A theory of convexity for probability measures

Mokshay Madiman

November 27, 2016

Abstract

While convexity of sets and functions on linear spaces is bread-and-butter to any mathematician, the notion of convexity for measures– although deeply studied by Christer Borell starting in the 1970's– is far less well known. The goal of these lectures will be to provide an exposition as well as motivation for the theory of convex measures, focusing in particular on the important and ubiquitous subclass of log-concave measures. The lectures will choose an eclectic path through some of the many aspects of the theory that have been developed in the last 40 years, covering some classical topics (such as the Prékopa-Leindler inequality), some developed around the turn of the millennium (such as connections to the concentration of measure phenomenon), and some very recent topics (such as Klartag's central limit theorem and the concentration of information content).

Contents

1 Lecture 1: What is a convex measure? 3

1.1 Convex sets and functions 3

1.1.1 Convex sets 3

1.1.2 Convex and s-concave functions 5

1.2 Convex measures 6

1.2.1 The definitions 6

1.2.2 Measurability issues 7

1.3 Exogenous motivations for studying convex measures 8

1.3.1 Geometric functional analysis 8

1.3.2 Stochastic optimization 10

1.3.3 Statistical inference 10

2 Lecture 2: Borell's characterization of convex measures 11

2.1 Convexity properties of Lebesgue measure 11

2.1.1 The Brunn-Minkowski inequality 11

2.1.2 Isoperimetric Inequality for Lebesgue measure 12

2.2 The characterization (without proofs) 13

2.3 Consequences of Borell's characterization theorem 15

2.4 The Prékopa-Leindler inequality 16

2.5 Proof of the characterization 18

3 Lecture 3: Reverse Hölder inequalities and information concentration **19**
3.1 Motivations . . . 19
3.1.1 The Gaussian thin shell property and its possible generalizations . . . 19
3.1.2 The Shannon-McMillan-Breiman theorem and its implications . . . 20
3.2 A basic reverse Hölder inequality . . . 21
3.3 Sharp varentropy bound . . . 22
3.4 Sharp concentration of information . . . 26
3.5 What happens for s-concave case . . . 30

4 Lecture 4: Convexity, curvature, and concentration **34**
4.1 Weighted Prékopa-Leindler inequality . . . 34
4.2 Properties of uniformly log-concave measures . . . 34
4.3 Aside: Ways of capturing curvature . . . 34

5 Lecture 5: Klartag's CLT **35**
5.1 Marginals and Gaussian mixtures . . . 35
5.2 The thin shell phenomenon . . . 35

Note to readers: These notes had their genesis in a course for Ph.D. students taught at Yale University in the spring of 2011, titled "Probabilistic Convex Geometry", They were further extended during lectures at TIFR CAM in Bengaluru during December 2014, and put into their present form in preparation for the Lectures on Probability and Statistics XI organized at ISI Delhi in November 2016.

Eventually these notes may wind up in some published form. For now, they are only being made available for interested readers– it should be emphasized that some parts are thin on details (some of the more involved proofs are missing, for instance), that the bibliography is incomplete, and that the later sections are almost entirely missing. The author would very much appreciate being notified of errors, and also would welcome other suggestions.

1 Lecture 1: What is a convex measure?

We begin with a discussion of convex sets in linear spaces, skipping over details and some proofs, because a central idea of interest to us– that of Minkowski summation– is naturally tied to basic consideration of convex sets.

1.1 Convex sets and functions

1.1.1 Convex sets

Let V be a linear space. A fundamental construction on linear spaces is that of the *convex hull* of a subset A of V, defined as the set of all convex combinations of finitely many points of A. Denoting by $[x, y]$ the line segment $\{\lambda x + (1-\lambda)y : \lambda \in [0,1]\}$, the set $K \subset V$ is convex if $[x, y] \subset K$ whenever $x, y \in K$. It is a standard exercise to check that the convex hull of A is the "smallest" convex set containing A, in the sense that it is the intersection of all convex sets containing A.

Convex sets arise naturally in linear algebra and functional analysis because they allow for separation theorems (Hahn-Banach being the prime example), which form a very powerful tool in analysis as well as optimization. However, we want to emphasize that convex sets also arise naturally just by iterating the basic operations of summation and scaling that are available on a linear space– this fact is the Emerson-Greenleaf-Shapley-Folkmann-Starr theorem, which we now describe.

Minkowski summation is a basic and ubiquitous operation on sets. Indeed, the *Minkowski sum* $A + B = \{a + b : a \in A, b \in B\}$ of sets A and B makes sense as long as A and B are subsets of an ambient set in which the operation $+$ is defined. In particular, this notion makes sense in any group, and there are multiple fields of mathematics that are preoccupied with studying what exactly this operation does. For example, much of classical additive combinatorics studies the cardinality of Minkowski sums (called sumsets in this context) of finite subsets of a group and their interaction with additive structure of the concerned sets, and the study of the Lebesgue measure of Minkowski sums in $\mathbb{R}^n$ is central to much of convex geometry and geometric functional analysis. Here we only aim to expose the qualitative effect of Minkowski summation in $\mathbb{R}^n$– specifically, the "convexifying" effect that it has.

The fact that Minkowski summation produces sets that look "more convex" is easy to visualize by drawing a non-convex set[1] in the plane and its self-averages $A(k)$ defined by

$$A(k) = \left\{ \frac{a_1 + \cdots + a_k}{k} : a_1, \ldots, a_k \in A \right\} = \frac{1}{k}\Big(\underbrace{A + \cdots + A}_{k \text{ times}}\Big). \tag{1}$$

This intuition was first made precise in the late 1960's independently by Starr [63], who credited Shapley and Folkmann for the main result, and by Emerson and Greenleaf [25]. Denote by $\mathrm{conv}(A)$ the convex hull of A, by B_2^n the n-dimensional Euclidean ball, and by $d(A) = \inf\{r > 0 : \mathrm{conv}(A) \subset A + rB_2^n\}$ the *Hausdorff distance* between a set A and its convex hull.

Theorem 1 (Emerson-Greenleaf-Shapley-Folkmann-Starr)**.** *If A is a compact subset of $\mathbb{R}^n$, $A(k)$ converges in Hausdorff distance to $\mathrm{conv}(A)$ as $k \to \infty$, at rate $O(1/k)$. More precisely, $d(A(k)) = O\big(\frac{\sqrt{n}}{k}\big)$.*

[1] The simplest nontrivial example is three non-collinear points in the plane, so that $A(k)$ is the original set A of vertices of a triangle together with those convex combinations of the vertices formed by rational coefficients with denominator k.

For sets of nonempty interior, this convergence of Minkowski averages to the convex hull can also be expressed in terms of the *volume deficit* $\Delta(A)$ of a compact set A in $\mathbb{R}^n$, which is defined as:

$$\Delta(A) := \mathrm{vol}_n(\mathrm{conv}(A) \setminus A) = \mathrm{vol}_n(\mathrm{conv}(A)) - \mathrm{vol}_n(A),$$

where vol_n denotes Lebesgue measure in $\mathbb{R}^n$.

Theorem 2 (Emerson-Greenleaf)**.** *If $A \subset \mathbb{R}^n$ is compact with nonempty interior, then the volume deficit of $A(k)$ also converges to 0. More precisely, when n is fixed*[2]*, $\Delta(A(k)) = O(1/k)$.*

Our geometric intuition would suggest that in some sense, as k increases, the set $A(k)$ is getting progressively more convex, or in other words, that the convergence of $A(k)$ to $\mathrm{conv}(A)$ is, in some sense, monotone. It was conjectured[3] in [11] that if A is a compact set in $\mathbb{R}^n$, then the sequence $\Delta(A(k))$ is non-increasing in k, or equivalently, $\{\mathrm{vol}_n(A(k))\}_{k\geq 1}$ is non-decreasing. Note that the trivial inclusion[4] $A \subset \frac{A+A}{2}$ immediately implies that the subsequence $\{\mathrm{vol}_n(A(2^k))\}_{k\in\mathbb{N}}$ is non-decreasing; however, the challenge appears at the very next step since $A(2)$ is generally not a subset of $A(3)$. Remarkably, the conjecture is false [29].

