows from Lemma 12. □

Corollary 14. *There exist universal constants $c_1, c_2 > 0$ with the following property. Let $d \in \mathbb{N}$ and let μ be an absolutely continuous isotropic log-concave probability measure. For all $\delta < e^{-11d\log d - 7}$,*

$$D_\delta \subset c_1^d d^{d/2}(\log \delta^{-1})B_2^d$$

In particular, $\mathrm{vol}_d(D_\delta) \leq c_2 \exp(d^2 \log d)(\log \delta^{-1})^d$.

Proof. By (1.2.11), $D_\delta \neq \emptyset$. By the bounds on δ, it follows that $11d \log d + 7 \leq \log \delta^{-1}$. The result now follows from lemmas 13 and 12. $\square$

Lemma 15. *There exists a universal constant $c > 0$ with the following property. Let $d \in \mathbb{N}$ and let μ be an isotropic log-concave probability measure with density f. Let $r > 1$ and $x \in \mathbb{R}^d$. If $f(x) < 2^{-8d}$ then*

$$f(rx) \leq f(x) \exp(-c^d d^{-d/2}(r-1)||x||_2) \tag{1.5.3}$$

Proof. Let $g = -\log f$. By Lemma 13 and Lemma 12, there exists a universal constant $c_2 > 0$ such that $f(\widetilde{x}) \leq e^{-6d}$ for all $\widetilde{x}$ with $||\widetilde{x}||_2 \geq c_2^d d^{d/2}$, see in particular (1.5.2). Let $x \in \mathbb{R}^d$ be the point specified in the statement of the lemma. We consider two cases. In the first case $||x||_2 \geq c_2^d d^{d/2}$. Let $\widetilde{x} = c_2^d d^{d/2} ||x||_2^{-1} x$. By inequality (1.2.11), $f(0) \geq 2^{-7d}$. By convexity of g and the definition of c_2,

$$\frac{g(rx) - g(x)}{(r-1)||x||_2} \geq \frac{g(\widetilde{x}) - g(0)}{||\widetilde{x}||} = ||\widetilde{x}||_2^{-1} \ln \frac{f(0)}{f(\widetilde{x})} \geq c_2^{-d} d^{1-d/2}$$

In the second case, $||x||_2 < c_2^d d^{d/2}$. Recall that, by hypothesis, $f(x) < 2^{-8d}$. Therefore,

$$\frac{g(rx) - g(x)}{(r-1)||x||_2} \geq \frac{g(x) - g(0)}{||x||} \geq ||x||_2^{-1} \ln \frac{f(0)}{f(x)} \geq \ln(2) c_2^{-d} d^{1-d/2}$$

from which the result follows with $c = (2c_2)^{-1}$. $\square$

Lemma 16. *There exists a universal constant $c_1 > 0$ with the following property. Let $d \in \mathbb{N}$ and let μ be an isotropic log-concave probability measure with a continuous density function f. For all $\delta < e^{-11d \log d - 7}$,*

$$\mu(\mathbb{R}^d \backslash D_\delta) \leq \alpha_d \delta (\log \delta^{-1})^d \tag{1.5.4}$$

where $\alpha_d = c_1 \exp(3d^2 \log d)$.

Proof. Since f is continuous, for all $\theta \in \partial D_\delta$ we have $f(\theta) = \delta$. By the polar integration formula (1.2.9) and inequality (1.5.3),

$$\begin{aligned} \mu(\mathbb{R}^d \backslash D_\delta) &= \int_{\mathbb{R}^d \backslash D_\delta} f(x)dx \\ &= d\int_1^\infty \int_{\partial D_\delta} r^{d-1} f(r\theta) d\mu_{D_\delta}(\theta)dr \\ &\leq d\int_1^\infty \int_{\partial D_\delta} r^{d-1}\delta \exp(-c^d d^{-d/2}(r-1)||\theta||_2)d\mu_{D_\delta}(\theta)dr \end{aligned}$$

By (1.2.13) and the fact that $\delta < 2^{-8d}$, we have $1/9B_2^d \subset D_\delta$. By Corollary 14,

$$\begin{aligned} \mu(\mathbb{R}^d \backslash D_\delta) &\leq d\int_1^\infty \int_{\partial D_\delta} r^{d-1}\delta \exp(-c^d d^{-d/2}(r-1)9^{-1})d\mu_{D_\delta}(\theta)dr \\ &= \delta \mathrm{vol}_d(D_\delta) d \int_1^\infty r^{d-1}\exp(-c_2^d d^{-d/2}(r-1))dr \\ &\leq \beta_d \delta (\log \delta^{-1})^d d \exp(d^2 \log d + c_3)) \end{aligned}$$

where

$$\beta_d = \int_1^\infty r^{d-1} \exp(-c_2^d d^{-d/2}(r-1))dr$$

Set $\omega_d = c_2^d d^{-d/2}$ and $t = \omega_d r$. Recall the definition of the gamma function $\Gamma(z) = \int_0^\infty e^{-r} r^{z-1} dr$. By a change of variables and Stirling's formula,

$$\begin{aligned} \beta_d &\leq \exp(\omega_d) \int_0^\infty r^{d-1}\exp(-\omega_d r)dr \\ &\leq c_4 \omega_d^{-d} \int_0^\infty t^{d-1} e^{-t} dt \\ &\leq \exp(2^{-1} d^2 \log d + c_5 d^2) \end{aligned}$$

from which the result follows. □

Lemma 16 is optimal in δ up to a factor $\log \delta^{-1}$ as can be seen from the example $f(x) = 2^{-d}\exp(-||x||_2)$, in which case $\mu(\mathbb{R}^d \backslash D_\delta) \geq c_d \delta (\log \delta^{-1})^{d-1}$ for $\delta < \delta_0(d)$. To see this, apply the polar integration formula just as in the proof of Lemma 16.

Lemma 17. *There exists a universal constant $c > 0$ such that for all $d \in \mathbb{N}$, if $t > d^{5d}$ then $\sqrt{t} \geq c(\log t)^d$.*

Note that the inequality fails for $t = d^{2d}$.

Proof. There exists $d_0 \in \mathbb{N}$ such that for all $d > d_0$, $(2d)^{4d} < d^{5d}$. Consider any such d. Set $T = 2d$ and $x = \log t$. Since $(2d)^{4d} < d^{5d} < t$, it follows that $2T \log T < x$. By Lemma 6, $(\log x)/x < T^{-1}$, or equivalently

$$\frac{\log \log t}{\log t} < \frac{1}{2d}$$

which is in turn equivalent to $\sqrt{t} \geq (\log t)^d$. By elementary analysis, the number

$$c' = \inf\{t^{1/2}(\log t)^{-d} : d \leq d_0, t > d^{5d}\}$$

is strictly positive and the result follows with $c = \min\{c', 1\}$. $\square$

Lemma 18. *There exists a universal constant $\widetilde{c} > 0$ with the following property. Let $d \in \mathbb{N}$ and let μ be an isotropic log-concave probability measure with a continuous density function f. For all $\delta < \widetilde{c} \exp(-5d^2 \log d)$,*

$$\mu(\mathbb{R}^d \backslash 2D_{\tau^{-1}9^d\delta}) < \delta$$

where $\tau = \tau_d = \mathrm{vol}_{d-1}(B_2^{d-1}) \int_{1/2}^1 (1-t^2)^{(d-1)/2} dt$.

Proof. Consider the quantity $\alpha_d = \exp(d^2(2\log d + c_1))$. By concavity, $1 - t^2 \geq 3(1-t)/2$ for all $1/2 \leq t \leq 3/4$. By a change of variables and Lemma 12, one sees that $\tau > c_2^d d^{-d/2}$. Let $\kappa = \tau^{-1}9^d\delta$. Consider any $y \in \partial(2D_\kappa)$. Then $x = y/2 \in \partial D_\kappa$ and we have the convex combination $x = \frac{1}{2}0 + \frac{1}{2}y$. By log-concavity, $f(x) \geq f(0)^{1/2} f(y)^{1/2}$ and by inequality (1.2.11),

$$f(y) \leq \frac{f(x)^2}{f(0)} < 2^{8d}\kappa^2$$

and $y \notin D_\varepsilon$ with $\varepsilon = 2^{8d}\kappa^2$. Since this is true for all $y \in \partial(2D_\kappa)$, $D_\varepsilon \subset 2D_\kappa$. For a sufficiently small choice of $\widetilde{c}$ (chosen independently of d), $\varepsilon < e^{-11d\log d - 7}$. By Lemma 16,

$$\begin{aligned} \mu(\mathbb{R}^d \backslash 2D_\kappa) &\le \mu(\mathbb{R}^d \backslash D_\varepsilon) \le \alpha_d \varepsilon (\log \varepsilon^{-1})^d \\ &\le e^{10d}\alpha_d \tau^{-2}\delta^2(2\log\delta^{-1} - \log(c_3^d\tau^{-2}))^d \\ &\le e^{10d}\alpha_d\tau^{-2}\delta^2(\log\delta^{-1})^d \\ &= \delta(ce^{10d}\alpha_d\tau^{-2}\delta^{1/2})c^{-1}\delta^{1/2}(\log\delta^{-1})^d \end{aligned}$$

where c is the constant from Lemma 17. By the bound imposed on δ, $ce^{10d}\alpha_d\tau^{-2}\delta^{1/2} < 1$. The result now follows from Lemma 17. $\square$

Recall that $\mathcal{E}_K$ denotes the John ellipsoid of a convex body K and that $\mathcal{E}_K \subset K \subset d(\mathcal{E}_K - x_0) + x_0$, where x_0 is the center of $\mathcal{E}_K$.

Lemma 19. *Let $K \subset \mathbb{R}^d$ be a convex body with $0 \in K$. Then $2K \subset 3d(\mathcal{E}_K - x_0) + x_0$.*

Proof. By applying a suitable linear transformation, we may assume that $\mathcal{E}_K = B_2^d + x_0$. Take any $x \in K$. Since $\max\{||x_0 - x||_2, ||x_0||_2\} \le d$, it follows that $||x||_2 \le ||x_0||_2 + ||x - x_0||_2 \le 2d$ and that $||x_0 - 2x||_2 \le ||x_0 - x||_2 + ||x - 2x||_2 \le 3d$. $\square$

Lemma 20. *Let $\mathcal{E}$ be an ellipsoid with centroid $\mathcal{O}$ and let $\mathfrak{H}$ be a hyperplane with* $\mathrm{vol}_d(\mathfrak{H} \cap \mathcal{E}) \times \mathrm{vol}_d(B_2^d) < \tau_d \mathrm{vol}_d(\mathcal{E})$. *Then $\mathfrak{H}$ and $\frac{1}{2}(\mathcal{E} - \mathcal{O}) + \mathcal{O}$ are disjoint.*

Proof. The truth of the lemma is invariant under affine transformations of $\mathcal{E}$ and we may therefore assume that $\mathcal{E} = B_2^d$. The result now follows from the definition of τ_d (see equation (1.4.1) and the fact that $\tau_d = \mathrm{vol}_d\{x \in \mathbb{R}^d : ||x||_2 \le 1,\, x_1 \ge 1/2\}$. $\square$

Proof of Lemma 8. Consider $\tau = \tau_d$ defined by (1.4.1). We may assume that μ is in isotropic position, which implies that $D^{\natural}_{\delta} = D_{\tau^{-1}9^d\delta}$. Lemma 18 and Lemma 19 together imply that $\mathfrak{H} \cap \mathcal{E}^{\sharp}_{\delta} \neq \emptyset$. For each $x \in D^{\natural}_{\delta}$, $f(x) \geq \tau^{-1}9^d\delta$. Therefore $\delta \geq \mu(\mathfrak{H} \cap \mathcal{E}_{D^{\natural}_{\delta}}) \geq \tau^{-1}9^d\delta \text{vol}_d(\mathfrak{H} \cap \mathcal{E}_{D^{\natural}_{\delta}})$ which implies that $\text{vol}_d(\mathfrak{H} \cap \mathcal{E}_{D^{\natural}_{\delta}}) \leq \tau 9^{-d}$. Since the density function f is continuous and $\tau^{-1}9^d\delta < 2^{-8d}$, inequality (1.2.11) implies that $(9^{-1} + \kappa)B^d_2 \subset D^{\natural}_{\delta}$ for some $\kappa > 0$. Since $\mathcal{E}_{D^{\natural}_{\delta}}$ is the ellipsoid of maximal volume inside $D^{\natural}_{\delta}$, we have $\text{vol}_d(\mathcal{E}_{D^{\natural}_{\delta}}) > 9^{-d}\text{vol}_d(B^d_2)$. From the definition of $\mathcal{E}^{\flat}_{\delta}$ and Lemma 20, we see that $\mathfrak{H} \cap \mathcal{E}^{\flat}_{\delta} = \emptyset$. Finally, the claim that $\mathcal{E}^{\flat}_{\delta} \subset F_{\delta}$ follows from the definition of F_{δ} while the claim that $F_{\delta} \subset \mathcal{E}^{\sharp}_{\delta}$ follows from the Hahn-Banach theorem (any $x \notin \mathcal{E}^{\sharp}_{\delta}$ lies in an open half-space $\mathfrak{H}$ with $\mathfrak{H} \cap \mathcal{E}^{\sharp}_{\delta} = \emptyset$ and therefore $\mu(\mathfrak{H}) < \delta$ and $x \notin F_{\delta}$). $\square$