Theorem 3 (Fradelizi-Madiman-Marsiglietti-Zvavitch)**.** *If $A \subset \mathbb{R}^1$ is compact, $\Delta(A(k))$ is non-increasing in k. If $n \geq 12$, there exist sets $A \subset \mathbb{R}^n$ such that $\Delta(A(k))$ is not monotone in k.*

It is natural to wonder if monotonicity holds for Hausdorff distance. It was proved in [29] that if A is a compact set in $\mathbb{R}^n$ and $k \geq n$, then $d\left(A(k+1)\right) \leq \frac{k}{k+1} d\left(A(k)\right)$. Note this also yields the $O(1/k)$ convergence rate as a corollary. To prove this fact, we need a barometer of non-convexity that was introduced by Schneider.

The "non-convexity index" of Schneider [61] is defined as

$$c(A) = \inf\{\lambda \geq 0 : A + \lambda\, \mathrm{conv}(A) \text{ is convex}\}.$$

Clearly $c(A) = 0$ if A is convex (just select $\lambda = 0$). On the other hand, if $c(A) = 0$ and A is compact, then $\{A + \frac{1}{m}\mathrm{conv}(A)\}_{m=1}^{\infty}$ is a sequence of compact convex sets, converging in Hausdorff metric to A, thus A must be convex. In other words, for compact sets A, $c(A) = 0$ if and only if A is convex, which makes it a legitimate barometer of non-convexity.

Lemma 1 (Schneider[5])**.** *For any subset A of $\mathbb{R}^n$,*

$$c(A) \leq n.$$

Proof. We want to prove that $A + n\mathrm{conv}(A)$ is convex. We only need to prove that $(n+1)\mathrm{conv}(A) \subset n\mathrm{conv}(A) + A$. Let $x \in (n+1)\mathrm{conv}(A)$. From Carathéodory's theorem, there exists $a_1, \ldots, a_{n+1} \in A$ and $\lambda_1, \ldots, \lambda_{n+1} \geq 0$ such that $x = (n+1)\sum_{i=1}^{n+1} \lambda_i a_i$ with $\sum_{i=1}^{n+1} \lambda_i =$

[2] The dependence on n is a much more complicated matter for the volume deficit.

[3] In fact, they conjectured a more general monotonicity property in a probabilistic limit theorem, namely the law of large numbers for random sets due to Artstein and Vitale [3].

[4] It is an easy exercise to check that $A = \frac{A+A}{2}$ characterizes convexity (indeed, $A \supset \frac{A+A}{2}$ is just a restatement of the fact that midpoint-convexity is equivalent to convexity).

[5] Schneider [61] showed, moreover, that $c(A) = n$ if and only if A consists of $n+1$ affinely independent points.

1. Since $1 = \sum_{i=1}^{n+1} \lambda_i \leq (n+1)\max_i \lambda_i$, there exists i_0 such that $\lambda_{i_0} \geq \frac{1}{n+1}$. Define $\mu_{i_0} = \frac{(n+1)\lambda_{i_0}-1}{n}$ and $\mu_i = \frac{n+1}{n}\lambda_i$ for $i \neq i_0$. Then $\mu_i \geq 0$ for all i, $\sum_i \mu_i = 1$ and

$$x = a_{i_0} + ((n+1)\lambda_{i_0} - 1)a_{i_0} + (n+1)\sum_{i\neq i_0} \lambda_i a_i = a_{i_0} + n\sum_i \mu_i a_i,$$

which shows that $x \in A + n\mathrm{conv}(A)$. □

Let us note an alternative representation of Schneider's non-convexity index. For any set $A \subset \mathbb{R}^n$, define

$$A_\lambda = \frac{1}{1+\lambda}[A + \lambda\,\mathrm{conv}(A)],$$

and observe that $A_\lambda \subset \mathrm{conv}(A)$, so that in particular $\mathrm{conv}(A_\lambda) \subset \mathrm{conv}(A)$. Moreover, $A \subset A_\lambda$ implies that $\mathrm{conv}(A) \subset \mathrm{conv}(A_\lambda)$. From the preceding two conclusions, we have that $\mathrm{conv}(A) = \mathrm{conv}(A_\lambda)$ for any $\lambda \geq 0$. Since A_λ is convex if and only if $A + \lambda\,\mathrm{conv}(A)$ is convex, we can express

$$c(A) = \inf\{\lambda \geq 0 : A_\lambda \text{ is convex}\} = \inf\{\lambda \geq 0 : A_\lambda = \mathrm{conv}(A)\}. \tag{2}$$

Theorem 4 (Fradelizi-Madiman-Marsiglietti-Zvavitch)**.** *Let A be a compact set in $\mathbb{R}^n$ and $k \geq c(A)$ be an integer. Then*

$$d\left(A(k+1)\right) \leq \frac{k}{k+1} d\left(A(k)\right).$$

Proof. Let $k \geq c(A)$, then, from the definitions of $c(A)$ and $d(A(k))$, one has

$$\begin{aligned} \mathrm{conv}(A) = \frac{A}{k+1} + \frac{k}{k+1}\mathrm{conv}(A) \quad &\subset \quad \frac{A}{k+1} + \frac{k}{k+1}\left(A(k) + d(A(k))B_2^n\right) \\ &= \quad A(k+1) + \frac{k}{k+1} d(A(k))B_2^n. \end{aligned}$$

This immediately implies the desired statement. □

There remain surprisingly simple-to-state open questions. For $2 \leq n \leq 12$, it is unknown if $\Delta(A(k))$ is non-increasing in k for compact $A \subset \mathbb{R}^n$. Also it is unknown if for fixed dimension n, $\Delta(A(k))$ is non-increasing once k exceeds some number depending on n.

1.1.2 Convex and s-concave functions

If V is a linear space, $f : V \to \mathbb{R} \cup \{\infty\}$ is a *convex function* if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for all $x, y \in V$ and all $\lambda \in [0,1]$. A function f is *concave* if $-f$ is convex. It is convenient to consider a 1-parameter generalization of this standard definition.

As a prerequisite, we define the κ-mean of two numbers, for $a, b \in (0,\infty)$, $t \in (0,1)$ and $\kappa \in (-\infty, 0) \cup (0,\infty)$ define

$$M_\kappa^t(a,b) = ((1-t)a^\kappa + tb^\kappa)^{\frac{1}{\kappa}}. \tag{3}$$

For $\kappa \in \{-\infty, 0, \infty\}$ define $M_\kappa^t(a,b) = \lim_{\kappa' \to \kappa} M_{\kappa'}^t(a,b)$ corresponding to

$$\{\min(a,b), a^{1-t}b^t, \max(a,b)\}$$

respectively. M_κ can be extended to $a, b \in [0,\infty)$ via direct evaluation when $\kappa \geq 0$ and again by limits when $\kappa < 0$ so that $M_\kappa(a,b) = 0$ whenever $ab = 0$.

Definition 1. *Fix $s \in [-\infty, \infty]$. A function $f : \mathbb{R}^n \to [0, \infty)$ is a s-concave function if*

$$f((1-t)x + ty) \geq M_s^t(f(x), f(y)),$$

whenever $f(x)f(y) > 0$ and $t \in (0,1)$.

The function $s \to M_s^{(t)}(a, b)$ is non-decreasing, so the defining inequality is getting stronger as s grows. In particular:

1. The strongest in the hierarchy– a ∞-concave function– is just a constant function on $\mathbb{R}^n$.

2. The weakest– a $-\infty$-concave function– is one satisfying $f((1-t)x+ty) \geq \min\{f(x), f(y)\}$. Such functions are said to be *quasi-concave*, and are precisely the functions such that the superlevel set $\{x \in \mathbb{R}^n : f(x) \geq \alpha\}$ is convex for each $\alpha \in \mathbb{R}$.

In dimension 1, it is easy to check (exercise!) that quasi-concavity coincides with unimodality (i.e., the function is non-decreasing up to its global maximum and then non-increasing afterwards). In general dimension, quasi-concavity implies unimodality (understood via rays through the global maximum) but the reverse may not be true. Also one can check that in any dimension, a monotonically increasing function of a concave function is quasi-concave (e.g., $(x, y) \mapsto xy + x^2y^2 + x^3y^3$).

A strictly quasi-concave function (i.e., where the inequality in the definition is strict) has a unique maximizer. More generally, the set of maximizers of a quasi-concave function must be a convex set.

Convex functions are, of course, ubiquitous in analysis and probability (where Jensen's inequality is one of the most powerful tools at our disposal). Because of the simplicity afforded by the fact that local minima are necessarily global minima, they are also central to optimization– both its theoretical aspects in the sense that it forms the threshold for polynomial-time optimization algorithms, and to applications, where casting a given optimization problem as a convex program allows one to bring to bear several algorithms that are remarkably efficient in practice and not just theoretically.

Grötschel, Lovász and Schrijver [34] (cf., [64]) showed that the ellipsoid method[6] solves the problem of minimizing a convex function over a convex set in $\mathbb{R}^n$ specified by a separation oracle, i.e., a procedure that reports, given a point x, either that the set contains x or a halfspace that separates the set from x. Bertsimas and Vempala [6] gave another polynomial-time algorithm based on sampling, which applies more generally to the maximization of quasi-concave functions. There are also variants (such as subgradient projection algorithms) of classical subgradient methods that work well on these functions even in higher dimensions.

1.2 Convex measures

1.2.1 The definitions

How would one capture convexity of a measure? Since measures act on sets, comparing with the definition of s-concave functions suggests that a natural definition might involve replacing a convex combination of points by a Minkowski convex combination of sets.

[6] The ellipsoid method was invented by Yudin-Nemirovski [67] for linear programming, and was shown by Khachiyan [37] to be polynomial-time. If one wants to optimize a linear function over a convex set, one now has a wider selection of algorithms to choose from, such as interior point methods introduced by Karmarkar [40] and studied deeply by Nesterov and Nemirovski [52] (see [18] for recent developments utilizing ideas related to the study of convex measures).

Definition 2. *Fix $-\infty \leq \kappa \leq +\infty$. A finite measure μ on $\mathbb{R}^n$ is called* κ-concave *if it satisfies a geometric inequality*

$$\mu(tA + (1-t)B) \geq [t\mu(A)^\kappa + (1-t)\mu(B)^\kappa]^{1/\kappa}, \tag{4}$$

in the class of all non-empty Borel subsets $A, B \subset \mathbb{R}^n$ of positive μ-measure, with arbitrary $t \in (0,1)$.

If a finite measure is κ-concave for some $-\infty \leq \kappa \leq +\infty$, and hence for $\kappa = -\infty$, then it is called a convex measure.

When the law of a random vector X is a κ-concave measure, we will refer to X as a κ-concave random vector.

As we saw, the function $\kappa \to M_\kappa^{(t)}(a,b)$ is non-decreasing, so the inequality (4) is getting stronger, when κ grows. Thus, the case $\kappa = -\infty$ describes the largest class, defined by

$$\mu\big(tA + (1-t)B\big) \geq \min\{\mu(A), \mu(B)\},$$

and its members are the convex measures (also sometimes called "hyperbolic measures").

Remarks:

1. Thus, the κ-concave measures are those that distribute volume in such a way that the vector space average of two sets is larger than the κ-mean of their respective volumes.

2. By Jensen's inequality μ being κ-concave implies μ is κ'-concave for $\kappa' \leq \kappa$.

A theory is entirely useless without examples. Later we will see that many of the most useful and ubiquitous measures on $\mathbb{R}^n$– including Lebesgue measure, all Gaussian measures, and uniform measures on convex bodies– are convex.

1.2.2 Measurability issues

There are interesting subtleties that require a measurability assumption on $A+B$, although one no longer needs this assumption if A and B are Borel sets. If one is interested only in convex sets, measurability questions are unimportant because all convex sets are measurable (see, e.g., [44]), and convex sets are trivially closed under Minkowski addition.

It is useful in this context to consider several distinguished classes of subsets of $\mathbb{R}^n$, their relationships, and the properties of each class– the usefulness arises from the fact that no one class has all the good properties one would hope for. The classes of interest are the σ-algebra $\mathcal{L}$ of Lebesgue-measurable subsets (or simply measurable sets), the σ-algebra $\mathcal{B}$ of Borel-measurable subsets (or simply Borel sets), the class $\mathcal{A}$ of analytic sets, and the class $\mathcal{U}$ of universally measurable sets, all of whose definitions and properties we will remind the reader of shortly without proof. One has the (proper) inclusions

$$\mathcal{B} \subset \mathcal{A} \subset \mathcal{U} \subset \mathcal{L} \subset 2^{\mathbb{R}}.$$

The Borel σ-algebra $\mathcal{B}$ is the smallest σ-algebra that contains the open sets of $\mathbb{R}^n$. While any measure defined on $\mathcal{B}$ is called a Borel measure, we usually label the choice of Borel measure which assigns $\mu([a_1,b_1] \times \ldots \times [a_n,b_n]) = \prod_{i=1}^n (b_i - a_i)$ for every box as "the Borel measure" on $\mathbb{R}^n$. However, one undesirable fact about the Borel σ-algebra $\mathcal{B}$ is that it is not complete. [Recall that a complete measure space is a measure space in which every subset of every measure-zero set is measurable (having measure zero).] Since one cannot avoid completeness questions when considering, for example, the construction of product measures, it is usual to define a completion procedure by looking at the σ-algebra generated by the sets in $\mathcal{B}$ and the class $\mathcal{N}$ of all subsets of measure-zero sets in $\mathcal{B}$; then every member of the expanded σ-algebra is of the form $A \cup B$ for some $A \in \mathcal{B}$ and some $B \in \mathcal{N}$, and $\mu_0(A \cup B) = \mu(A)$. The completion of the Borel measure is precisely the Lebesgue measure (which implies $\mathcal{B} \subset \mathcal{L}$). Recall that Lebesgue measure is independently defined by a two-step procedure based on C. Carathéodory's extension theorem: first, one defines the outer measure (for all sets) as the smallest sum of box volumes corresponding to covering the set using boxes, and then, one defines a set A to be measurable if for every other set K, the outer measure of K is the sum of outer measures of $K \setminus A$ and $K \cap A$.

A set $A \subset \mathbb{R}^n$ is an *analytic set* if it is the continuous image of a Borel set in a Polish space. Analytic subsets of Polish spaces are closed under countable unions and intersections, continuous images, and inverse images (but not under complementation). Any Borel set is analytic (i.e., $\mathcal{B} \subset \mathcal{A}$).

A subset A of $\mathbb{R}^n$ is *universally measurable* if it is measurable with respect to every complete probability measure on $\mathbb{R}^n$ that measures all Borel sets in $\mathbb{R}^n$. In particular, a universally measurable set of reals is necessarily Lebesgue measurable (i.e., $\mathcal{U} \subset \mathcal{L}$). Every analytic set is universally measurable (i.e., $\mathcal{A} \subset \mathcal{U}$), and hence Lebesgue measurable.

We can now explain why it is sufficient in the definition of convex measures to only require that the sets A and B are Borel-measurable. Indeed, if A and B are Borel, they are analytic; hence $A \times B$ is an analytic set in $\mathbb{R}^n \times \mathbb{R}^n$. Now, by projecting from $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ to $x + y \in \mathbb{R}^n$ (which is a continuous map), the resultant set $A + B$ is analytic. In particular, since $\mathcal{A} \subset \mathcal{L}$, $A + B$ is measurable, and no additional assumption is needed.