Proof of Lemma 9. Note that $1+\rho \leq (1-\rho)^{-1} \leq 1+2\rho$ and $1-\rho \leq (1+\rho)^{-1} \leq 1-\rho/2$, and the same inequalities hold for ε. Since $rB^d_2 \subset K \subset RB^d_2$, we have that

$$R^{-1}||x||_2 \leq ||x||_K \leq r^{-1}||x||_2$$

for all $x \in \mathbb{R}^d$. Combining this with (1.4.6) gives

$$R^{-1}(1+\rho)^{-1} \leq ||\omega||_L \leq r^{-1}(1-\rho)^{-1}$$

for all $\omega \in \mathcal{N}$. Consider any $\theta \in S^{d-1}$. By the series representation (1.2.2) and the triangle inequality,

$$||\theta||_L \leq r^{-1}(1-\rho)^{-1}(1-\varepsilon)^{-1}$$

Hence $||x||_L \leq r^{-1}(1-\rho)^{-1}(1-\varepsilon)^{-1}||x||_2$ for all $x \in \mathbb{R}^d$. Using the triangle inequality in a bit of a different way,

$$\begin{aligned}
||\theta||_L &\geq ||\omega_0||_L - \sum_{i=1}^{\infty} \varepsilon_i ||\omega_i||_L \\
&\geq R^{-1}(1+\rho)^{-1} - r^{-1}\varepsilon(1-\varepsilon)^{-1}(1-\rho)^{-1} \\
&\geq R^{-1}/2 - 4r^{-1}\varepsilon \\
&= R^{-1}(1 - 8Rr^{-1}\varepsilon)/2 \\
&\geq (4R)^{-1}
\end{aligned}$$

which holds since $8Rr^{-1}\varepsilon \leq 1/2$. Thus,

$$\begin{aligned}
||\theta||_L &\leq ||\omega_0||_L + ||\theta - \omega_0||_L \\
&\leq (1-\rho)^{-1}||\omega_0||_K + r^{-1}(1-\rho)^{-1}(1-\varepsilon)^{-1}\varepsilon \\
&\leq (1-\rho)^{-1}(||\theta||_K + ||\omega_0 - \theta||_K) + r^{-1}(1-\rho)^{-1}(1-\varepsilon)^{-1}\varepsilon \\
&\leq (1-\rho)^{-1}||\theta||_K + r^{-1}(1-\rho)^{-1}\varepsilon(1 + (1-\varepsilon)^{-1}) \\
&\leq (1-\rho)^{-1}||\theta||_K + Rr^{-1}(1-\rho)^{-1}\varepsilon(1 + (1-\varepsilon)^{-1})||\theta||_K \\
&\leq (1+2\rho)(1 + 3Rr^{-1}\varepsilon)||\theta||_K \\
&\leq (1 + 2\rho + 6Rr^{-1}\varepsilon)||\theta||_K
\end{aligned}$$

where ω_0 is the element of $\mathcal{N}$ that minimizes $||\theta - \omega_0||_2$. On the other hand,

$$\begin{aligned}
||\theta||_K &\leq ||\omega_0||_K + ||\theta - \omega_0||_K \\
&\leq (1+\rho)||\omega_0||_L + r^{-1}\varepsilon \\
&\leq (1+\rho)(||\theta||_L + ||\omega_0 - \theta||_L) + r^{-1}\varepsilon \\
&\leq (1+\rho)||\theta||_L + r^{-1}(1+\rho)(1-\rho)^{-1}(1-\varepsilon)^{-1}\varepsilon + r^{-1}\varepsilon \\
&\leq (1+\rho)||\theta||_L + 7r^{-1}\varepsilon \cdot 4R||\theta||_L \\
&\leq (1 + \rho + 28Rr^{-1}\varepsilon)||\theta||_L
\end{aligned}$$

The result follows by positive homogeneity. $\square$

1.6 Proof of Theorem 3

Fix μ and d as in the statement of Theorem 3. Let f be the density of μ and let $g = -\log f$. All variables in this section depend on both d and μ.

Lemma 21. *There exist $c, \varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$,*

$$\mu(\mathbb{R}^d \backslash D_\varepsilon) < c\varepsilon \left(\log \varepsilon^{-1}\right)^d \tag{1.6.1}$$

Proof. Since μ has a log-concave density, it necessarily has a nonsingular covariance matrix, and there exists an affine map T such that $\mu' = T\mu$ is isotropic. The density of μ' is

$$\widetilde{f}(x) = \det(T^{-1}) f(T^{-1}x)$$

and $D_\varepsilon = T^{-1}\widetilde{D}_{\widetilde{\varepsilon}}$, where $\widetilde{\varepsilon} = \varepsilon \det T^{-1}$ and $\widetilde{D}_{\widetilde{\varepsilon}} = \{x : \widetilde{f}(x) \geq \widetilde{\varepsilon}\}$. Since μ' is isotropic, we may use Lemma (16), which gives

$$\begin{aligned} \mu(\mathbb{R}^d \backslash D_\varepsilon) &= \mu'(\mathbb{R}^d \backslash \widetilde{D}_{\widetilde{\varepsilon}}) \\ &\leq c'\widetilde{\varepsilon}(\log \widetilde{\varepsilon}^{-1})^d \\ &\leq c\varepsilon \left(\log \varepsilon^{-1}\right)^d \end{aligned}$$

$\square$

Lemma 22. *For any $x \in \mathbb{R}^d$ there exist $c', \delta_0 > 0$ and a function $p : (0, \delta_0) \to (0, \infty)$ such that for all $\delta \in (0, \delta_0)$,*

$$p(\delta) \leq c' \frac{\log \log \delta^{-1}}{\log \delta^{-1}} \tag{1.6.2}$$

and

$$F_\delta \subset (1+p)(D_\delta - x) + x \tag{1.6.3}$$

Proof. Let $c > 0$ be the constant in (1.6.1). A brief analysis of the function $t \mapsto ct\left(\log t^{-1}\right)^d$ shows that there exists $\delta_0 > 0$ and a function $\varepsilon = \varepsilon(\delta)$ defined implicitly for all $\delta \in (0, \delta_0)$ by the equation $\delta = c\varepsilon\left(\log \varepsilon^{-1}\right)^d$. We can take δ_0 small enough to ensure that $\varepsilon < \delta$ and that $\log \delta^{-1} < \log \varepsilon^{-1} < 2\log \delta^{-1}$. If we define

$$p(\delta) = 3\frac{\log \varepsilon^{-1} - \log \delta^{-1}}{\log \delta^{-1}}$$

then $\delta^{1+p/2} < \varepsilon$ and (1.6.2) holds. Since D_ε is both compact and convex, for any point $y \in D_\varepsilon$ there exists (by the Hahn-Banach theorem), a closed halfspace $\mathfrak{H}$ with $y \in \mathfrak{H}$ and $\mathfrak{H} \cap D_\varepsilon = \emptyset$. Since $\mathfrak{H} \subset \mathbb{R}^d \backslash D_\varepsilon$, (1.6.1) implies that $\mu(\mathfrak{H}) < \delta$ and by definition of F_δ, $y \notin F_\delta$. This goes to show that $F_\delta \subset D_\varepsilon$. Let $x \in \mathbb{R}^d$. For any $\theta \in S^{d-1}$ consider the function $f_\theta(t) = f(x + t\theta) = e^{-g_\theta(t)}$, $t \in \mathbb{R}$. This notation differs slightly from that in the proof of theorem 1. If ε is small enough then for all $\theta \in S^{d-1}$ there is a unique $v > 0$ such that $f_\theta(v) = \varepsilon$; we denote this number by $f_\theta^{-1}(\varepsilon)$. We may assume that $\delta_0 < \min\{1, f(x)^2\}$. Note that $1 < \delta/\varepsilon < \delta^{-p/2}$ and $\log \delta^{-1} + \log f(x) \geq 1/2 \log \delta^{-1}$. By convexity of g_θ, for any $0 < s < v$ we have $s^{-1}(g_\theta(s) - g_\theta(0)) \leq (v-s)^{-1}(g_\theta(v) - g_\theta(s))$. Taking $v = f_\theta^{-1}(\varepsilon)$ and $s = f_\theta^{-1}(\delta)$, this becomes

$$\begin{aligned} \frac{f_\theta^{-1}(\varepsilon) - f_\theta^{-1}(\delta)}{f_\theta^{-1}(\delta)} &\leq \frac{\log \varepsilon^{-1} - \log \delta^{-1}}{\log \delta^{-1} + \log f(x)} \\ &< p \end{aligned} \tag{1.6.4}$$

Inequality (1.6.4) reduces to $f_\theta^{-1}(\varepsilon) \leq (1+p) f_\theta^{-1}(\delta)$. Since this holds for any $\theta \in S^{d-1}$, $D_\varepsilon \subset (1+p)(D_\delta - x) + x$ and (1.6.3) follows. □

Lemma 23. *There exists* $\delta_0 > 0$ *such that for all* $\delta \in (0, \delta_0)$ *we have the relation*

$$(1 + 8\lambda^{1/d})^{-1}(D_\delta - x') + x' \subset F_\delta \tag{1.6.5}$$

where $\lambda = \mathrm{vol}_d(D_\delta)^{-1}$ *and* x' *is the centroid of* D_δ.

Proof. Let δ_0 be such that $\mathrm{vol}_d(D_{\delta_0}) > 8^d$. We use the notation $(D_\delta)_\lambda$ for the convex floating body with parameter $\lambda > 0$ corresponding to the uniform probability measure on D_δ. If $\mathfrak{H}$ is any half-space with $\mu(\mathfrak{H}) < \delta$, then $\mathrm{vol}_d(\mathfrak{H} \cap D_\delta) < 1$. Hence $(D_\delta)_\lambda \subset F_\delta$, where $\lambda = \mathrm{vol}_d(D_\delta)^{-1}$. The result now follows from inequality (1.2.8). □

Lemma 24. *Let* $K, L \subset \mathbb{R}^d$ *be convex bodies such that there exist* $x, x' \in int(K \cap L)$ *and* $0 < r < (8d)^{-1}$ *for which*

$$(1 + r)^{-1}(K - x) + x \subset L \subset (1 + r)(K - x') + x' \tag{1.6.6}$$

Then

$$d_{\mathfrak{L}}(K, L) \leq 1 + 8dr \tag{1.6.7}$$

Proof. Since the statement of the lemma is invariant under affine transformations of K and L, we may assume without loss of generality that the John ellipsoid of K is B_2^d. Hence $B_2^d \subset K \subset dB_2^d$ and $||x||_2, ||x'||_2 \leq d$. Note also that $L \subset 3dB_2^d$. Using these facts and manipulating (1.6.6) in the obvious way, we see that both of the following relations hold

$$\begin{aligned} L &\subset K + 2drB_2^d \\ K &\subset L + 4drB_2^d \end{aligned}$$

By definition of the Hausdorff distance, $d_{\mathcal{H}}(K, L) \leq 4dr$. Since $B_2^d \subset K$, $d_{\mathfrak{L}}(K, L) \leq (1 - 4dr)^{-1} \leq 1 + 8dr$. □

Proof of equation (1.1.4). Since $\lim_{\delta\to 0} p(\delta) = \lim_{\delta\to 0}\lambda(\delta) = 0$, equation (1.1.4) now follows from (1.6.3), (1.6.5) and (1.6.7). □

Remark 1. *There is no lower bound on the growth rate of* $\mathrm{vol}_d(D_\delta)$*, indeed the function could grow arbitrarily slowly. However in the case of the Schechtman-Zinn distributions,* $\mathrm{vol}_d(D_\delta) = (\log(c_p^d/\delta))^{d/p}\mathrm{vol}_d(B_p^d)$ *and we leave it to the reader to combine this with (1.6.3), (1.6.2) and (1.6.5) to obtain a quantitative upper bound on* $d_{\mathfrak{L}}(F_\delta, D_\delta)$*.*

Proof of equation (1.1.5). Let $\varepsilon > 0$ be given. Using the notation from the proof of theorem 1, for any $\theta \in S^{d-1}$ we define

$$f_\theta(t) = -\frac{d}{dt}\mu(\mathfrak{H}_{\theta,t})$$

This function is the density of a log-concave probability measure on $\mathbb{R}$ with cumulative distribution function $J_\theta(t) = 1 - \mu(\mathfrak{H}_{\theta,t})$. By Fubini's theorem we have

$$f_\theta(t) = Rf(\mathcal{H}_{\theta,t})$$

Define $\alpha = \inf\{f_\theta(0) : \theta \in S^{d-1}\}$. By (1.2.10) there exists $t_0 > 0$ such that if $\beta = \sup\{f_\theta(t_0) : \theta \in S^{d-1}\}$, then $\beta < \alpha$. Since f is non-vanishing, S^{d-1} is compact and the function $\theta \mapsto f_\theta(t)$ is continuous, $\beta > 0$. Define $g_\theta(t) = -\log f_\theta(t)$ and let $\lambda = t_0^{-1}(\log\alpha - \log\beta)$ and $\Delta = \max\{1, \lambda^{-1}\log\lambda^{-1}\}$. By definition of α, β and λ, for all $\theta \in S^{d-1}$ we have $t_0^{-1}(g_\theta(t_0) - g_\theta(0)) \geq \lambda$. By convexity of g_θ, if $u > v \geq t_0$ then $g_\theta(u) \geq g_\theta(v) + \lambda(u - v)$, which translates into $f_\theta(u) \leq f_\theta(v)e^{-\lambda(u-v)}$. Let $\delta_0 < \inf\{f_\theta(t_0 + 1) : \theta \in S^{d-1}\}$ be such that $\Delta\varepsilon^{-1}B_2^d \subset F_{\delta_0}$. Consider any $\delta < \delta_0$ and momentarily fix $\theta \in S^{d-1}$. Let $s = J_\theta^{-1}(\delta)$ and denote by $t = f_\theta^{-1}(\delta)$ the unique positive number such that $f_\theta(t) = \delta$. Consider the hyperplane $\mathcal{H}_{\theta,t}$ and the half-space $\mathfrak{H}_{\theta,s}$. Note that

$$\mu(\mathfrak{H}_{\theta,s}) = Rf(\mathcal{H}_{\theta,t}) = \delta$$

By log-concavity we have $f_\theta(u) \geq \delta_0$ for all $0 < u < t_0 + 1$, hence $t > t_0 + 1$. By the fundamental theorem of calculus and the fact that $f_\theta(u) \geq \delta$ for all $u \in [t-1, t]$ we have

$$\begin{aligned} \mu(\mathfrak{H}_{\theta,t-1}) &> \mu\{x \in \mathbb{R}^d : t-1 \leq \langle \theta, x \rangle \leq t\} \\ &= \int_{t-1}^{t} f_\theta(u) du \\ &\geq \delta \end{aligned}$$

hence $\mathfrak{H}_{\theta,s} \subset \mathfrak{H}_{\theta,t-1}$ which implies that $s > t - 1 > t_0$. Thus, if $s \leq t$ then $|s - t| \leq 1$. If $s > t$ then

$$\begin{aligned} \delta &= \int_{s}^{\infty} f_\theta(u) du \\ &\leq f_\theta(s) \int_{s}^{\infty} e^{-\lambda(u-s)} du \\ &\leq \delta e^{-\lambda(s-t)} \lambda^{-1} \end{aligned}$$

from which it follows that $s - t \leq \lambda^{-1} \log \lambda^{-1}$. Either way, $|s - t| \leq \max\{1, \lambda^{-1} \log \lambda^{-1}\} = \Delta$. Since $\Delta \varepsilon^{-1} B_2^d \subset F_{\delta_0}$, it follows that $(1-\varepsilon)s \leq t \leq (1+\varepsilon)s$. Since this holds for all $\theta \in S^{d-1}$ we have

$$(1-\varepsilon)F_\delta \leq R_\delta \leq (1+\varepsilon)F_\delta$$

□

1.7 Optimality

Let Φ denote the cumulative standard normal distribution on $\mathbb{R}$,

$$\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-\frac{1}{2}s^2} ds$$

By (1.4.14) there exists $c > 0$ such that for all $n \geq 3$,

$$\Phi^{-1}(1 - 1/n) \geq c(\log n)^{1/2} \tag{1.7.1}$$

Lemma 25. *For all $q > 0$ and all $d \in \mathbb{N}$, there exists $c, \widetilde{c} > 0$ such that for all $n \geq d+1$, if $(x_i)_1^n$ is an i.i.d. sample from the standard normal distribution on $\mathbb{R}^d$ and $P_n = conv\{x_i\}_1^n$, then with probability at least $1 - \widetilde{c}(\log n)^{-q(d-1)/2}$ both of the following events occur,*

$$\begin{aligned} d_{\mathcal{H}}(P_n, F_{1/n}) &\geq c(\log n)^{-\frac{1}{2}-q} && (1.7.2) \\ d_{\mathfrak{L}}(P_n, F_{1/n}) &\geq 1 + c(\log n)^{-1-q} && (1.7.3) \end{aligned}$$

Proof. A standard result in approximation theory ([40] p. 326) is that for any polytope $K_m \subset \mathbb{R}^d$ with at most m vertices,

$$d_{\mathcal{H}}(K_m, B_2^d) > c_d \left(\frac{1}{m}\right)^{\frac{2}{d-1}} \tag{1.7.4}$$

Since $F_{1/n} = \Phi^{-1}(1 - 1/n)B_2^d$, inequality (1.7.1) implies that

$$d_{\mathcal{H}}(K_m, F_{1/n}) > c_d(\log n)^{1/2} \left(\frac{1}{m}\right)^{\frac{2}{d-1}}$$

By a result of Raynaud [61], the number of vertices of P_n, denoted by $f_0(P_n)$, obeys the inequality $\mathbb{E}f_0(P_n) < \widetilde{c}_d(\log n)^{(d-1)/2}$. By Chebyshev's inequality we have

$$\mathbb{P}\{f_0(P_n) > (\log n)^{\frac{(d-1)(q+1)}{2}}\} \leq \frac{\mathbb{E}f_0(P_n)}{(\log n)^{\frac{(d-1)(q+1)}{2}}} < \widetilde{c}_d(\log n)^{-\frac{q(d-1)}{2}}$$

and if the complement of this event occurs, then so does (1.7.2). By (1.7.4) and (1.2.6) we get

$$d_{\mathfrak{L}}(K_m, B_2^d) > 1 + c_d \left(\frac{1}{m}\right)^{\frac{2}{d-1}}$$

Since $d_{\mathfrak{L}}$ is preserved by invertible affine transformations (as per (1.2.5)), the same inequality holds for all Euclidean balls. This gives (1.7.3). □

We can choose q to be arbitrarily small, in which case (1.7.2) and (1.7.3) complement Theorem 1 and Theorem 2.

1.8 Proof of Theorem 4

If Ω is a convex subset of a real vector space and $\mathcal{K}_d$ is the collection of all convex bodies in $\mathbb{R}^d$, then we define a function $\kappa : \Omega \to \mathcal{K}_d$ to be concave if for all $x, y \in \Omega$ and all $\lambda \in (0,1)$ we have

$$\lambda\kappa(x) + (1-\lambda)\kappa(y) \subset \kappa(\lambda x + (1-\lambda)y)$$

If Ω has an ordering then we define κ to be non-decreasing if for all $x, y \in \Omega$ with $x \leq y$ we have $\kappa(x) \subset \kappa(y)$.

Lemma 26. *If $\kappa : [0,\infty) \to \mathcal{K}_d$ is concave, non-decreasing and $\cup_{t\in[0,\infty)}\kappa(t) = \mathbb{R}^d$, then the function $g : \mathbb{R}^d \to [0,\infty)$ defined by*

$$g(x) = \inf\{t \geq 0 : x \in \kappa(t)\} \tag{1.8.1}$$

is convex. Furthermore, κ is continuous with respect to the Hausdorff distance and for all $t > 0$

$$\kappa(t) = \{x \in \mathbb{R}^d : g(x) \leq t\} \tag{1.8.2}$$

Proof. By hypothesis, $\kappa(0) \neq \emptyset$. If $0 \notin \kappa(0)$, then we define $\kappa^\sharp(t) = \kappa(t) - x_0$, where $x_0 \in \kappa(0)$. The function $\kappa^\sharp$ enjoys all of the properties that κ does, and the function

$$g^\sharp(x) = \inf\{t \geq 0 : x \in \kappa^\sharp(t)\}$$

is related to g by the equation $g^\sharp(x) = g(x + x_0)$. Note that $0 \in \kappa^\sharp(0)$. If the lemma holds for the functions $\kappa^\sharp$ and $g^\sharp$, it will necessarily hold for κ and g. We may therefore, without loss of generality, assume that $0 \in \kappa(0)$. For any $0 < \varepsilon < t$ we have the convex combination

$$t = \frac{\varepsilon}{t+\varepsilon} 0 + \frac{t}{t+\varepsilon}(t+\varepsilon)$$

Exploiting the concavity of κ, this leads to

$$\kappa(t+\varepsilon) \subset \frac{t+\varepsilon}{t}\kappa(t)$$

Similarly,

$$\frac{t-\varepsilon}{t}\kappa(t) \subset \kappa(t-\varepsilon)$$

Hence κ is continuous with respect to the Hausdorff distance. By definition of g, $\kappa(t) \subset \{x \in \mathbb{R}^d : g(x) \leq t\}$. Since $\kappa(t)$ is a closed set, if $x \notin \kappa(t)$ then $d(x, \kappa(t)) > 0$ and by continuity of κ, $g(x) > t$. This implies (1.8.2). Consider any $x, y \in \mathbb{R}^d$ and $\lambda \in (0,1)$. Let $t = g(x)$ and $s = g(y)$. By (1.8.2), $x \in \kappa(t)$ and $y \in \kappa(s)$. Therefore

$$\begin{aligned} \lambda x + (1-\lambda)y &\in \lambda\kappa(t) + (1-\lambda)\kappa(s) \\ &\subset \kappa(\lambda t + (1-\lambda)s) \end{aligned}$$

This implies that $g(\lambda x + (1-\lambda)y) \leq \lambda t + (1-\lambda)s$ which shows that g is convex. $\square$

Note that the function g is a generalization of the Minkowski functional of a convex body K, in which case $\kappa(t) = tK$. Including $\{0\}$ as an honorary member of $\mathcal{K}_d$ does no harm to the preceding lemma. If $(K_n)_{n=1}^{\infty}$ is a sequence of convex bodies then we define the corresponding Minkowski series as,

$$\sum_{n=1}^{\infty} K_n = \left\{ \sum_{n=1}^{\infty} x_n : \forall n,\ x_n \in K_n \right\}$$

where we take $\sum x_n$ to have meaning only if it converges. We leave the easy proof of the following lemma to the reader.

Lemma 27. *For each* $n \in \mathbb{N}$, *let* $\alpha_n : [0, \infty) \to [0, \infty)$ *be a concave function and let* K_n *be a convex body with* $0 \in K_n$. *Provided that*

$$\sum_{n=1}^{\infty} \alpha_n(t) diam(K_n) < \infty$$

for all $t \geq 0$, *then the function* $\kappa : [0, \infty) \to \mathcal{K}_d$ *defined by*

$$\kappa(t) = \sum_{n=1}^{\infty} \alpha_n(t) K_n$$

is concave.

The space $\mathcal{K}_d$ is separable with respect to d_{BM} and we shall use a dence sequence $(K_n)_{n=1}^{\infty}$ that is dense in $\mathcal{K}_d$. Since d_{BM} is blind to affine transformations we can assume that the John ellipsoid of each K_n is B_2^d. As coefficients, we shall use the functions

$$\alpha_n(t) = \begin{cases} 2^{-n^2} t & : \quad 0 \leq t \leq 2^{2n^2} \\ 2^{n^2} & : \quad 2^{2n^2} < t < \infty \end{cases}$$

Note that for large values of n, the dominant coefficient at the value $t = 2^{2n^2}$ is

α_n. In fact $\sum_{j\neq n}\alpha_j(2^{2n^2})$ is much smaller than $\alpha_n(2^{2n^2})$,

$$\begin{aligned}\sum_{j\neq n}\alpha_j(2^{2n^2}) &= \sum_{j=1}^{n-1}2^{j^2}+2^{2n^2}\sum_{j=n+1}^{\infty}2^{-j^2}\\ &\leq \sum_{j=1}^{n-1}2^{nj}+2^{2n^2}\sum_{j=n+1}^{\infty}2^{-nj}\\ &\leq 2^{n^2-n+2}\\ &= 2^{-n+2}\alpha_n(2^{2n^2})\end{aligned}$$

Hence,

$$d_{BM}(\kappa(2^{2n^2}),K_n)\leq 1+2^{-n+2}d$$

Thus the sequence $(\kappa(n))_{n=1}^{\infty}$ is dense in $\mathcal{K}_d$. Since each coefficient α_n is non-decreasing and concave, κ is concave and the function g as defined by (1.8.1) is convex. Clearly, $\lim_{x\to\infty}g(x)=\infty$. For some $c>0$, the function

$$f(x)=2^{-g(cx)}$$

is the density of a log-concave probability measure μ on $\mathbb{R}^d$. For each $n\in\mathbb{N}$, $D_{2^{-n}}=\{x\in\mathbb{R}^d : f(x)\geq 2^{-n}\}=\{x\in\mathbb{R}^d : g(cx)\leq n\}=c^{-1}\kappa(n)$. Hence the sequence $(D_{1/n})_{n=1}^{\infty}$ is dense in $\mathcal{K}_d$. By (1.1.4), the sequence $(F_{1/n})_{n=3}^{\infty}$ is also dense in $\mathcal{K}_d$.

We now use Theorem 1 with $q=1$. Let $\widetilde{\mathcal{K}}_d$ denote a countably dense subset of $\mathcal{K}_d$ and let $K\in\widetilde{\mathcal{K}}_d$. Note that there exists an increasing sequence of natural numbers $(k_n)_1^{\infty}$ such that $\lim_{n\to\infty}d_{BM}(F_{1/k_n},K)=1$ and $\sum_{n=1}^{\infty}\widetilde{c}(\log k_n)^{-1}<\varepsilon$. By (1.1.2),

$$\lim_{n\to\infty}d_{BM}(P_{k_n},K)=1$$

with probability at least $1-\varepsilon$. Since this holds for all $\varepsilon>0$, $K\in cl_{BM}\{P_n : n\in\mathbb{N}, n\geq d+1\}$ almost surely, where cl_{BM} denotes closure in $\mathcal{K}_d$ with respect to d_{BM}.

Since this holds for all $K \in \widetilde{\mathcal{K}}_d$ and $\widetilde{\mathcal{K}}_d$ is countable, $\widetilde{\mathcal{K}}_d \subset cl_{BM}\{P_n : n \in \mathbb{N}, n \geq d+1\}$ almost surely. The result now follows since $\widetilde{\mathcal{K}}_d$ is dense in $\mathcal{K}_d$.

Chapter 2

Simultaneous concentration of order statistics

Let μ be a probability measure on $\mathbb{R}$ with cumulative distribution function F and let $(x_i)_1^\infty$ denote an i.i.d. sequence of random variables with distribution μ. For each $n \in \mathbb{N}$ let F_n denote the empirical cumulative distribution function

$$F_n(t) = \frac{1}{n}|\{i \in \mathbb{N} : i \leq n,\, x_i \leq t\}|$$

where $|A|$ denotes the cardinality of a set A. The Glivenko-Cantelli theorem (see e.g. [20]) states that with probability 1,

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F(t) - F_n(t)| = 0$$

The Dvoretzky-Kiefer-Wolfowitz inequality ([21] and [53]) provides a quantitative formulation of this and states that for all $n \in \mathbb{N}$ and all $\lambda > 0$, with probability at least $1 - 2\exp(-2\lambda^2)$,

$$\sup_{t\in\mathbb{R}} \sqrt{n}|F(t) - F_n(t)| \leq \lambda$$

This titanic theorem is well deserving of the name '*the fundamental theorem of statistics*' as it is the theoretical foundation behind the idea that a large independent sample is representative of the population (see e.g. [65] p.1). Loève refers to it as the '*central statistical theorem*' ([50] p.20) while Pitman ([60] p.79) calls it '*the existence theorem for Statistics as a branch of Applied Mathematics*'. There is a substantial body of literature devoted to the Glivenko-Cantelli theorem; the papers [16], [68], [70] and [72] are of particular interest.

There is, however, a certain crudeness in this noble theorem. Asymptotically, individual points play a negligible role and we learn very little about the finer structure of the sample $\{x_i\}_1^n$. For instance, it gives us almost no information about either the maximum or the minimum. We could take any subset of $o(n)$ points and perturb them as we please without affecting the convergence. Donsker's theorem (see e.g. [19], [49] and [52]) gives more insight into the structure of the sample. Consider the stochastic process X_n defined on $\mathbb{R}$ by

$$X_n(t) = \sqrt{n}(F_n(t) - F(t))$$

Provided that F is strictly increasing and continuous, X_n converges to a re-scaled Brownian bridge (more precisely, $X_n \circ F^{-1}$ converges to a Brownian bridge on $[0, 1]$). However Donsker's theorem is plagued by a similar insensitivity to the cries of the minority. Through the eyes of Donsker's theorem, we can 'see' subsets as small as $\sqrt{n}$ but are blind to anything smaller such as subsets of size $\log(n)$.