1.3 Exogenous motivations for studying convex measures

1.3.1 Geometric functional analysis

A key category in functional analysis is that of Banach spaces. Recall that a (real) Banach space V is a complete normed vector space, where complete means that any Cauchy sequence converges, and a norm $\|\cdot\|$ is defined by the three conditions: $\|x\| \geq 0$, $\|\lambda x\| = |\lambda|\|x\|$ for any $\lambda \in \mathbb{R}$, and $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

Key examples include

$$C(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R} : \ f \text{ continuous}, \|f\|_\infty := \sup_x |f(x)| < \infty\},$$

$$L^p(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R} : \ \|f\|_p := \left(\int_{\mathbb{R}^n} |f(x)|^p dx\right)^{1/p} < \infty\}.$$

An idea pioneered by V. Milman and his contemporaries was to focus on the "local", or "asymptotic" theory of Banach spaces. The approach here is to understand the geometry of infinite-dimensional spaces by studying large but finite-dimensional Banach spaces (i.e., $\mathbb{R}^n$ with some norm, for large n). An immediate connection between Banach space theory and geometry comes from considering unit balls. Let us recall some basic definitions. The *unit ball* of a normed space (say $\mathbb{R}^n$ with some norm) is defined by $\{x \in \mathbb{R}^n : \|x\| \leq 1\}$. We say that a set $A \subset \mathbb{R}^n$ is *symmetric* if $x \in A$ implies that $-x \in A$. A compact convex set with non-empty interior is called a *convex body*.

Theorem 5. *For any Banach space, the unit ball is a symmetric, convex body. Moreover, every symmetric, convex body in $\mathbb{R}^n$ is the unit ball for some norm on $\mathbb{R}^n$; indeed, one can define the Minkowski or gauge functional $\|x\|_K = \inf\{t > 0 : \frac{x}{t} \in K\}$.*

Observe that by definition, $\|x\|_K \leq r$ iff $x/r \in K$ or $x \in rK$.

Let l_p^n denotes the $\mathbb{R}^n$ equipped with l_p norm, $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. Observe the shapes of the unit balls.

As an illustration, a major theorem[7] in this local theory of Banach spaces is Dvoretzky's theorem [20, 21]. For background, note that although all norms on $\mathbb{R}^n$ are equivalent, the constants that appear would typically depend on n; for example, the vector $(1, \ldots, 1)$ has l^∞ norm 1 but l^2 norm $\sqrt{n}$.

Theorem 6. *For any n-dimensional Banach space V with n large enough, there exists $m(n) \to \infty$ as $n \to \infty$, and an $m(n)$ dimensional subspace E of V such that E looks like $l_2^{m(n)}$. More precisely, for every $\epsilon > 0$, there exists an $m(n, \epsilon)$ dimensional subspace E and a positive quadratic form Q on E such that the corresponding Euclidean norm $|\cdot|_2$ satisfies*

$$|x|_2 \leq \|x\| \leq (1 + \epsilon)|x|_2$$

for every $x \in E$.

In fact one can always choose $m(n)$ growing like $\epsilon^2 \log n$; the sharp dependence on ϵ here was proved by Gordon [33] (cf. [60]).

An alternative formulation using convex bodies is: For any convex body in $\mathbb{R}^n$ with n large enough, there exist universal constants c, C, and $m(n) \to \infty$ as $n \to \infty$, and an $m(n)$ dimensional subspace E of $\mathbb{R}^n$ such that

$$\frac{c}{M(K)}(B_2^n \cap E) \subset K \cap E \subset \frac{C}{M(K)}(B_2^n \cap E),$$

where $M(K) = \int_{\mathbb{S}^{n-1}} \|x\| \sigma(dx)$.

Another major theorem about high-dimensional convex bodies was obtained much more recently by Klartag [41, 42]. It says that measure projections (and not just geometric projections) are "nice".

Theorem 7. *Let $\ell(n) = cn^\tau \in [n]$, where c, τ are universal (but unspecified) constants. If X is a random vector in $\mathbb{R}^n$ that is distributed uniformly in a convex body K, then there exists a $\ell(n)$-dimensional subspace $E \subset \mathbb{R}^n$ and $r > 0$ such that*

$$d_{TV}(\mathcal{L}(\pi_E X), N(0, r)) \leq n^{-\tau},$$

where $\pi_E X$ denotes the orthogonal projection of X onto E.

Observe that there is a somewhat surprising distinction between the measure projection and the geometric projection of high-dimensional convex bodies; geometric projections are only nice up to dimension logarithmic in n, while measure projections remain nice up to a fractional power of n.

[7] Here is a nice extract from K. Ball's review of Milman's survey article [51] on Dvoretzky's theorem: "Dvoretzky's theorem initiated an avalanche of work on finite-dimensional normed spaces, guided by the heuristic principle: "All convex bodies behave like ellipsoids, and ever more so as the dimension increases." This principle runs completely counter to one's initial experience. As the dimension n increases, cubes seem less and less like spheres. The ellipsoid of largest volume inside an n-dimensional cube is the obvious solid sphere, and occupies only a tiny fraction, about $(\pi e/2n)^{n/2}$, of the cube's volume. Nevertheless there are remarkable theorems embodying the heuristic principle and providing powerful tools for the study of high-dimensional geometry."

As we will see, convex bodies can be identified with the class of $1/n$-concave probability measures on $\mathbb{R}^n$ (the latter are precisely the uniform measures on the former). Thus one may wonder if the "natural" setting for results in geometric functional analysis (such as Klartag's theorem) is really that of convex measures. That this does indeed seem to be the case has been realized, and there is much active research on completing this bigger picture.

We will develop enough machinery in these lectures to at least understand some of the ingredients that go into proving Theorem 7 and its generalizations to convex measures, which can be seen as a powerful variant of the central limit theorem that imposes the global dependence structure of convexity rather than a coordinate-dependent dependence structure (such as a Markov or martingale structure).

1.3.2 Stochastic optimization

One motivation for the study of convex measures comes from the fact that just as convex (or more generally, quasi-concave) functions are the threshold between easy and difficult optimization problems, convex measures are the threshold between easy and difficult *stochastic* optimization problems. There are several kinds of stochastic optimization problems that can be considered– one example, which was the original motivation for Prékopa's study [58] of log-concave functions, is when the goal is to minimize a convex function $f(x)$ defined on $\mathbb{R}^n$ subject to the constraints

$$\mathbf{P}\{g_1(x) \geq A_1, \ldots, g_m(x) \geq A_m\} \geq p, \quad h_1(x) \geq 0, \ldots, h_k(x) \geq 0,$$

where $p \in (0,1)$ is a prescribed probability, the functions $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$ are concave, and the parameters $A_1, \ldots, A_m$ are random variables. In this case, the problem reduces to a deterministic optimization problem whose nature depends on that of the function

$$h(x) = \mathbf{P}\{g_1(x) \geq A_1, \ldots, g_m(x) \geq A_m\},$$

which in turn depends on the joint distribution of $(A_1, \ldots, A_m) \in \mathbb{R}^m$. As we will see, if $(A_1, \ldots, A_m)$ has a log-concave distribution, then the function $h(x)$ will turn out to be a log-concave function on $\mathbb{R}^n$, which makes the stochastic optimization problem amenable to standard convex programming algorithms.

1.3.3 Statistical inference

Statistical decision theory explores fundamental limits on the performance of statistical estimators, and there is an enormous theory that explores this, both in parametric and nonparametric settings. Non-parametric statistical inference under shape constraints has attracted a huge amount of interest lately [4, 19]. A prototypical example of such an inference problem is to estimate a log-concave (or more generally, s-concave with $s < 0$) density from data drawn from it; it turns out that a maximum likelihood approach can be used despite the infinite dimensionality.

2 Lecture 2: Borell's characterization of convex measures

2.1 Convexity properties of Lebesgue measure

2.1.1 The Brunn-Minkowski inequality

The most basic measure on $\mathbb{R}^n$ is the Lebesgue measure ℓ. Does it have convexity properties? Enter the Brunn-Minkowski inequality[8] (BMI).