In this chapter we provide refined forms of the Glivenko-Cantelli theorem which, under certain conditions, guarantee tight control over all or most points in the sample, not only individually but *simultaneously*. The sequence of order statistics $(x_{(i)})_1^n$ is the non-decreasing rearrangement of the sample $(x_i)_1^n$. Super-exponential decay of the distribution provides simultaneous concentration of *all* order statistics (see theorem 28) while exponential decay provides simultaneous concentration of *most* order statistics and slightly weaker control over the rest (see theorems 29 and 30). We provide quantitative bounds for log-concave distributions (see theorem 31).

Our results extend the Gnedenko law of large numbers [34], which guarantees concentration of $\max\{x_i\}_1^n$. They may be compared to the results in [25] where the Gnedenko law of large numbers is extended to the multi-dimensional setting, to the paper [36] that provides estimates of order statistics in terms of Orlicz norms and to the article [1] that concerns optimal matchings of random points uniformly distributed within the unit square. We refer the reader to [33] and [66] for an extensive treatment of empirical process theory and to [4], [15] and [63] for information on order statistics.

The generalized inverse of F is the function $F^{-1} : (0,1) \to \mathbb{R}$ defined by

$$F^{-1}(x) = \inf\{t : F(t) \geq x\}$$

In theorems 28 and 29 we define $0/0 = 0$ to allow for the case when the measure has bounded support.

Theorem 28. *Let μ be any probability measure on $\mathbb{R}$ with connected support and cumulative distribution function F such that for all $\varepsilon > 0$*

$$\lim_{t\to\infty} \frac{1 - F(t+\varepsilon)}{1 - F(t)} = \lim_{t\to-\infty} \frac{F(t)}{F(t+\varepsilon)} = 0 \tag{2.0.1}$$

Then there exists a sequence $(\delta_n)_1^\infty$ with $\lim_{n\to\infty} \delta_n = 0$ such that for all $n \in \mathbb{N}$, if $(x_i)_1^n$ is an i.i.d. sample from μ with corresponding order statistics $(x_{(i)})_1^n$, then with probability at least $1 - \delta_n$,

$$\sup_{1\leq i\leq n} |x_{(i)} - x_{(i)}^*| \leq \delta_n \tag{2.0.2}$$

where $x_{(i)}^* = F^{-1}(i/(n+1))$.

Theorem 29. *Let μ be any probability measure on $\mathbb{R}$ with connected support and cumulative distribution function F such that for all $\varepsilon > 0$*

$$\begin{aligned} \limsup_{t\to\infty} \frac{1-F(t+\varepsilon)}{1-F(t)} &< 1 \qquad (2.0.3) \\ \limsup_{t\to-\infty} \frac{F(t)}{F(t+\varepsilon)} &< 1 \qquad (2.0.4) \end{aligned}$$

Let $(\omega_n)_1^\infty$ be any sequence in $\mathbb{N}$ with $\lim_{n\to\infty} \omega_n = \infty$. Then there exists a sequence $(\delta_n)_1^\infty$ with $\lim_{n\to\infty} \delta_n = 0$, such that for all $n \in \mathbb{N}$, if $(x_i)_1^n$ is an i.i.d. sample from μ with corresponding order statistics $(x_{(i)})_1^n$, then with probability at least $1 - \delta_n$,

$$\sup_{\omega_n \le i \le n-\omega_n} |x_{(i)} - x^*_{(i)}| \le \delta_n$$

where $x^*_{(i)} = F^{-1}(i/(n+1))$.

Examples of probability distributions that satisfy the conditions of theorem 28 include the normal distribution, the Weibull distribution with shape parameter $c > 1$ and the chi distribution (including the Rayleigh and Maxwell distributions). Examples of distributions that satisfy the conditions of theorem 29 include the exponential distribution, the chi-squared distribution (and more generally the gamma distribution), the Weibull distribution with shape parameter $c = 1$, the Laplace distribution, the logistic distribution, and the Gumbel distribution. Note that in theorem 29 we can take $(\omega_n)_1^\infty$ to grow arbitrarily slowly, for example let $\omega_n = \log\log\log n$. We thus have tight control over almost the entire data set with the exception of a *very* small proportion of points. This is substantially better than the $\sqrt{n}$ 'visibility' of Donsker's theorem. To account for the few data points that escape theorem 29 we provide the following result.

Theorem 30. *Let μ be any probability measure on $\mathbb{R}$ that obeys the conditions of theorem 29. Then there exists $k > 0$ such that for all $T > 10^6$ and all $n \in \mathbb{N}$, if $(x_i)_1^n$ is an i.i.d. sample from μ with corresponding order statistics $(x_{(i)})_1^n$, then with probability at least $1 - 400T^{-1/2}$,*

$$\sup_{1 \leq i \leq n} |x_{(i)} - x^*_{(i)}| \leq kT$$

A probability measure μ is called p-log-concave for some $p \in [1, \infty)$ if it has a density function of the form $f(x) = c\exp(-g(x)^p)$ where g is non-negative and convex. The 1-log-concave distributions are simply referred to as log-concave. If μ is p-log-concave then it is also q-log-concave for all $1 \leq q \leq p$. Many of the probability distributions in statistics (including most of those listed above, see [3]) are log-concave. Log-concave probability measures are of great interest in pure mathematics, especially in the study of convex bodies (see for example [43] and [44]). In economics, reliability theory and several other fields, log-concave distributions play a very natural role, as log-concavity is intimately connected to monotonicity of the mean-residual-lifetime function (see section 7 in [3] for more details and an extensive list of references) . The generalized normal distributions with density functions

$$f_p(x) = \frac{p}{2\Gamma(p^{-1})} e^{-|x|^p}$$

are the primary examples of p-log-concave distributions.

Theorem 31. *Let $p > 1$, $q > 0$ and let μ be a p-log-concave probability measure on $\mathbb{R}$ with cumulative distribution function F. Then there exists $c > 0$ such that for any $n \in \mathbb{N}$ and any i.i.d. sample $(x_i)_1^n$ from μ with order statistics $(x_{(i)})_1^n$, with probability* *at least* $1 - c(\log n)^{-q}$,

$$\sup_{1\leq i\leq n} |x_{(i)} - x^*_{(i)}| \leq c\frac{\log\log n}{(\log n)^{1-1/p}}$$

where $x^*_{(i)} = F^{-1}(i/(n+1))$.

The main idea behind the proof of these theorems is to first analyze the uniform distribution on $[0,1]$. We do this using a powerful representation of the empirical point process via independent random variables that allows us to use classical results such as the law of large numbers (in the form of Chebyshev's inequality) and the law of the iterated logarithm. A key step in this analysis is to exploit the inherent regularity of order statistics which allows for control over all points based on an inspection of merely $\log n$ carefully chosen points. We then transform the points under the action of F^{-1} to analyze the general case. We introduce a new class of metrics on $(0,1)$ defined by

$$\theta_p(x,y) = \max\left\{\frac{\log(x^{-1}y)}{(\log x^{-1})^{1-1/p}}, \frac{\log((1-y)^{-1}(1-x))}{(\log(1-y)^{-1})^{1-1/p}}\right\} \tag{2.0.5}$$

for $1 \leq p < \infty$ and $0 < x \leq y < 1$. To see that each θ_p is indeed a metric, note that $\theta_p(x,y)$ is decreasing in x and increasing in y throughout the triangular region $\{(x,y) \in (0,1)^2 : x < y\}$. We show that F^{-1} is either Lipschitz or uniformly continuous with respect to these metrics (depending on the assumptions imposed on μ). After this, our main results become straightforward to prove.

There are endless variations on the main theme of this chapter. Our intention is simply to highlight a phenomenon and introduce methods by which to study it. Note that our results are purely asymptotic in nature and we can (and do) assume throughout the chapter that $n > n_0$ for some $n_0 \in \mathbb{N}$.

2.1 The uniform distribution

Let $(\gamma_i)_1^n$ denote an i.i.d. sample from the uniform distribution on $[0,1]$ with corresponding order statistics $(\gamma_{(i)})_1^n$ and let $(z_i)_1^{n+1}$ be an i.i.d. sequence of random variables that follow the standard exponential distribution. For $1 \le i \le n$ define

$$y_i = \left(\sum_{j=1}^{i} z_j\right)\left(\sum_{j=1}^{n+1} z_j\right)^{-1}$$

It is of great interest to us that $(y_i)_1^n$ and $(\gamma_{(i)})_1^n$ have the same distribution in $\mathbb{R}^n$ (see chapter 5 in [17]). This is nothing but an expression of the fact that the empirical point process locally resembles the Poisson point process. Also of interest is the fact that these random vectors have the same distribution as the partial sums of a random vector uniformly distributed (with respect to Lebesgue measure) in the standard simplex $\Delta^n = \{w \in \mathbb{R}^{n+1} : w_i \ge 0 \ \forall i, \ \sum_i w_i = 1\}$. The power of this representation is that we have an expression for $(\gamma_{(i)})_1^n$ in terms of independent random variables. Note that

$$y_i = \frac{i}{n+1}\left(\frac{1}{i}\sum_{j=1}^{i} z_j\right)\left(\frac{1}{n+1}\sum_{j=1}^{n+1} z_j\right)^{-1} \tag{2.1.1}$$

Both lemma 32 and lemma 34 below can be compared to the results in [73].

Lemma 32. *Let $T > 10^6$ and $n \in \mathbb{N}$. With probability at least $1 - 400T^{-1/2}$ the following inequalities hold simultaneously for all $1 \le i \le n$,*

$$T^{-1} \le \gamma_{(i)} \left(\frac{i}{n+1}\right)^{-1} \le T \tag{2.1.2}$$

$$T^{-1} \le (1-\gamma_{(i)}) \left(1 - \frac{i}{n+1}\right)^{-1} \le T \tag{2.1.3}$$

Proof. Let $Q = 2^{-1}T^{1/2}$ and momentarily fix $1 \leq i \leq n+1$. The random variable $i^{-1}\sum_{j=1}^{i} z_j$ has mean 1 and variance i^{-1}. Using Chebyshev's inequality, with probability at least $1 - i^{-1}Q^{-2}$ we have

$$-Q < 1 - \frac{1}{i}\sum_{j=1}^{i} z_j < Q$$

The random variable

$$U_i = |\{j \in \mathbb{N} : j \leq i, z_j \leq 2Q^{-1}\}|$$

follows a binomial distribution with i trials and success probability $1-\exp(-2Q^{-1}) \leq 2Q^{-1}$. Using Chebyshev's inequality again, with probability at least $1 - 32i^{-1}Q^{-1}$ we have $U_i < i/2$, which implies that $i^{-1}\sum_{j=1}^{i} z_j > Q^{-1}$. Hence, with probability at least $1 - 33i^{-1}Q^{-1}$ we have

$$Q^{-1} < \frac{1}{i}\sum_{j=1}^{i} z_j < Q + 1 \tag{2.1.4}$$

Let $M = \lfloor \log_2(n) \rfloor$. With probability at least $1 - 33Q^{-1}\sum_{j=0}^{M} 2^{-j} - 33(n+1)^{-1}Q^{-1} \geq 1 - 100Q^{-1}$equation (2.1.4) holds simultaneously for $i = 1, 2, 2^2, 2^3 \ldots 2^M$ and for $i = n+1$. Hence, by (2.1.1), with probability at least $1 - 100Q^{-1}$ we have that for all such i

$$\frac{1}{2}Q^{-2}\frac{i}{n+1} \leq y_i \leq 2Q^2\frac{i}{n+1}$$

Since $(y_i)_1^n$ is an increasing sequence, control over the values $(y_{2^j})_{j=1}^{M}$ leads to control over the entire sequence and, recalling the representation of $(\gamma_{(i)})_1^n$ in terms of $(y_i)_1^n$, the bound (2.1.2) follows for all $1 \leq i \leq n$. The bound (2.1.3) then follows by symmetry. $\square$

Lemma 33. *Let* $t \in (0,1)$ *and* $n \in \mathbb{N}$. *With probability at least* $1 - 2\exp(-nt^2/5)$ *the following inequality holds simultaneously for all* $1 \leq i \leq n$,

$$\left|\gamma_{(i)} - \frac{i}{n+1}\right| \leq t \tag{2.1.5}$$

Proof. We can assume without loss of generality that $n^{-1} \leq 2t/3$ (otherwise the probability bound becomes trivial). Note that since our sample is taken from the uniform distribution we have

$$\begin{aligned} \sup_{1\leq i\leq n} |\gamma_{(i)} - i(n+1)^{-1}| &\leq n^{-1} + \sup_{1\leq i\leq n} |\gamma_{(i)} - in^{-1}| \\ &= n^{-1} + \sup_{0\leq t\leq 1} |F_n(t) - F(t)| \end{aligned}$$

where $F(t) = t$ is the cumulative distribution function and F_n is the empirical distribution function. By the Dvoretzky-Kiefer-Wolfowitz inequality (as mentioned in the introduction), with probability at least $1 - 2\exp(-5^{-1}nt^2)$ we have

$$\sup_{0\leq t\leq 1} |F_n(t) - F(t)| \leq t/3$$

and the result follows. $\square$

Note that in the preceding proof one can also use Doob's martingale inequality (in the form of Kolmogorov's inequality) and the representation of $(\gamma_{(i)})_1^n$ in terms of $(y_n)_1^n$, although this approach yields an inferior probability bound.

Lemma 34. *Let* $(\omega_n)_1^\infty$ *be any sequence in* $\mathbb{N}$ *such that* $\lim_{n\to\infty} \omega_n = \infty$. *Then for all* $T > 1$ *and all* $\delta \in (0,1)$ *there exists* $n_0 \in \mathbb{N}$ *such that for all* $n > n_0$, *if* $(\gamma_{(i)})_1^n$ *are the order statistics from an i.i.d. sample from the uniform distribution on* $[0,1]$, *then with probability at least* $1 - \delta$, *(2.1.2) and (2.1.3) hold for all* $\omega_n \leq i \leq n - \omega_n$.