Theorem 8. *If A, B are Borel subsets of $\mathbb{R}^d$,*

$$|A+B|^{\frac{1}{d}} \geq |A|^{\frac{1}{d}} + |B|^{\frac{1}{d}}.$$

It is useful to record several equivalent forms of the BMI.

Lemma 2. *The following are equivalent:*

1. *For the class of Borel-measurable subsets of $\mathbb{R}^d$,*

$$|A+B|^{\frac{1}{d}} \geq |A|^{\frac{1}{d}} + |B|^{\frac{1}{d}}$$

2. *ℓ is $1/d$-concave, i.e.,*

$$|\lambda A + (1-\lambda)B| \geq \left(\lambda |A|^{\frac{1}{d}} + (1-\lambda)|B|^{\frac{1}{d}}\right)^d$$

3. *ℓ is 0-concave, i.e.,*

$$|\lambda A + (1-\lambda)B| \geq |A|^{\lambda}|B|^{1-\lambda}$$

4. *ℓ is $(-\infty)$-concave, i.e.,*

$$|\lambda A + (1-\lambda)B| \geq \min\{|A|, |B|\}.$$

Proof. Let us indicate why the inequalities (5)–(5) are equivalent. Making use of the arithmetic mean-geometric mean inequality, we immediately have (5) ⇒ (5) ⇒ (5). Applying (5) to $\tilde{A} = \lambda A$, $\tilde{B} = (1-\lambda)B$ we have

$$\begin{aligned}
|\lambda A + (1-\lambda)B| &= |\tilde{A} + \tilde{B}| \\
&\geq (|\tilde{A}|^{\frac{1}{d}} + |\tilde{B}|^{\frac{1}{d}})^d \\
&= \left(|\lambda A|^{\frac{1}{d}} + |(1-\lambda)B|^{\frac{1}{d}}\right)^d \\
&= \left(\lambda |A|^{\frac{1}{d}} + (1-\lambda)|B|^{\frac{1}{d}}\right)^d,
\end{aligned}$$

where the last equality is by homogeneity of the Lebesgue measure. Thus (5) ⇒ (5). It remains to prove that (5) ⇒ (5). First notice that (5) is equivalent to

$$\begin{aligned}
|A+B| &\geq \min\{|A/\lambda|, |B/(1-\lambda)|\} \\
&= \min\{|A|/\lambda^d, |B|/(1-\lambda)^d\}.
\end{aligned}$$

[8] While Brunn and Minkowski proved the inequality (**??**) for convex sets in the late 19th century, Lusternik [48] extended it to the general case. Therefore, strictly speaking, one should refer to the Brunn-Minkowski inequality only when the objects of interest are convex sets, but in practice, this appellation is commonly used for the general inequality as well.

It is easy to see that the right hand side is maximized when $|A|/\lambda^d = |B|/(1-\lambda)^d$, or

$$\lambda = \frac{|A|^{\frac{1}{d}}}{|A|^{\frac{1}{d}} + |B|^{\frac{1}{d}}}.$$

Inserting λ into the above yields (5). □

There are now numerous proofs of the BMI. A particularly nice one is that of Hadwiger and Ohmann [38] (recounted as part of a beautiful survey of the BMI by Gardner [32]). There are also nice proofs using Steiner symmetrization, rearrangements, optimal transport, etc. (which all need some machinery to explain that we do not want to bother with).

Later we will give an elegant proof of the BMI that goes through its functional version, the Prékopa-Leindler inequality. For now, we merely observe that the BMI in one dimension is trivial.

Lemma 3. *If $A, B \subset \mathbb{R}$ are compact, then $|A+B| \geq |A| + |B|$.*

Proof. By translation-invariance of Lebesgue measure, we can move A and B so that the minimum element of A and the maximum element of B (both of which exist by compactness) are 0. Then $A + B \supset A \cup B$, and so

$$|A+B| \geq |A \cup B| = |A| + |B| - |A \cap B| = |A| + |B| - |\{0\}| = |A| + |B|.$$

□

2.1.2 Isoperimetric Inequality for Lebesgue measure

Define the surface area of a set A (if it exists) as

$$S(A) = \lim_{\epsilon \to 0} \frac{\mathrm{Vol}(A + \epsilon B_2^n) - \mathrm{Vol}(A)}{\epsilon}.$$

Here B_2^n is the unit ball in $\mathbb{R}^n$ using the Euclidean metric. Observe that another way to write the enlargement $A^\epsilon = A + \epsilon B_2^n$ is as $\{x \in \mathbb{R}^n : d(x, A) \leq \epsilon\}$, where $d(x, A) = \inf\{d(x, y) : y \in A\}$ is the distance from x to the set A.

The classical isoperimetric principle has a very long history, and underlies for example the fact that water bubbles are spherical (since surface tension causes them to minimize surface area for the fixed volume of trapped air). Two ways to state the isoperimetric principle for Lebesgue measure on $\mathbb{R}^n$:

1. For fixed $\mathrm{Vol}(A)$, the surface area of A, given by $S(A)$ is minimized when A is a Euclidean ball; this can be restated as $S(A) \geq S(A^*)$, where A^* is the spherical symmetrization of A.

2. For fixed surface area $S(A)$, the volume is maximized when A is a Euclidean ball.

More precisely, we capture this by proving that the functional $\mathrm{Vol}(A)/S(A)^{\frac{n}{n-1}}$ is maximized by the ball.

Theorem 9. *For sets $A \subset \mathbb{R}^n$ whose volume and surface area exist,*

$$\left[\frac{Vol(A)}{Vol(B_2)}\right]^{1/n} \leq \left[\frac{S(A)}{S(B_2)}\right]^{1/n-1}.$$

Proof. We show that the Brunn-Minkowski inequality implies the Isoperimetric inequality for the Lebesgue measure. Recall that the Brunn-Minkowski inequality states that, if A and B are two non-empty compact subsets of $\mathbb{R}^n$, then,

$$\mathrm{Vol}^{1/n}(A+B) \geq \mathrm{Vol}^{1/n}(A) + \mathrm{Vol}^{1/n}(B).$$

From the above we get that for any $s,\, t > 0$

$$\mathrm{Vol}^{1/n}(sA+tB) \geq s\mathrm{Vol}^{1/n}(A) + t\mathrm{Vol}^{1/n}(B).$$

The above immediately seen by noting that $\mathrm{Vol}^{1/n}(sA) = s\mathrm{Vol}^{1/n}(A)$. From this we get that

$$\begin{aligned} S(A) &= \lim_{\epsilon\to 0} \frac{\mathrm{Vol}(A+\epsilon B_2) - \mathrm{Vol}(A)}{\epsilon} \\ &\geq \lim_{\epsilon\to 0} \frac{\left[\mathrm{Vol}^{1/n}(A) + \epsilon\mathrm{Vol}^{1/n}(B_2)\right]^n - \mathrm{Vol}(A)}{\epsilon} \qquad (5) \\ &= n\mathrm{Vol}^{(n-1)/n}(A)\mathrm{Vol}^{1/n}(B_2) \qquad (6) \end{aligned}$$

Here (5) follows from the form of Brunn-Minskowski given above. Further (6) follows by noting that the limit is $\mathrm{Vol}^{1/n}(B_2)$ times the derivative of the function x^n evaluated at $\mathrm{Vol}^{1/n}(A)$. Accordingly, using that $S(B_2) = n\mathrm{Vol}(B_2)$ one gets that the right side of (6) is $\mathrm{Vol}^{\frac{n-1}{n}}(A)S(B_2)/\mathrm{Vol}^{(n-1)/n}(B_2)$. Rearranging terms leads to us to the Isoperimetric inequality. $\square$

A particularly powerful proof approach, however, which also has many other applications is to start with the more general Brunn-Minkowski inequality.

2.2 The characterization (without proofs)

It is a nontrivial fact, due to Borell [14], that concavity properties of a measure can equivalently be described pointwise in terms of its density (for $\kappa = 0$, this was understood slightly earlier by Prékopa). First, one has to show that convex measures have densities in the first place, with respect to Lebesgue measure on some affine subspace/linear variety.

Any hyperbolic measure μ on a locally convex space E has a convex support $\mathrm{supp}(\mu)$. Let H_μ denote the smallest affine subspace in E containing $\mathrm{supp}(\mu)$. The dimension $k = \dim(H_\mu)$ is called the dimension of μ and is denoted $\dim(\mu)$.

Theorem 10 (Borell)**.** *If a locally finite hyperbolic measure μ on E has a finite dimension, it is absolutely continuous with respect to Lebesgue measure on H_μ.*

The proof of Theorem 10 involves some careful analysis and we will not reproduce it here. The main point of it is that, without loss of generality, we can restrict our attention to the relevant affine subspace, and assume our convex measure has a density.

Theorem 11 (Borell)**.** *A measure μ on $\mathbb{R}^d$ is κ-concave and absolutely continuous with respect to the Lebesgue measure if and only if it has a density that is a $s_{\kappa,d}$-concave function, in the sense that*

$$f((1-t)x+ty) \geq M^t_{s_{\kappa,d}}(f(x), f(y))$$

whenever $f(x)f(y) > 0$ and $t \in (0,1)$, and where

$$s_{\kappa,d} := \frac{\kappa}{1-\kappa d}.$$

Examples:

1. If X is the uniform distribution on a convex body K, it has an ∞-concave density function $f = |K|^{-1}\mathbb{1}_K$ and thus the probability measure is $1/d$-concave. Let us note that by our requirement that μ is "full-dimensional" (i.e., has support with nonempty interior), the only $1/d$-concave probability measures on $\mathbb{R}^d$ are of this type.

2. A measure that is 0-concave is also called a *log-concave measure.* Since $s_{0,d} = 0$ for any positive integer d, Theorem 11 implies that an absolutely continuous measure μ is log-concave if and only if its density is a log-concave function (as defined in Definition **??**). In other words, X has a log-concave distribution if and only if its density function can be expressed on its support as $e^{-V(x)}$ for V convex. When $V(x) = \frac{1}{2}|x|^2 - \frac{d}{2}\log(2\pi)$, one has the standard Gaussian distribution; when $V(x) = x$ for $x \geq 0$ and $V(x) = \infty$ for $x < 0$, one has the standard exponential distribution; and so on.

3. If X is log-normal distribution with density function
$$f(x) := \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$
Then for $\sigma < 4$, the density function of X is $-\sigma/4$-concave, and the probability measure is $\frac{-\sigma}{4-\sigma}$-concave.

4. If X is a Beta distribution with density function
$$\frac{x^\alpha(1-x)^\beta}{B(\alpha,\beta)}$$
with shape parameters $\alpha \geq 1$ and $\beta \geq 1$, then the density function of X is $\min(\frac{1}{\alpha-1}, \frac{1}{\beta-1})$-concave, and the probability measure is $\frac{1}{\max(\alpha,\beta)}$-concave.

5. If X is a d-dimensional Student's t-distribution with density function
$$f(x) := \frac{\Gamma(\frac{\nu+d}{2})}{\nu^{\frac{d}{2}}\pi^{\frac{d}{2}}\Gamma(\frac{\nu}{2})}\left(1 + \frac{|x|^2}{\nu}\right)^{-\frac{\nu+d}{2}}$$
with $\nu > 0$, then the density function of X is $-\frac{1}{\nu+d}$-concave, and the probability measure is $-\frac{1}{\nu}$-concave.

6. If X is a d-dimensional Pareto distribution of the first kind with density function
$$f(x) := a(a+1)\cdots(a+d-1)\left(\prod_{i=1}^d \theta_i\right)^{-1}\left(\sum_{i=1}^d \frac{x_i}{\theta_i} - d + 1\right)^{-(a+d)}$$
for $x_i > \theta_i > 0$ with $a > 0$, then the density function of X is $-\frac{1}{a+d}$-concave, and the probability measure is $1/a$-concave.

The optimal κ for the distributions above can be found through direct computation on densities.

Observe that if $\kappa > \frac{1}{n}$, a κ-concave measure cannot be absolutely continuous with respect to Lebesgue measure.

Observe that $\kappa = -\infty$ corresponds to $s = -\frac{1}{n}$, which means that s-concave densities with $s < -\frac{1}{n}$ may not induce convex measures. In particular, the theory of convex measures does not currently cover quasiconcave densities (except in dimension 1, where we will see that a stronger form of the characterization theorem holds).

2.3 Consequences of Borell's characterization theorem

Borell's characterization theorem is a rather powerful tool, especially in conjunction with the following elementary observation.

Lemma 4. *For a measure on $\mathbb{R}^n$, κ-concavity is an affine invariant. For probability measures, this can be stated as follows: If X is κ-concave and T is affine, then TX is κ-concave as well.*

Proof. Obvious once we notice that $T^{-1}(\lambda A + (1-\lambda)B) = \lambda T^{-1}A + (1-\lambda)T^{-1}B$. $\square$

We now list some corollaries.

Corollary 1 (Brunn's section theorem)**.** *Let K be a convex body in $\mathbb{R}^n$, and H be a hyperplane passing through 0. If $g(t)$ represents the volume of the cross-section of K by the hyperplane $t+H$, then*

$$g(s+t)^{\frac{1}{n-1}} \geq g(s)^{\frac{1}{n-1}} + g(t)^{\frac{1}{n-1}}.$$

In particular, if K is also symmetric, the central section is the largest.

Proof. Since X is $1/n$-concave, so is the real-valued random variable $\langle \theta, X\rangle$, which means its density is $s_{1/n,1} = \frac{1}{n-1}$-concave. But the value of the density of $\langle \theta, X\rangle$ at t is just $\mathrm{vol}_{n-1}(K \cap (t+H))$. $\square$

More generally, we have:

Corollary 2 (Convexity of marginals)**.** *Let $f : \mathbb{R}^n \to [0,\infty)$ be an integrable s-concave function on $\mathbb{R}^n$. Let E be a subspace of $\mathbb{R}^n$ of dimension k, and define, for $t \in E$,*

$$f_E(t) = \int_{E^\perp} f(t+y)dy.$$

Then f_E is a $\frac{s}{1+s(n-k)}$-concave function.

In particular, marginals of a log-concave density are log-concave.

Proof. Note that f induces a $\frac{s}{1+ns}$-concave measure μ, and f_E is the density of the projection of $X \sim \mu$ onto E. Hence f_E is $s_{\frac{s}{1+ns},k} = \frac{s}{1+s(n-k)}$-concave. $\square$

Corollary 3. *If f is a log-concave density on $\mathbb{R}$, then its cumulative distribution function is also log-concave.*

Proof. Since f is log-concave, so is the induced measure μ. The CDF is defined by $F(x) = \mu(-\infty, x] = \mu(A_x)$ where $A_x = (-\infty, x]$. By log-concavity of the measure,

$$F(\lambda x + (1-\lambda)y) = \mu(A_{\lambda x+(1-\lambda)y}) = \mu(\lambda A_x + (1-\lambda)A_y) \geq \mu(A_x)^\lambda \mu(A_y)^{1-\lambda} = F(x)^\lambda F(y)^{1-\lambda}.$$

$\square$

Lemma 5. *Under the condition*

$$s_1 + s_2 > 0, \qquad \frac{1}{s} = \frac{1}{s_1} + \frac{1}{s_2}, \tag{7}$$

for all real positive numbers a', a'', b, b'' and any $\lambda \in (0,1)$,

$$M_{s_1}^\lambda(a',b')\, M_{s_2}^\lambda(a'',b'') \;\geq\; M_s^\lambda(a'a'', b'b'').$$

Proof. Hölder. $\square$

Lemma 6. *If a finite measure μ is κ'-concave on $\mathbb{R}^{n'}$ and a finite measure ν is κ''-concave on $\mathbb{R}^{n''}$, and if $\kappa', \kappa'' \in (-\infty, +\infty]$ satisfy (7), then the product measure $\lambda = \mu \otimes \nu$ is κ-concave on $\mathbb{R}^{n'+n''}$.*