Proof. We use the representation (2.1.1). Let $T > 1$ and $\delta \in (0,1)$ be given. Without loss of generality we may assume that $T \leq 2$. Let $(\widetilde{z}_i)_1^\infty$ denote any i.i.d. sequence of random variables that follow the standard exponential distribution. Define the deterministic sequence $(\lambda_j)_1^\infty$ as follows,

$$\lambda_j = \mathbb{P}\{\sup_{i\geq j}(2i \log\log i)^{-1/2} \left|\sum_{k=1}^{i}(\widetilde{z}_k - 1)\right| \leq 2\}$$

Note that $(\lambda_j)_1^\infty$ is an increasing sequence and by the law of the iterated logarithm, $\lim_{j\to\infty} \lambda_j = 1$. Fix $n_0 \in \mathbb{N}$ with $n_0 \geq 64\delta^{-1}(T^{1/2}-1)^{-2}$ such that for all $n > n_0$ we have the following inequalities,

$$\begin{aligned} \lambda_{\omega(n)} &\geq 1 - \delta/4 \\ \left(\frac{8\log\log\omega_n}{\omega_n}\right)^{1/2} &\leq T^{1/2} - 1 \end{aligned}$$

Now consider any $n > n_0$ and let $(\gamma_{(i)})_1^n$ denote the order statistics mentioned in the statement of the lemma. With probability at least $1 - \delta/4$, for all $\omega(n) \leq i \leq n$,

$$\begin{aligned} \left|1 - \frac{1}{i}\sum_{j=1}^{i} z_j\right| &\leq \left(\frac{8\log\log\omega_n}{\omega_n}\right)^{1/2} \\ &\leq T^{1/2} - 1 \end{aligned}$$

By Chebyshev's inequality and the fact that the function $u \mapsto u^{-1}$ is 4-Lipschitz on $[1/2, \infty)$, with probability at least $1 - 16n^{-1}(T^{1/2}-1)^{-2} \geq 1 - \delta/4$

$$\left|1 - \left(\frac{1}{n+1}\sum_{j=1}^{n+1} z_j\right)^{-1}\right| < T^{1/2} - 1$$

By (2.1.1), with probability at least $1 - \delta/2$, (2.1.2) holds for all $\omega(n) \leq i \leq n$. By symmetry, with the same probability (2.1.3) holds for all $1 \leq i \leq n - \omega(n)$. The lemma is thus proven. □

2.2 The general case

We now consider the general setting of a probability measure μ with connected support in $\mathbb{R}$ and cumulative distribution function F. The generalized inverse F^{-1} is continuous, non-decreasing and satisfies the equation

$$F^{-1}(F(t)) = t$$

for all $t \in \mathbb{R}$.

Lemma 35. *Let μ be a probability measure on $\mathbb{R}$ with connected support and cumulative distribution function F that satisfies (2.0.1). Then for all $T > 1$ and all $\delta > 0$ there exists $\eta \in (0,1)$ such that for all $x, y \in (0, \eta)$ with $T^{-1} \leq xy^{-1} \leq T$ and all $x, y \in (1-\eta, 1)$ with $T^{-1} \leq (1-x)(1-y)^{-1} \leq T$ we have $|F^{-1}(x) - F^{-1}(y)| \leq \delta$.*

Proof. Consider any $T > 1$ and $\delta > 0$. By (2.0.1) there exists $t_0 \in \mathbb{R}$ such that for all $t \leq t_0$, $TF(t) < F(t+\delta)$. Let $\eta_1 = F(t_0)$. Consider any $x, y \in (0, \eta_1)$ such that $T^{-1} \leq xy^{-1} \leq T$. Without loss of generality, $x < y$. Let $s = F^{-1}(x)$ and $t = F^{-1}(y)$. Then $s \leq t_0$, hence $y \leq Tx = TF(s) < F(s+\delta)$, from which it follows by applying F^{-1} that $t \leq s + \delta$ and that $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. Analysis of the right hand tail is identical and provides us with $\eta_2 > 0$ such that for all $x, y \in (1-\eta_2, 1)$ with $T^{-1} \leq (1-x)(1-y)^{-1} \leq T$ we have $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. The result follows with $\eta = \min\{\eta_1, \eta_2\}$. $\square$

Lemma 36. *Let μ be a probability measure on $\mathbb{R}$ with connected support and cumulative distribution function F that satisfies both (2.0.3) and (2.0.4). Then for all $\delta > 0$ there exists $T > 1$ such that for all $x, y \in (0,1)$ with $T^{-1} \leq xy^{-1} \leq T$ and*

$T^{-1} \leq (1-x)(1-y)^{-1} \leq T$ *we have* $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. *In particular,* F^{-1} *is uniformly continuous with respect to the metric* θ_1 *(see (2.0.5)).*

Proof. Consider any $\delta > 0$. By (2.0.4) there exists $T_1 > 1$ and $t_0 \in \mathbb{R}$ such that for all $t < t_0$, $T_1 F(t) \leq F(t+\delta)$. Let $\eta_1 = \min\{F(t_0), 2^{-1}\}$. As in the proof of the previous lemma, it follows that for all $x, y \in (0, \eta_1)$ with $T_1^{-1} \leq xy^{-1} \leq T_1$ we have $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. Similarly (using (2.0.3)), there exists $T_2 > 1$ and $\eta_2 \in (2^{-1}, 1)$ such that for all $x, y \in (\eta_2, 1)$ with $T_2^{-1} \leq (1-x)(1-y)^{-1} \leq T_2$ we have $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. By continuity of F^{-1} relative to the standard topology on $(0,1)$, and by compactness of $[2^{-1}\eta_1, 1 - 2^{-1}\eta_2]$ there exists $0 < \delta' < 10^{-1}\min\{\eta_1, \eta_2\}$ such that for all $x, y \in [2^{-1}\eta_1, 1-2^{-1}\eta_2]$ with $|x-y| < \delta'$ we have $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. We leave it to the reader to verify that the result holds with

$$T = \min\{T_1, T_2, 1 + \delta'\}$$

□

Proof of theorem 28. We shall construct a function h that takes an arbitrary $\delta \in (0,1)$ and produces an appropriate $n_0 = h(\delta) \in \mathbb{N}$. Then, using this function we shall define the desired sequence $(\delta_n)_1^\infty$ that is mentioned in the statement of the theorem. To this end, let $\delta \in (0,1)$ be given. Define

$$T = 10^6 \delta^{-2} \tag{2.2.1}$$

By lemma 35 there exists $\eta \in (0,1)$ such that if $x, y \in (0, \eta)$ and $T^{-1} \leq xy^{-1} \leq T$, or $x, y \in (1-\eta, 1)$ and $T^{-1} \leq (1-x)(1-y)^{-1} \leq T$, then $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. By compactness, F^{-1} is uniformly continuous on $[\eta/2, 1-\eta/2]$, which implies the existence of $t \in (0, \eta/2)$ such that if $x, y \in [\eta/2, 1 - \eta/2]$ and $|x - y| \leq t$, then $|F^{-1}(x) - F^{-1}(y)| \leq \delta$. Define

$$n_0 = \left\lceil 5t^{-2} \log(4\delta^{-1}) \right\rceil \tag{2.2.2}$$

and consider any $n \geq n_0$. Let $(\gamma_{(i)})_1^n$ denote the order statistics corresponding to an i.i.d. sample from the uniform distribution on $[0, 1]$. Note that we have the representation

$$x_{(i)} = F^{-1}(\gamma_{(i)}) \tag{2.2.3}$$

valid for all $1 \leq i \leq n$. By lemmas 32 and 33, as well as equations (2.2.1) and (2.2.2), with probability at least $1 - \delta$ inequalities (2.1.2), (2.1.3) and (2.1.5) hold simultaneously for all $1 \leq i \leq n$. Suppose that these inequalities do indeed hold and consider any fixed $1 \leq i \leq n$. Since $t \leq \eta/2$, one of the three sets $[0, \eta]$, $[\eta/2, 1 - \eta/2]$ and $[1 - \eta, 1]$ contains both $\gamma_{(i)}$ and $i(n+1)^{-1}$, which implies that $|F^{-1}(\gamma_{(i)}) - F^{-1}(i(n+1)^{-1})| \leq \delta$, which is inequality (2.0.2).

Define the non-decreasing sequence $(\kappa_n)_1^\infty$ by $\kappa_n = \max\{h(e^{-i}) : 1 \leq i \leq n\}$ and set

$$\delta_n = \exp(-\max\{i \in \mathbb{N} : \kappa_i \leq n\})$$

where we define $\max \emptyset = 0$. It is clear that $\lim_{n \to \infty} \delta_n = 0$. Consider any fixed $n \in \mathbb{N}$. If $\{i \in \mathbb{N} : \kappa_i \leq n\} = \emptyset$ then the probability bound is trivial, otherwise let $j = \max\{i \in \mathbb{N} : \kappa_i \leq n\}$. The result follows by the inequality $h(\delta_n) = h(e^{-j}) \leq \kappa_j \leq n$ and by definition of the function h. □

Proof of theorems 29 and 30. The proof is very similar to that of theorem 28. We use the representation (2.2.3). The main difference is that we use lemmas 34 and 36 instead of lemmas 32 and 35. The details are left to the reader. □

2.3 Log-concave distributions

The following two lemmas are modifications of lemmas 6 and 9 in [25].

Lemma 37. *Let μ be a log-concave probability measure on $\mathbb{R}$ with cumulative distribution function F. Then there exists $c > 0$ such that for all $0 < x < y < 1$,*

$$|F^{-1}(y) - F^{-1}(x)| \leq c \max\left\{ \left|F^{-1}(y)\right| \frac{\log(x^{-1}y)}{\log y^{-1}}, \left|F^{-1}(x)\right| \frac{\log((1-x)/(1-y))}{\log(1-x)^{-1}} \right\} \tag{2.3.1}$$

Proof. By theorem 5.1 in [51] (see lemma 5 in [25] for a proof) F is log-concave. Hence the function $u(t) = -\log F(t)$ is convex (and strictly decreasing). Let $\mathbb{E}\mu$ denote the centroid of μ (the expected value of a random variable with distribution μ). By lemma 5.12 in [51] (see also lemma 3.3 in [12]) $F(\mathbb{E}\mu) \geq e^{-1}$, hence $u(\mathbb{E}\mu) \leq 1$. By convexity of u we have the inequality $(t-s)^{-1}(u(t)-u(s)) \leq (\mathbb{E}\mu - t)^{-1}(u(\mathbb{E}\mu)-u(t))$, which is valid for all $s < t < \mathbb{E}\mu$. Let $0 < x < y < \min\{e^{-2}, F(0), F(-2\mathbb{E}\mu)\}$ and define $s = F^{-1}(x)$ and $t = F^{-1}(y)$. Then we have

$$F^{-1}(y) - F^{-1}(x) \leq (\mathbb{E}\mu - F^{-1}(y)) \frac{\log(x^{-1}y)}{\log y^{-1} - u(\mathbb{E}\mu)}$$

It follows from the restrictions on y that $F^{-1}(y) < 0$ and that $|F^{-1}(y)| \geq 2\,|\mathbb{E}\mu|$. Since $y < F(\mathbb{E}\mu)^2$, it follows that $\log y^{-1} > 2u(\mathbb{E}\mu)$ and (2.3.1) follows for such x and y with $c = 4$. For other values of x and y, inequality (2.3.1) follows by compactness, continuity and symmetry. □

Lemma 38. *Let $p \geq 1$ and let μ be a p-log-concave probability measure on $\mathbb{R}$ with cumulative distribution function F. Then there exists $c > 0$ such that for all $x \in (0,1)$,*

$$|F^{-1}(x)| \leq c \max\{(\log x^{-1})^{1/p}, (\log(1-x)^{-1})^{1/p}\} \tag{2.3.2}$$

As a consequence of (2.3.2) and (2.3.1), F^{-1} is Lipschitz with respect to the metric θ_p (see (2.0.5)).

Proof. By lemma 9 in [25] (which holds for $p \geq 1$) there exists $c_1, c_2 > 0$ and $t_0 > 1$ such that for all $t < -t_0$, $F(t) \leq c_1|t|^{1-p}\exp(-c_2|t|^p)$. Let $\eta_1 = \min\{F(-t_0), c_1^{-1}\}$ and consider any $x \in (0, \eta_1)$. Let $t = F^{-1}(x)$. Hence $x = F(t) \leq c_1|t|^{1-p}\exp(-c_2|t|^p)$, which implies that

$$\begin{aligned} |F^{-1}(x)| &= -t \\ &\leq (c_2^{-1}(\log c_1 + \log x^{-1}))^{1/p} \\ &\leq 2^{1/p} c_2^{-1/p} (\log x^{-1})^{1/p} \end{aligned}$$

The result now follows by symmetry, compactness and continuity. □

Lemma 39. *Let F be a cumulative distribution function associated to a log-concave probability measure. Then there exists $c > 0$ such that for all $\varepsilon \in (0, 1/2)$ and all $x, y \in [\varepsilon, 1-\varepsilon]$,*

$$|F^{-1}(x) - F^{-1}(y)| \leq c\varepsilon^{-1}|x - y|$$

Proof. This follows from lemmas 37 and 38 with $p = 1$ and the inequality $\log t \leq t - 1$. □

Proof of theorem 31. By lemmas 32, 37 and 38, with probability at least $1-400(\log n)^{-q}$, for all $i \leq n^{3/4}$ and all $i \geq n - n^{3/4}$ we have

$$|x_{(i)} - x^*_{(i)}| \leq c\frac{\log\log n}{(\log n)^{1-1/p}}$$

Let $I = [2^{-1}n^{-1/4}, 1 - 2^{-1}n^{-1/4}]$. By lemma 39, for all $x, y \in I$ we have

$$|F^{-1}(x) - F^{-1}(y)| \leq cn^{1/4}|x - y|$$

By lemma 33, with probability at least $1 - 2\exp(-5n^{1/4})$, for all $1 \leq i \leq n$ we have

$$|\gamma_{(i)} - i(n+1)^{-1}| \leq n^{-3/8}$$

Hence for all $n^{3/4} \leq i \leq n - n^{3/4}$ both $\gamma_{(i)}$ and $i(n+1)^{-1}$ are elements of I and the result follows. □

Chapter 3

Concentration inequalities for Lipschitz functions into an arbitrary metric space

Typical concentration results include Lévy's concentration of Lipschitz functions with respect to Haar measure on high dimensional spheres and Talagrand's concentration of convex Lipschitz functions with respect to product measures on high dimensional cubes. These results assume that the range of the Lipschitz function is contained in $\mathbb{R}$ and break down when the range is allowed to be high dimensional. The identity map provides the simplest such example. As long as the dimension of the range is much less than that of the domain, however, one can reclaim a similar concentration inequality.

Proposition 40. *Let $n \in \mathbb{N}$, let $(X, ||\cdot||_K)$ denote any Banach space of dimension $N < \infty$ and let $f : S^{n-1} \to X$ be a Lipschitz function. Let x and y be independent random vectors uniformly distributed on S^{n-1}. Then for all $r > 0$, with probability at least $1 - C\exp(c_1 N - c_2 r^2 n)$,*

$$||f(x) - f(y)||_K < r Lip(f) \tag{3.0.1}$$

where $C, c_1, c_2 > 0$ are universal constants.