Proof. By Lemma 5, if $A = A' \otimes A''$ and $B = B' \otimes B''$ with standard parallelepipeds A', B' in $\mathbb{R}^{n'}$ of positive μ-measure, and with standard parallelepipeds A'', B'' in $\mathbb{R}^{n''}$ of positive ν-measure, then $tA + sB = (tA' + sB') \times (tA'' + sB'')$ and

$$\begin{aligned}
\lambda(tA + sB) &= \mu(tA' + sB')\, \nu(tA'' + sB'') \\
&\geq M_{\kappa'}^{(t)}(\mu(A'), \mu(B'))\, M_{\kappa''}^{(t)}(\nu(A''), \nu(B'')) \\
&= M_{\kappa}^{(t)}(\mu(A')\nu(A''), \mu(B')\nu(B'')) \;=\; M_{\kappa}^{(t)}(\lambda(A), \lambda(B)).
\end{aligned}$$

That is, the Brunn-Minkowski-type inequality (4) is fulfilled for the measure λ on $\mathbb{R}^n$ in the class of all standard parallelepipeds (of positive measure), which implies the κ-concavity of λ. $\square$

Since all (affine) projections of κ-concave measures are κ-concave, we obtain from Lemma 6 the following corollary (cf. [14, Theorem 4.5], where it is however assumed additionally that $0 < \kappa' < 1/n$ and $0 < \kappa'' < 1/n$).

Corollary 4. *Under the condition (7), if a finite measure μ is κ'-concave on $\mathbb{R}^n$ and a finite measure ν is κ''-concave on $\mathbb{R}^n$, then their convolution $\mu * \nu$ is κ-concave.*

Indeed, $\mu * \nu$ represents the image of the product measure $\lambda = \mu \otimes \nu$ on $\mathbb{R}^n \times \mathbb{R}^n$ under the linear map $(x, y) \to x + y$.

Finally, let us deduce an Anderson-type inequality for convex measures. Whenever we talk about a symmetric set or measure, we always mean centrally symmetric.

Theorem 12. *If C is a symmetric convex set in $\mathbb{R}^n$ and μ is a symmetric convex measure, then*

$$\mu(C) \geq \mu(C + x),$$

for any $x \in \mathbb{R}^n$. In fact, for any fixed $x \in \mathbb{R}^n$, the function $t \mapsto \mu(C + tx)$ is non-increasing for $t \in [0, \infty)$.

Proof. Just take average of $C + x$ and $C - x$ for symmetric convex C, and apply definition of convex measure. $\square$

2.4 The Prékopa-Leindler inequality

The Prékopa-Leindler inequality (PLI) [58, 45, 59] states:

Theorem 13. *If $f, g, h : \mathbb{R}^d \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0, 1)$,*

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

for every $x, y \in \mathbb{R}^d$, then

$$\int h \geq \left(\int f\right)^{\lambda} \left(\int g\right)^{1-\lambda}. \tag{8}$$

Remarks:

1. To see the connection with the BMI, one simply has to observe that $f = 1_A, g = 1_B$ and $h = 1_{\lambda A+(1-\lambda)B}$ satisfy the hypothesis, and in this case, the conclusion is precisely the BMI in its "geometric mean" form $|\lambda A + (1-\lambda)B| \geq |A|^\lambda |B|^{1-\lambda}$.

2. It is natural to try to formulate a "best possible" version of the PLI, in the sense of using the h that gives the tightest bound. However, if we define $h(z) = \sup\{f^\lambda(x)g^{1-\lambda}(y) : \lambda x + (1-\lambda)y = z\}$, we end up with a serious measurability problem. One way to get around this is to define
$$h^*(z) = \operatorname{ess\,sup}\{f^\lambda(x)g^{1-\lambda}(y) : \lambda x + (1-\lambda)y = z\},$$
which is always measurable. (Another way is to use the upper integral to write the conclusion.)

3. It can be seen as a reversal of Hölder's inequality, which can be written as
$$\left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda} \geq \int f^\lambda g^{1-\lambda}.$$

4. If one prefers, the PLI can also be written more explicitly as a kind of convolution inequality, as implictly observed in [16] and explicitly in [43]. Indeed, if one defines the Asplund product of two nonnegative functions by
$$(f \star g)(x) = \sup_{x_1+x_2=x} f(x_1)g(x_2),$$
and the scaling $(\lambda \cdot f)(x) = f^\lambda(x/\lambda)$, then the left side of (26) can be replaced by the integral of $[\lambda \cdot f] \star [(1-\lambda) \cdot g]$.

In dimension 1, we can prove a stronger inequality under an additional condition.

Lemma 7. *If $f, g, h : \mathbb{R} \to [0, \infty)$ satisfy*
$$h(\lambda x + (1-\lambda)y) \geq \min\{f(x), g(y)\}$$
for each $x, y \in \mathbb{R}$, and $\sup f = \sup g$, then
$$\int h \geq \lambda \int f + (1-\lambda) \int g.$$

Proof. Use 1-d BMI and super level sets. □

Let us observe that the PLI in dimension 1 follows from this, since the PLI is homogenous, which means the assumption $\sup f = \sup g$ can be made without loss of generality.

2.5 Proof of the characterization

More generally, one has the Borell-Brascamp-Lieb inequality (BBLI).

Theorem 14. *Let $\gamma \geq -\frac{1}{n}$, $\lambda \in [0,1]$ and $f, g, h : \mathbb{R}^n \to [0, +\infty)$ be measurable functions. If the inequality*

$$h((1-\lambda)x + \lambda y) \geq M_\gamma^\lambda(f(x), g(y))$$

holds for every $x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)$, then

$$\int_{\mathbb{R}^n} h \geq M^\lambda_{\frac{\gamma}{1+\gamma n}} \left(\int_{\mathbb{R}^n} f, \int_{\mathbb{R}^n} g \right).$$

Proof. We already have a result of this type when $n = 1$.

The induction step follows from what is called the "tensorization property" of the BBLI. Indeed, suppose f, g, h are functions on $\mathbb{R}^{n+1}$ with F, G, H. TO BE ADDED... □

Remark: Tensorization works for PLI but not BMI. This is related to the fact that marginals of indicators are not indicators; it is key to move to a larger more flexible class of objects. On the other hand, we cannot have an AM form of PLI in general dimension. This is because going from indicators to nonnegative measurable functions (via indicators, in principle) allows only GM form to persist.

3 Lecture 3: Reverse Hölder inequalities and information concentration

3.1 Motivations

3.1.1 The Gaussian thin shell property and its possible generalizations

Consider a random vector Z taking values in $\mathbb{R}^n$, drawn from the standard Gaussian distribution γ, whose density is given by

$$\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$$

for each $x \in \mathbb{R}^n$, where $|\cdot|$ denotes the Euclidean norm. It is well known that when the dimension n is large, the distribution of Z is highly concentrated around the sphere of radius $\sqrt{n}$; that $\sqrt{n}$ is the appropriate radius follows by the trivial observation that $\mathbb{E}|Z|^2 = \sum_{i=1}^n \mathbb{E}Z_i^2 = n$. One way to express this concentration property is by computing the variance of $|Z|^2$, which is easy to do using the independence of the coordinates of Z:

$$\mathrm{Var}(|Z|^2) = \mathrm{Var}\bigg(\sum_{i=1}^n Z_i^2\bigg) = \sum_{i=1}^n \mathrm{Var}(Z_i^2) = 2n.$$

In particular, the standard deviation of $|Z|^2$ is $\sqrt{2n}$, which is much smaller than the mean n of $|Z|^2$ when n is large. Another way to express this concentration property is through a deviation inequality:

$$\mathbf{P}\bigg\{\frac{|Z|^2}{n} - 1 > t\bigg\} \leq \exp\bigg\{-\frac{n}{2}[t - \log(1+t)]\bigg\} \tag{9}$$

for the upper tail, and a corresponding upper bound on the lower tail. These inequalities immediately follow from Chernoff's bound, since $|Z|^2/n$ is just the empirical mean of i.i.d. random variables.