The proof of proposition 40 is routine and can also be applied to the case when X is a metric space that embeds well into ℓ_∞^M for $M = e^{o(n)}$. The study of concentration of functions into more general spaces was, to our knowledge, introduced by Gromov [37] [38] [39] and has also been studied by Funano [30] [31] [32].

In all of the existing theory, the structure of the range plays a fundamental role. We are not aware of any existing results for functions taking values in a truly arbitrary metric space. As far as Talagrand's inequality is concerned, where convexity of f is required, it is not entirely clear even what the statement of such an extension would be.

In this chapter we show that if the Lipschitz function f is invariant under coordinate permutations, then it obeys concentration inequalities that are independent of the range. We are particularly interested in the case when the range is very complicated, such as an infinite dimensional space, and the existing theory breaks down completely.

3.1 Main results

In the spirit of Lévy's inequality we have the following result,

Theorem 41. *Let* (Ω, ρ) *denote any metric space and let* $f : S^{n-1} \to \Omega$ *be a Lipschitz function that is invariant under coordinate permutations. Let* x *and* y *be independent random vectors uniformly distributed on* S^{n-1}*. Then with probability at least* $1 - C(\log n)^{-1000}$,

$$\rho(f(x), f(y)) \leq \frac{c \log \log n}{\sqrt{\log n}} Lip(f) \tag{3.1.1}$$

where $C, c > 0$ *are universal constants and* $Lip(f)$ *is taken with respect to either the*

Euclidean or the geodesic distance on S^{n-1}*.*

We expect to improve this result to a bound of the form $\rho(f(x), f(y)) \leq cn^{-\alpha} Lip(f)$ for some $\alpha > 0$, perhaps $\alpha = 1/4$.

For each $x \in \mathbb{R}^n$ let $Tx = (x_{(i)})_{i=1}^n$ denote the non-decreasing rearrangement of the coordinates of x, also known as the order statistics of x. The function $T : \ell_2^n \to \ell_2^n$ is rearrangement invariant as well as Lipschitz, with $Lip(T) = 1$. If $f : \mathbb{R}^n \to \Omega$ is any rearrangement invariant function , then $f = f \circ T$. Hence theorem 41 is equivalent to the following lemma, which claims that most points on the unit sphere have essentially the same coordinates, just in a different order.

Lemma 42. *Let* x *and* y *be independent random vectors uniformly distributed on* S^{n-1}*. Then with probability at least* $1 - C(\log n)^{-1000}$*,*

$$||Tx - Ty||_2 \leq \frac{c \log\log n}{\sqrt{\log n}}$$

where $C, c > 0$ *are universal constants.*

These results can easily be generalized as follows:

Theorem 43. *Let* $1 < p < \infty$ *and* $1 \leq q \leq \infty$*. Let* (Ω, ρ) *denote any metric space and let* $f : B_p^n \to \Omega$ *be a Lipschitz function that is invariant under coordinate permutations. Let* $Lip_q(f)$ *denote the Lipschitz constant of* f *with respect to the* ℓ_q *metric. Let* x *and* y *be independent random vectors uniformly distributed inside* B_p^n*. Then with probability at least* $1 - C_p(\log n)^{-1000}$*,*

$$\rho(f(x), f(y)) \leq c_p n^{\frac{1}{q} - \frac{1}{p}} \frac{\log\log n}{(\log n)^{1-\frac{1}{p}}} Lip_q(f)$$

where $C_p, c_p > 0$ are constants that depend on p but not on n, q or (Ω, ρ). The same result holds when $f : \partial B_p^n \to \Omega$ and x and y are distributed according to the cone measure on ∂B_p^n.

The final result is a variation of the Dvoretzky-Kiefer-Wolfowitz inequality from empirical process theory that we write in a form which mirrors Talagrand's inequality.

Proposition 44. *Let (Ω, ρ) denote any metric space and let $f : [0,1]^n \to \Omega$ be a Lipschitz function that is invariant under coordinate permutations. Let x and y be independent random vectors uniformly distributed in $[0,1]^n$. Then for all $\lambda > 0$, with probability at least $1 - 4\exp(-2\lambda^2)$ we have,*

$$\rho(f(x), f(y)) \leq \lambda Lip(f)$$

where $Lip(f)$ is defined relative to the Euclidean norm on $[0,1]^n$.

Note that the bounds are independent of n and that with probability $1-o(1)$, $||x-y||_2 > c\sqrt{n}$. As before, the prototypical example is the function $Tx = (x_{(i)})_{i=1}^n$. The diameter of the unit cube $[0,1]^n$ is $\sqrt{n}$, however if we ignore the order of coordinates by allowing appropriate permutations to act on points, most of the cube is contained in a set of bounded diameter (bounded with respect to n).

3.2 Proofs

Proof of proposition 40. A standard result is that there exists a linear embedding $T : X \hookrightarrow \ell_\infty^M$ with $M = 3^N$ such that for all $x \in X$

$$||x||_K \leq ||Tx||_\infty \leq 2||x||_K$$

To see this, consider any 1-net $\mathcal{N} \subset \partial K$. The standard bound yields $|\mathcal{N}| \leq 3^N$. By the Hahn-Banach theorem, for each $y \in \mathcal{N}$ there exists a functional $y^* \in \partial B_{X^*}$ such that $y^*(y) = 1$. The embedding is then given by $x \mapsto 2(y^*(x))_{y \in \mathcal{N}}$. Note that $Lip(y^* \circ f) \leq Lip(f)$. The result follows from Lévy's concentration inequality, the union bound and the formula

$$||f(\theta) - f(\omega)||_K \leq ||Tf(\theta) - Tf(\omega)||_\infty = 2 \sup_{y \in \mathcal{N}} |y^*(f(\theta)) - y^*(f(\omega))|$$

□

Lemma 45. *Let θ and θ' be independent vectors uniformly distributed on $\sqrt{n}S^{n-1}$. Let $T\theta = (\theta_{(i)})_1^n$ and $T\theta' = (\theta'_{(i)})_1^n$ be the corresponding order statistics. Then with probability at least* $1 - c(\log n)^{-1000}$, $||T\theta - T\theta'||_\infty \leq c(\log \log n)/\sqrt{\log n}$.

Proof. We can simulate $\theta_i = \left(\frac{1}{n}\sum_{j=1}^n \gamma_j^2\right)^{-1/2} \gamma_i$ and $\theta'_i = \left(\frac{1}{n}\sum_{j=1}^n \widehat{\gamma}_j^2\right)^{-1/2} \widehat{\gamma}_i$ where $\gamma = (\gamma_i)_1^n$ and $\widehat{\gamma} = (\widehat{\gamma}_i)_1^n$ are independent i.i.d. $N(0,1)$ samples. By theorem 31 in chapter 3 (see also [27]), with probability at least $1 - c(\log n)^{-1000}$, $||\gamma - \widehat{\gamma}||_\infty < c(\log \log n)/\sqrt{\log n}$. The coefficients $\frac{1}{n}\sum_{j=1}^n \gamma_j^2$ and $\frac{1}{n}\sum_{j=1}^n \widehat{\gamma}_j^2$ have expected value 1 and variance cn^{-1}. By Chebyshev's inequality, with probability at least $1 - cn^{-1/2}$, we have $1 - n^{-1/4} < \frac{1}{n}\sum_{j=1}^n \gamma_j^2 < 1 + n^{-1/4}$. By theorem 31 again and the fact that $\Phi^{-1}(1 - c_1/n) < c_2(\log n)^{1/2}$, with probability at least $1 - c(\log n)^{-1000}$, $||\gamma||_\infty < c(\log n)^{1/2}$. Hence, $\max\{||\gamma - X||_\infty, ||\widehat{\gamma} - Y||_\infty\} < c(\log n)^{1/2} n^{-1/4}$. The result follows by the triangle inequality. □

Proof of lemma 42. Take $x = n^{-1/2}\theta$ and $y = n^{-1/2}\theta'$. Then $||Tx - Ty||_2 = n^{-1/2}||T\theta - T\theta'||_2 \leq ||T\theta - T\theta'||_\infty$. The result then follows from lemma 45. □

Proof of theorem 41. This follows from lemma 42 and the inequality $f(x), f(y)) \leq \rho(f(Tx), f(Ty)) \leq Lip(f)||Tx - Ty||_2$. $\square$

Proof of theorem 43. By theorem 1 in [11], we can simulate x and y as

$$\begin{aligned} x_i &= \frac{\gamma_i}{\left(\sum_{j=1}^n |\gamma_j|^p + W_1\right)^{1/p}} = \frac{\gamma_i}{n^{1/p}\left(\frac{1}{n}\sum_{j=1}^n |\gamma_j|^p + \frac{1}{n}W_1\right)^{1/p}} \\ y_i &= \frac{\widehat{\gamma}_i}{\left(\sum_{j=1}^n |\widehat{\gamma}_i|^p + W_2\right)^{1/p}} = \frac{\widehat{\gamma}_i}{n^{1/p}\left(\frac{1}{n}\sum_{j=1}^n |\widehat{\gamma}_i|^p + \frac{1}{n}W_2\right)^{1/p}} \end{aligned}$$

where $\gamma = (\gamma_i)_1^n$ and $\widehat{\gamma} = (\widehat{\gamma}_i)_1^n$ are independent random vectors each with a density given by $f(x) = c_p^n \exp(-||x||_p^p)$, while W_1 and W_2 each have the standard exponential distribution and are also independent of x and y, and each other. By theorem 31 in chapter 3 (see also [27]), with probability at least $1 - c(\log n)^{-1000}$, $||\gamma - \widehat{\gamma}||_\infty < c(\log\log n)(\log n)^{-1+\frac{1}{p}}$. The coefficients $\frac{1}{n}\sum_{j=1}^n |\gamma_j|^p + \frac{1}{n}W_1$ and $\frac{1}{n}\sum_{j=1}^n |\widehat{\gamma}_i|^p + \frac{1}{n}W_2$ have expected value $c_p + n^{-1}$ and variance $c_p' n^{-1} + n^{-2}$. By Chebyshev's inequality, with probability at least $1 - c_p n^{-1/2}$ we have

$$c_p + n^{-1} - n^{-1/4} \leq \frac{1}{n}\sum_{j=1}^n |\gamma_j|^p + \frac{1}{n}W_1 \leq c_p + n^{-1} + n^{-1/4}$$

Let F be the cumulative distribution of the coordinates of x and y,

$$F(t) = \int_{-\infty}^t c_p e^{-|u|^p}\, du$$

From chapter 1 we know that $F^{-1}(1 - c_p n^{-1}) \leq c_p'(\log n)^{1/p}$, and combining this and theorem 31, with probability at least $1 - c_p(\log n)^{-1000}$ we have $||\gamma||_\infty < c_p'(\log n)^{1/p}$. Hence, $\max\{||\gamma - n^{1/p}x||_\infty, ||\widehat{\gamma} - n^{1/p}y||_\infty\} < c_p(\log n)^{1/p} n^{-1/4}$. By the triangle inequality, $||n^{1/p}x - n^{1/p}y||_\infty \leq c(\log\log n)(\log n)^{-1+\frac{1}{p}}$ and the result follows from the inequality $||\cdot||_q \leq n^{1/q}||\cdot||_\infty$. When x and y each have the cone measure on ∂B_p^n, we use the same representation and the same proof, except without the variables W_1 and W_2. $\square$

Proof of proposition 44. With probability 1, all coordinates of x and y are distinct. Consider $F_x(t) = n^{-1}|\{i : x_i \leq t\}|$ and $F_y(t) = n^{-1}|\{i : y_i \leq t\}|$. The Dvoretzky-Kiefer-Wolfowitz inequality states that with probability at least $1 - 2\exp(-2\lambda^2)$, $\sup_{0\leq t\leq 1} |F_n(t) - t| \leq \lambda n^{-1/2}$, and likewise for F_y. Thus for all i,

$$\begin{aligned} |\frac{i}{n} - x_{(i)}| &= |F(x_{(i)}) - x_{(i)}| \leq \lambda n^{-1/2} \\ |\frac{i}{n} - y_{(i)}| &= |F(y_{(i)}) - y_{(i)}| \leq \lambda n^{-1/2} \end{aligned}$$

By the triangle inequality, $||Tx - Ty||_\infty \leq \lambda n^{-1/2}$ (with probability at least $1 - 4\exp(-2\lambda^2)$). The result then follows as before from the expression $f = f \circ T$ and the inequality $|| \cdot ||_2 \leq n^{1/2} || \cdot ||_\infty$. $\square$

Chapter 4

A non-asymptotic central limit theorem

Let $n \in \mathbb{N}$ and let $X = (X_i)_1^n$ be an i.i.d. sequence of random variables each with cumulative distribution function F such that $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = 1$. The distribution of X is thus an isotropic product measure μ. The condition that μ be isotropic is a mild one and entails that X has mean zero and identity covariance matrix. On the other hand, independence is a very strong condition. Define

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

Classical central limit theory dictates that for large values of n the distribution of Z approximates the standard normal distribution. Under the identification $t \leftrightarrow t\theta$, where $t \in \mathbb{R}$ and $\theta = (1/\sqrt{n}, 1/\sqrt{n} \ldots 1/\sqrt{n})$, the random variable Z is the orthogonal projection of X onto the one dimensional subspace spanned by θ. Many other projections of μ have this property, including projections onto subspaces of higher dimension. For $1 \leq k \leq n$, the Grassmanian manifold $G_{n,k}$ is the collection of all linear subspaces of $\mathbb{R}^n$ of dimension k. It is a compact homogeneous space under the action of the group O_n of orthogonal matrices $T : \mathbb{R}^n \to \mathbb{R}^n$ and is thus endowed with a unique rotationally invariant probability measure $\lambda_{n,k}$ called Haar measure. For each $E \in G_{n,k}$ let P_E denote the orthogonal projection of $\mathbb{R}^n$ onto E. A result of Romik [62] is that there exist $c_1, c_2, c_3 > 0$ (that depend on F but not on n) such that for all $\varepsilon > 0$ and all $k < c_1 \varepsilon^4 n$ there exists a set $\mathcal{E} \subset G_{n,k}$ with $\lambda_{n,k}(\mathcal{E}) > 1 - c_2 \varepsilon^{-1} \exp(-c_3 \varepsilon^4 n)$ such that for all $E \in \mathcal{E}$,

$$\sup_{\mathcal{H}} |\mathbb{P}\{P_E X \in \mathcal{H}\} - \Phi_E(\mathcal{H})| < \varepsilon \tag{4.0.1}$$

where the supremum is taken over all half spaces $\mathcal{H} \subset E$ and Φ_E is the standard normal measure on E. The metric defined by (4.0.1) is called the Tsirelson distance. A similar result holds for the Kolmogorov metric.

The condition of independence is not fundamental to the central limit theorem and can be replaced by a number of other regularity properties such as convexity [2] [22] [44] [45] and symmetry [55] [56]. These regularity properties all come down to a more fundamental property called the thin shell property [12] [18] [23] [24] [41] [46] [67] [71]. The random vector X has the thin shell property if the random variable $||X||_2/\mathbb{E}||X||_2$ is concentrated around the value 1. In other words, most of the mass of μ is contained in a spherical shell of thickness much smaller than its radius. By the weak law of large numbers, product measures have the thin shell property. The fact that isotropic convex bodies have the thin shell property was a profound contribution of [44]. Provided that X is isotropic and has the thin shell property, most one dimensional projections of X are approximately Gaussian. The converse is also true. An excellent exposition of these ideas as well as a detailed list of references is contained in [44], [47] and [41].

We define a section of a function f as the restriction of f to an affine subspace. This is a natural functional generalization of a section of a convex body. If $X \in \mathbb{R}^n$

is a random vector with a continuous density function f that decays rapidly to zero in all directions, then a suitable multiple of the section $f|_E$ can be thought of as the density function of X conditional on the event $\{\varphi(X) = T\}$, where φ is a linear functional, $E = \{x : \varphi(x) = T\}$ and the density is taken with respect to $n-1$ dimensional Lebesgue measure on E. Of course the usual definition of conditional distribution breaks down because the event $\{\varphi(X) = T\}$ has probability zero and the section becomes our *definition* of such a conditional distribution.

Sections and projections are natural counterparts in convex geometry (and functional analysis) and one is lead naturally to ask whether the central limit theory as described above has an analogue for sections. In this chapter we prove several such theorems, albeit in a different spirit to the central limit theory of projections. It is easily seen that central sections do not obey the central limit theorem and we shall consider sections far from the origin. A fundamental difference between the theory of sections and projections is that for sections we do not require high dimensionality.

Our results can also be interpreted in the setting of classical probability theory without reference to sections or projections. Conditioned on the event $\{\varphi(X) = T\}$, any other linear functional $\widetilde{\varphi}$ such that $null(\widetilde{\varphi}) \neq null(\varphi)$ has an approximately normal distribution.

4.1 Main results

For a function $g : \mathbb{R} \to \mathbb{R}$ such $\lim_{t\to\pm\infty} t^2 g_i''(t) = \infty$ we consider the following modulus

$$\xi_g(r,t) = \sup\left\{ \left| \frac{g'''(w+s)}{g''(w)^{3/2}} \right| : |w| \geq t, |s| \leq r g''(w)^{-1/2} \right\} \tag{4.1.1}$$

We will be interested in functions g such that for all $r > 0$,

$$\lim_{t\to\infty} \xi_g(r,t) = 0 \tag{4.1.2}$$

This is a relatively natural condition satisfied by many functions and we discuss it further in section 4.3. As an example, consider the functions $g(t) = |t|^p$ $(p > 1)$ for which we have

$$\xi_g(r,t) \le 2t^{-p/2}$$

valid for all r and t such that $r \le c_p t^{p/2}$. As another example, consider the function $g(t) = e^t + e^{-t}$ for which we have

$$\xi_g(r,t) \le \exp(2t - e^{3t/2})$$

valid for all r and t such that $r \le te^{t/2}$. Let ϕ_n denote the standard normal density function on $\mathbb{R}^n$,

$$\phi_n(x) = (2\pi)^{-n/2} e^{-||x||_2^2/2}$$

Note that the density of *any* absolutely continuous product measure can be written as

$$f(x) = \exp\left(-\sum_{i=1}^{n} g_i(x_i)\right)$$

Theorem 46. *Let $n \ge 2$ and for each $1 \le i \le n$ let $g_i : \mathbb{R} \to \mathbb{R}$ be a convex function with corresponding modulus ξ_i as defined by (4.1.1). Let $\xi(r,t) = \max_i \xi_i(r,t)$. Assume that there exist $\sigma > 1$, $\omega > 0$ such that for all $1 \le i \le n$ and $1 \le j \le n$, and all $t \in \mathbb{R}$ with $|t| > t_0$,*

$$g_j'(\sigma^{-1}t) < g_i'(\pm t) < g_j'(\sigma t) \tag{4.1.3}$$

$$g_i''(t) > \omega t^{-1} g_i'(t) \tag{4.1.4}$$

Consider the function $f(x) = \exp\left(-\sum_{i=1}^n g_i(x_i)\right)$. *There exist* $c, \widetilde{c} > 0$ *with the following property: for all* $\theta \in S^{n-1}$ *with* $q = \min_{1 \le i \le n} |\theta_i| \neq 0$, *all* $r > 0$, *and all* $T > 0$ *with* $nr^3\xi(r, \widetilde{c}q^cT) < 6$, *there exists* $\alpha > 0$, $y \in \mathbb{R}^n$ *and a linear injection* $Q : \mathbb{R}^{n-1} \to \{\theta\}^\perp$ *such that* $\langle \theta, y \rangle = T$ *and for all* $x \in \mathbb{R}^{n-1}$ *with* $||x||_2 \le r$,

$$\left| \frac{\alpha f(Qx + y)}{\phi_{n-1}(x)} - 1 \right| < nr^3\xi(r, \widetilde{c}q^cT)$$

In the special case of theorem 46 where $g_i = g$ for all $1 \le i \le n$ and $\theta = (1/\sqrt{n}, 1/\sqrt{n}, \ldots 1/\sqrt{n})$, we can take

$$\alpha = (2\pi)^{-(n-1)/2} \exp(ng(T/\sqrt{n}))$$

and we can express $Q = \beta\widetilde{Q}$ where $\widetilde{Q} : \mathbb{R}^{n-1} \to \theta^\perp$ is a linear isometry independent of T and

$$\beta = \frac{1}{\sqrt{g''(T/\sqrt{n})}}$$

Of particular interest are the functions $f(x) = e^{-||x||_p^p}$ $(1 < p < \infty)$. These functions were studied in the papers [64] and [25] and have an interesting geometry.

Corollary 47. *For any* $1 < p < \infty$, *define* $f : \ell_p \to \mathbb{R}$ *by* $f(x) = e^{-||x||_p^p}$. *For all* $n \in \mathbb{N}$, *all compact sets* $\Omega \subset \mathbb{R}^n$ *and all* $\varepsilon > 0$ *there exists* $\alpha > 0$ *and an affine injection* $T : \mathbb{R}^n \to \ell_p$ *such that*

$$\sup_{x \in \Omega} |\alpha f(Tx) - \phi_n(x)| < \varepsilon$$

4.2 A simple example

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x) = e^{-(x_1^4 + x_2^4)}$$

For $T > 0$ we parameterize the line $x_1 + x_2 = T$ as $x_1 = T/2 + st$, $x_2 = T/2 - st$, where $s > 0$ is a scale parameter to be determined in a moment. We express

$$f(T/2 + st, T/2 - st) = \exp\left(-\left(\frac{1}{8}T^4 + 3T^2s^2t^2 + 2s^4t^4\right)\right)$$

Setting $\alpha = \exp(T^4/8)/\sqrt{2\pi}$ and $s = 1/(T\sqrt{6})$ yeilds the correct variance

$$\alpha f(T/2 + st, T/2 - st) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}\exp\left(-\frac{1}{18T^4}t^4\right)$$

For large values of T, the factor $\exp\left(-T^{-4}t^4/18\right)$ is approximately equal to 1 for all t in the interval $[-\sqrt{T}, \sqrt{T}]$ and the approximation by the normal density function is seen to hold.

4.3 Further discussion

Consider a polynomial $g(x) = a + bx + cx^2 + dx^3$ and the corresponding normalization $g(c^{-1/2}x) = a + bc^{-1/2}x + x^2 + dc^{-3/2}x^3$. If $dc^{-3/2}$ is small, then g resembles a quadratic function around zero in the sense that the quadratic term becomes large while the cubic term is still negligible. This happens, for example, on the interval $-c^{-1/4}d^{-1/6} \leq x \leq c^{-1/4}d^{-1/6}$. If in addition $g'(0) = 0$, then for $\widetilde{c} = e^{-a}/\sqrt{2\pi}$ the function $f(x) = \widetilde{c}e^{-g(x/\sqrt{c})} = \exp(-dc^{-3/2}x^3)\phi_1(x)$ approximates the standard normal density function on an interval $[-R, R]$ for some large value of R, say $R = c^{1/4}d^{-1/6}$. This is the motivation for studying the modulus defined by (4.1.1).

To see that condition (4.1.2) is not too restrictive, consider the following lemma which we prove later on.

Lemma 48. *If* $\omega : [0, \infty) \to \mathbb{R}$ *is any differentiable function with* $\lim_{t\to\infty} t^2\omega(t) = \infty$, *then for all* $\varepsilon > 0$

$$\liminf_{t\to\infty} \left| \frac{\omega'(t)}{\omega(t)^{1+\varepsilon}} \right| = 0$$

The function ω plays the role of g'' and we take $\varepsilon = 1/2$. There are two reasons why not every function satisfies (4.1.2). The first is that there could be infinite oscillation in the tails of $|g'''(t)|g''(t)^{-3/2}$ whereby the $\liminf$ is zero but the $\limsup$ is strictly positive. The second is due to the perturbation s, however the condition $\lim_{t\to\pm\infty} t^2 g''(t) = \infty$ implies that $|s|$ is only a small proportion of $|w|$. The appearance of w is simply to insure that $\xi(r,t)$ is non-increasing in t; one could just as well erase it and use t instead.

Conditions (4.1.3) and (4.1.4) are not fundamental to theorem 46 and are only imposed to obtain a quantitative bound. Their role is to provide a linear lower bound on the growth of the coordinates of y as $T \to \infty$. Without them the coordinates of y would still converge to ∞, just not at a linear rate.

4.4 Proofs

The proof below is split into two parts for ease of reading. The fundamental ingredients are contained mainly in the second part.

Proof of theorem 46. **Part 1:** Without loss of generality $\theta_i > 0$ for all $1 \le i \le n$. Consider the function $g : \mathbb{R}^n \to \mathbb{R}$ defined by $g(x) = \sum_{i=1}^n g_i(x_i)$. By convolution with a smooth test function we may assume that $\nabla g(x)$ exists for all $x \in \mathbb{R}^n$ and that g is strictly convex. By (4.1.4) there exist $c_1, t_1 > 0$ such that $g_i'(t) > c_1 t^\omega$ for all $t > t_1$ and all $1 \le i \le n$. Define $T_1 = \sup\{g_i'(t) : 1 \le i \le n, t \le t_0\}$ and consider any $T > \max\{\sqrt{n}t_1, c_1^{-1/\omega}\sqrt{n}q^{-1/\omega}T_1^{1/\omega}\}$. Restricted to the affine subspace

$E = \{x \in \mathbb{R}^n : \langle x, \theta \rangle = T\}$, the function g attains a minimum at some point $y \in E$. By the theory of Lagrange multipliers, this point satisfies $\nabla g(y) = \lambda\theta$ for some $\lambda \neq 0$, i.e. $g_i'(y_i) = \lambda\theta_i$ for all $1 \leq i \leq n$. Since $T = \sum \theta_i y_i \leq ||\theta||_1 \max\{y_i\}_1^n$ and $1 \leq ||\theta||_1 \leq \sqrt{n}$, it follows that for some $1 \leq k \leq n$, $y_k = \max\{y_i\}_1^n \geq T/\sqrt{n} > t_1$. Hence $g_k'(y_k) > c_1 y_k^\omega \geq c_1 n^{-\omega/2} T^\omega$ and for all $1 \leq i \leq n$, $g_i'(y_i) = \lambda\theta_i \geq q\lambda\theta_k = q g_k'(y_k) \geq c_1 q n^{-\omega/2} T^\omega > T_1$ (by definition of T). By definition of T_1, $y_i > t_0$ (for all i). By (4.1.4), for all $k > 1$ and $s > t_0$, $g_i'(ks) > k^\omega g_i'(s)$. Hence, if $s_1 > s_2 > t_0$ then

$$\frac{s_1}{s_2} < \left(\frac{g_i'(s_1)}{g_i'(s_2)}\right)^{1/\omega} \tag{4.4.1}$$

We consider two cases. In case 1, $y_i \geq \sigma^{-1} y_k$. In case 2, $y_i < \sigma^{-1} y_k$ and we can apply inequality (4.4.1) which gives

$$\frac{\sigma^{-1} y_k}{y_i} < \left(\frac{g_i'(\sigma^{-1} y_k)}{g_i'(y_i)}\right)^{1/\omega} < \left(\frac{g_k'(y_k)}{g_i'(y_i)}\right)^{1/\omega} = \left(\frac{\theta_k}{\theta_i}\right)^{1/\omega} \leq q^{-1/\omega}$$

and $y_i > \sigma^{-1} q^{1/\omega} y_k \geq \sigma^{-1} q^{1/\omega} n^{-1/2} T$. In either case,

$$y_i > \widetilde{c} q^c T$$

Part 2: Any two Hilbert spaces of the same dimension are linearly isometric. Since the norm $|| \cdot ||_\sharp$ defined by

$$||z||_\sharp = \left(\sum_{i=1}^n z_i^2 g_i''(y_i)\right)^{1/2}$$

is Hilbertian, there exists a linear embedding $Q : \mathbb{R}^{n-1} \to \{\theta\}^\perp$ such that $||Qz'||_\sharp = ||z'||_2$ for all $z' \in \mathbb{R}^{n-1}$. Fix any $x \in \mathbb{R}^{n-1}$ and express $x = ru$ with $r \geq 0$ and $u \in S^{n-2}$. Let $\eta = Qu$ and define

$$\begin{aligned}\psi(s) &= -\log(\alpha(t) f(s\eta + y)) \\ &= (n-1)\log\sqrt{2\pi} - \sum_{i=1}^n g_i(y_i) + \sum_{i=1}^n g_i(s\eta_i + y_i)\end{aligned}$$

By definition of ξ, for all $t \in \mathbb{R}$, all $1 \leq i \leq n$ and all s' with $|s'| < r/\sqrt{g_i''(y_i)}$,

$$\left| \frac{g_i'''(y_i + s')}{g_i''(y_i)^{3/2}} \right| < \xi(r, y_i) \leq \xi(r, cq^c T)$$

By the chain rule, $\psi(0) = (n-1)\log\sqrt{2\pi}$, $\psi'(0) = \langle \eta, \nabla g(y) \rangle = 0$, $\psi''(0) = ||\eta||_\sharp^2 = 1$ and

$$\psi'''(s) = \sum_{i=1}^{n} \nu_i^3 \frac{g_i'''(y_i + s\eta_i)}{g''(y_i)^{3/2}}$$

where $\nu_i = \eta_i \sqrt{g''(y_i)}$ for all $1 \leq i \leq n$, and $||\nu||_2 = 1$. Since $||\nu||_\infty \leq 1$, for all $s \in [0, r]$ and all $1 \leq i \leq n$ the quantity $s' = s\eta_i$ obeys $|s'| \leq r/\sqrt{g_i''(y_i)}$. Hence

$$|\psi'''(s)| \leq n\xi(r, cq^c T)$$

and by Taylor's theorem,

$$\sup_{s \in [0,r]} |\psi(s) - \frac{1}{2}s^2 - (n-1)\log\sqrt{2\pi}| \leq 6^{-1} n r^3 \xi(r, cq^c T)$$

which gives

$$\left| \frac{\alpha(t) f(T_t x + t\theta)}{\phi_{n-1}(x)} - 1 \right| < \exp\left(6^{-1} n r^3 \xi(r, cq^c T)\right) - 1$$

the result follows from the inequality $1 + \delta \leq e^\delta \leq 1 + 3\delta$ valid for all $\delta \in [0, 1]$. $\square$

Proof of lemma 48. Suppose that the result does not hold. Then there exists $t_0, c > 0$ such that for all $t > t_0$, $\omega(t) > t^{-2}$ and $|\omega'(t)| > c\omega(t)^{1+\varepsilon}$. In particular, $\omega'(t) \neq 0$ and ω is either strictly increasing on (t_0, ∞), or strictly decreasing on (t_0, ∞). Hence ω' does not change sign on (t_0, ∞). If $\omega' < 0$ then ω decays exponentially (or quicker) violating the inequality $\omega(t) > t^{-2}$. Thus $\omega' > 0$. Since ω is injective on (t_0, ∞) it satisfies an autonomous differential equation $\omega'(t) = \Theta(\omega(t))$, where

$\Theta(s) = \omega'(\omega^{-1}(s))$ and $\Theta(s) > cs^{1+\varepsilon}$ for all $s > s_0$. Note that,

$$\int_{s_0}^{\omega(t)} \frac{1}{\Theta(s)} ds = t - \omega^{-1}(s_0) \tag{4.4.2}$$

(just differentiate both sides to see why). This is a contradiction because the left hand side of (4.4.2) is bounded (as a function of t) while the right hand side is not. □

Bibliography

[1] Ajtai, M., Komlòs, J., Tusnàdy, G.: On optimal matchings. Combinatorica **4,** 259-264 (1984)

[2] Anttila, M., Ball, K., Perissinaki, I.: The central limit problem for convex bodies. Trans. Amer. Math. Soc. **355** (12), 4723-4735 (2003)

[3] Bagnoli, M., Bergstrom, T.: Log-concave probability and its applications. Econom. Theory **26** (2) 445-469 (2005)

[4] Balakrishnan, N., Clifford Cohen, A.: Order Statistics and Inference. Statistical Modeling and Decision Science. Academic Press (1991)

[5] Ball, K.: Logarithmically concave functions and sections of convex sets in $\mathbb{R}^n$. Studia Math. **88** (1), 69-84 (1988)

[6] Ball, K.: An elementary introduction to modern convex geometry. Flavors of Geometry, MSRI Publications 31 (1997)

[7] Ball, K.: Ellipsoids of maximal volume in convex bodies. Geom. Dedicata **41** (2), 241-250 (1992)

[8] Bárány, I., Larman, D. G.: Convex bodies, economic cap coverings, random polytopes. Mathematika **35** (2), 274-291 (1988)

[9] Bárány, I., Vu, V.: Central limit theorems for Gaussian polytopes. Ann. Probab. **35** (4), 1593-1621 (2007)

[10] Bárány, I.: Random points and lattice points in convex bodies. Bull. Amer. Math. Soc. **45** (3) 339-365 (2008)

[11] Barthe, F., Guédon, O., Mendelson, S., Naor, A.: A probabilistic approach to the geometry of the ℓ_p^n-ball. Ann. Probab. **33** (2), 480-513 (2005)

[12] Bobkov, S.: On concentration of distributions of random weighted sums. Ann. Probab. **31** (1), 195-215 (2003)

[13] Bobkov, S., Madiman, M.: Concentration of the information in data with log-concave distributions. Ann. Probab. **39** (4), 1528-1543 (2011)

[14] Dafnis, N., Gianopoulis, Tsolomitis.: Asymptotic shape of a random polytope in a convex body. J. Funct. Anal. **257** (9), 2820-2839 (2009)

[15] David, H. A.: Order Statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons (1970)

[16] Dehardt, J.: Generalizations of the Glivenko-Cantelli theorem. Ann. Math. Statist. **42** (6), 2050-2055 (1971)

[17] Devroye, L.: Non-Uniform Random Variate Generation. Originally published with Springer-Verlag, New York (1986)

[18] Diaconis, P., Freedman, D.: Asymptotics of graphical projection pursuit. Ann. Stat. **12** (3), 793-815 (1984)

[19] Donsker, M. D.: Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. Ann. Math. Statist. **23**, 277–281 (1952)

[20] Dudley, R. M.: Real Analysis and Probability. Wadsworth & Brooks/Cole (1989)

[21] Dvoretzky, A., Kiefer, J., Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. **27** (3), 642–669 (1956)

[22] Eldan, R., Klartag, B.: Pointwise estimates for marginals of convex bodies. J. Funct. Anal. **254** (8), 2275-2293 (2008)

[23] Fleury, B.: Concentration in a thin Euclidean shell for log-concave measures. J. Func. Anal. **259** (4), 832-841 (2010)

[24] Fleury, B., Guédon, O., Paouris, G.: A stability result for mean width of L_p-centroid bodies. Adv. Math. **214** (2), 865-877 (2007)

[25] Fresen, D.: A multivariate Gnedenko law of large numbers. arXiv:1101.4887

[26] Fresen, D.: The floating body and the hyperplane conjecture. Arch. Math. **98**, 389-397 (2012)

[27] Fresen, D.: Simultaneous concentration of order statistics. arXiv:1102.1128

[28] Fresen, D.: Central limit theorems for non-central sections of log-concave product measures and star shaped functions. arXiv:108.5011

[29] Fresen, D.: Concentration inequalities for Lipschitz functions into an arbitrary metric space.

[30] Funano, K.: Observable concentration of mm-spaces into spaces with doubling measures. Geom. Dedicata **127**, 49-56 (2007)

[31] Funano, F.: Concentration of 1-Lipschitz maps into an infinite dimensional ℓ^p-ball with the ℓ^q-distance function. Proc. Amer. Math. Soc. **137** (7) 2407-2417 (2009)

[32] Funano, K.: Exponential and Gaussian concentration of 1-Lipschitz maps. Manuscripta Math. **130** (3), 273-285 (2009)

[33] Gaensller, P., Stute, W.: Empirical processes: a survey of results for independent and identically distributed random variables. Ann. Probab. **7** (2) 193-243 (1979)

[34] Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math. **44,** 423-453 (1943)

[35] Goodman, V.: Characteristics of normal samples. Ann. Probab. **16** (3), 1281-1290 (1988)

[36] Gordon, Y., Litvak, A., Schütt, C., Werner, E.: Uniform estimates for order statistics and Orlicz functions. arXiv:0809.2989v1

[37] Gromov, M.: CAT(κ)-spaces: construction and concentration (Russian summary) Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 280, Geom. i Topol. 7: 100-140, 299-300 (2001); translation in J. Math. Sci. (N. Y.) **119** (2), 178-200 (2004)

[38] Gromov, M.: Isoperimetry of waists and concentration of maps. Geom. Funct. Anal. **13** (1), 178-215 (2003)

[39] Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces, Based on the 1981 French original, with appendices by M. Katz, P. Pansu and S. Semmes. Translated from the French by Sean Michael Bates. Progress in Mathematics, vol. 152. Birkhäuser Boston, Inc., Boston, MA (1999)

[40] Gruber, P. M.: Aspects of approximation of convex bodies. Handbook of Convex Geometry, Vol. A. 319-345, North Holland, Amsterdam (1993)

[41] Guédon, O., Milman, E.: Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. Geom. Funct. Anal. **21** (5), 1043-1068 (2011)

[42] Klartag, B.: $5n$ Minkowski symmetrizations suffice to arrive at an approximate Euclidean ball. Ann. of Math. **156** (3), 947–960 (2002)

[43] Klartag, B., Milman, V.: Geometry of log-concave functions and measures. Geom. Dedicata **112,** 169-182 (2005).

[44] Klartag, B.: A central limit theorem for convex sets. Invent. Math. **168** (1) 91-131 (2007)

[45] Klartag, B.: Power-law estimates for the central limit theorem for convex sets. J. Funct. Anal. **245** (1), 284-310 (2007)

[46] Klartag, B.: A Berry-Esseen type inequality for convex bodies with an unconditional basis. Probab. Theory Relat. Fields. **145** (1-2), 1-33 (2009)

[47] Klartag, B.: High-dimensional distributions with convexity properties. A text based on a talk given by the author at the fifth European Congress of Mathematics. Available online at http://www.math.tau.ac.il/∼klartagb/index.html

[48] Koldobsky, A.: Fourier Analysis in Convex Geometry, Mathematical Surveys and Monographs, American Mathematical Society, Providence RI 2005.

[49] Komlòs, J., Major, P., Tusnàdy, G.: An approximation of partial sums of independent RV'-s and the sample DF. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 111–131. (1975)

[50] Loève, M.: Probability Theory. D. Van Nostrand Company (1955)

[51] Lovász, L., Vempala, S.: The geometry of logconcave functions and sampling algorithms. Random Structures Algorithms **30** (3), 307-358 (2007)

[52] Mason, D. M., van Zwet, W.: A refinement of the KMT inequality for the uniform empirical process. Ann. Probab. **15**, 871-884 (1987)

[53] Massart, P.: The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. Ann. Probab. **18** (3), 1269–1283 (1990)

[54] Matoušek, J.: Lectures on Discrete Geometry. Graduate Texts in Mathematics. Springer-Verlag, New York (2002)

[55] Meckes, M.: Gaussian marginals of convex bodies with symmetries. Beiträge Algebra Geom. **50** (1), 101-118 (2009)

[56] Meckes, E., Meckes, M.: The central limit problem for random vectors with symmetries. J. Theor. Probab. **20**, 697-720 (2007)

[57] Milman, V., Schechtman, G.: Asymptotic Theory of Finite-Dimensional Normed Spaces. Lect. Notes Math., vol. 1200. Springer, Berlin (1986)

[58] Naor, A., Romik, D.: Projecting the surface measure of the sphere of ℓ_p^n. Ann. Inst. H. Poincaré Probab. Statist. **39** (2), 241-261 (2003)

[59] Pisier, G.: The Volume of Convex Bodies and Banach Space Geometry. In: CambridgeTracts in Mathematics, vol. 94. Cambridge University Press, Cambridge1989.

[60] Pitman, E. J. G.: Some Basic Theory for Statistical Inference. Chapman and Hall (1979)

[61] Raynaud, H.: Sur l'enveloppe convexe des nauges de points aléatoires dans $\mathbb{R}^n$. J. Appl. Probab., **7**, 35-48 (1970)

[62] Romik, D.: Randomized Central Limit Theorems - Probabilistic and Geometric Aspects. PhD dissertation Tel Aviv University (2001)

[63] Sarhan, A. E., Greenberg, B. G. (eds.): Contributions to Order Statistics. Wiley Publications in Statistics. John Wiley & Sons (1962)

[64] Schechtman, G., Zinn, J.: On the volume of the intersection of two L_p^n balls. Proc. Amer. Math. Soc. **110** (1), 217-224 (1990)

[65] Schütt, C., Werner, E.: The convex floating body. Math. Scand. **66**, 275-290 (1990)

[66] Shorack, G., Wellner, J.: Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons (1986)

[67] Sudakov, V. N.: Typical distribution of linear functionals in finite-dimensional spaces of high-dimension. Dokl., Akad. Nauk Azerb. **243** (6), 1402-1405 (1978) (Russian) English translation in Sov. Math. Dokl. **19**, 1578-1582 (1978)

[68] Talagrand, M.: The Glivenko-Cantelli problem. Ann. Probab. **15** (3), 837-870 (1987)

[69] Talagrand, M.: Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math. **81**, 73-205 (1995)

[70] Talagrand, M.: The Glivenko-Cantelli problem, ten years later. J. Theoret. Probab. **9** (2), 371-384 (1996)

[71] von Weizsäcker, H.: Sudakov's typical marginals, random linear functionals and a conditional central limit theorem. Probab. Theory Relat. Fields **107** (3), 313-324 (1997)

[72] Wellner, J.: A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. Ann. Statist. **5** (3), 473-480 (1977)

[73] Wellner, J.: Limit theorems for the ratio of the empirical distribution function to the true distribution function. Probab. Theory Relat. Fields **45** (1) 73-88 (1978)

VITA

Daniel Fresen was born in Cape Town, South Africa, in 1983. He matriculated from St Alban's College in 2001 and recieved his BSc in Mathematics from the University of Pretoria in 2005. In 2006 he came to the U.S.A. to persue a PhD in mathematics at the University of Missouri under the supervision of Nigel Kalton. Nigel passed away in 2010 and Daniel continued his work under the supervision of Alexander Koldobsky and Mark Rudelson. He is expected to graduate in May 2012 and will move to New Haven, Connecticut, in July to begin as a J. Willard Gibbs Assistant Professor at Yale University.