It is natural to wonder if, like so many other facts about Gaussian measures, the above concentration property also has an extension to log-concave measures (or to some subclass of them). There are two ways one may think about extending the above concentration property. One is to ask if there is a universal constant C such that

$$\mathrm{Var}(|X|^2) \leq Cn,$$

for every random vector X that has an isotropic, log-concave distribution on $\mathbb{R}^n$. Here, we say that a distribution on $\mathbb{R}^n$ is isotropic if its covariance matrix is the identity matrix; this assumption ensures that $\mathbb{E}|X|^2 = n$, and provides the normalization needed to make the question meaningful. This question has been well studied in the literature, and is known as the "thin shell conjecture" in convex geometry. It is closely related to other famous conjectures: it implies the hyperplane conjecture of Bourgain [23, 24], is trivially implied by the Kannan-Lovasz-Simonovits conjecture, and also implies the Kannan-Lovasz-Simonovits conjecture up to logarithmic terms [22]. The best bounds known to date are those of Guédon and E. Milman [35], and assert that

$$\mathrm{Var}(|X|^2) \leq Cn^{4/3}.$$

The second way that one may try to extend the above concentration property from Gaussians to log-concave measures is to first observe that the quantity that concentrates, namely $|Z|^2$, is essentially the logarithm of the Gaussian density function. More precisely, since

$$-\log \phi(x) = \frac{n}{2}\log(2\pi) + \frac{|x|^2}{2},$$

the concentration of $|Z|^2$ about its mean is equivalent to the concentration of $-\log \phi(Z)$ about its mean. Thus one can ask if, for every random vector X that has a log-concave density f on $\mathbb{R}^n$,

$$\mathrm{Var}(-\log f(X)) \leq Cn \tag{10}$$

for some absolute constant C. An affirmative answer to this question was provided by Bobkov and Madiman [8]. The approach of [8] can be used to obtain bounds on C, but the bounds so obtained are quite suboptimal (around 1000). Recently V. H. Nguyen [53] (see also [54]) and L. Wang [65] independently determined, in their respective Ph.D. theses, that the sharp constant C in the bound (10) is 1. Soon after this work, simpler proofs of the sharp variance bound were obtained independently by us (presented in the proof of Theorem 16 in this paper) and by Bolley, Gentil and Guillin [12] (see Remark 4.2 in their paper). An advantage of our proof over the others mentioned is that it is very short and straightforward, and emerges as a consequence of a more basic log-concavity property (namely Theorem 15) of L^p-norms of log-concave functions, which may be thought of as an analogue for log-concave functions of a classical inequality of Borell [13] for concave functions.

If we are interested in finer control of the integrability of $-\log f(X)$, we may wish to consider analogues for general log-concave distributions of the inequality (9). Our second objective in this note is to provide such an analogue (in Theorem 18). A weak version of such a statement was announced in [9] and proved in [8], but the bounds we provide in this note are much stronger. Our approach has two key advantages: first, the proof is transparent and completely avoids the use of the sophisticated Lovasz-Simonovits localization lemma, which is a key ingredient of the approach in [8]; and second, our bounds on the moment generating function are sharp, and are attained for example when the distribution under consideration has i.i.d. exponentially distributed marginals.

While in general exponential deviation inequalities imply variance bounds, the reverse is not true. Nonetheless, our approach in this note is to first prove the variance bound (10), and then use a general bootstrapping result (Theorem 17) to deduce the exponential deviation inequalities from it. The bootstrapping result is of independent interest; it relies on a technical condition that turns out to be automatically satisfied when the distribution in question is log-concave.

3.1.2 The Shannon-McMillan-Breiman theorem and its implications

Let $\mathbb{X} = (X_1, X_2, \ldots)$ be a stochastic process with each X_i taking values on the real line $\mathbb{R}$. Suppose that the joint distribution of $X^n = (X_1, \ldots, X_n)$ has a density f with respect to either Lebesgue measure on $\mathbb{R}^n$. We are interested in the random variable

$$\tilde{h}(X^n) = -\log f(X^n).$$

In the discrete case, the quantity $\tilde{h}(X^n)$ (using f for the probability mass function in this case, thought of as the density with respect to counting measure on some discrete subset of $\mathbb{R}^n$) is essentially the number of bits needed to represent X by a coding scheme that minimizes average code length (cf. [62]), and therefore may be thought of as the (random) *information content* of X^n. Such an interpretation is not justified in the continuous case, but the quantity $\tilde{h}(X^n)$ remains of central interest in information theory, statistical physics, and statistics, and so we will with some abuse of terminology continue to call it the information content. Its importance in information theory comes from the fact that it is the building block for Pinsker's information density; its importance in statistical physics comes from the fact that it represents (up to an additive constant involving the logarithm of the partition function) the Hamiltonian or energy of a physical system under a Gibbs measure; and its importance in statistics comes from the fact that it represents the log-likelihood function in the nonparametric inference problem of density estimation.

The average value of the information content of X is the differential entropy. Indeed, the entropy of X^n is defined by

$$h(X^n) = -\int f(x) \log f(x)\, dx = -\,\mathbb{E} \log f(X^n),$$

when it exists. If the limit

$$h(\mathbb{X}) := \lim_{n\to\infty} \frac{h(X^n)}{n}$$

exists, it is called the entropy rate of the process $\mathbb{X}$.

The Shannon-McMillan-Breiman theorem [62, 49, 17] is a central result of information theory; indeed, an early form of this result was called by McMillan the "fundamental theorem of information theory" [49]. (McMillan also gave it the pithy and expressive title of the "Asymptotic Equipartition Property".) It asserts that for any stationary, ergodic process whose entropy rate exists, the information content per coordinate converges almost surely to the entropy rate. This version, and a generalization involving log likelihood ratios with respect to a Markov process, is due independently to Barron [5] and Orey [57]; the definitive version for asymptotically mean stationary processes is due to [5], and Algoet and Cover [1] give an elementary approach to it. The theorem implies in particular that for purposes of coding discrete data from ergodic sources, it is sufficient to consider "typical sequences", and that the entropy rate of the process plays an essential role in characterizing fundamental limits of compressing data from such sources.

3.2 A basic reverse Hölder inequality

Hölder's inequality implies the following basic fact about L^p-norms (on any measure space), which is sometimes attributed to Lyapunov: *For a nonnegative function f, the function $p \mapsto \int f^p$ is log-convex for $p \in [0,\infty]$.* Indeed, Hölder's inequality states that if f and g are measurable functions from S to $\mathbb{R}$, then

$$\|fg\|_{L^1(S,\mu)} \le \|f\|_{L^r(S,\mu)} \|g\|_{L^{r'}(S,\mu)},$$

for any $1 \le r \le \infty$ with the conjugate number r' to r being defined by $\frac{1}{r} + \frac{1}{r'} = 1$. Therefore, writing $\int f^p$ for $\int_S |f(x)|^p\, \mu(dx)$, we have

$$\int f^{\lambda p + (1-\lambda) q} = \int f^{\lambda p} f^{(1-\lambda) q} \le \left[\int f^{\lambda p r}\right]^{\frac{1}{r}} \left[\int f^{(1-\lambda) q r'}\right]^{\frac{1}{r'}} = \left[\int f^p\right]^{\lambda} \left[\int f^q\right]^{1-\lambda},$$

by choosing $r = 1/\lambda$, which is precisely the claimed convexity of $\log \int f^p$ with respect to $p \in (0,\infty)$.

The key fact in this section is the following theorem, which clearly can be seen as a describing behavior that is in the opposite direction to Lyapunov's observation.

Theorem 15. *If f is log-concave on $\mathbb{R}^n$, then the function*

$$G(\alpha) := \alpha^n \int f(x)^\alpha dx$$

is log-concave on $(0, +\infty)$.

Proof. Write $f = e^{-U}$, with U convex. Make the change of variable $x = z/\alpha$ to get

$$G(\alpha) = \int e^{-\alpha U(z/\alpha)} dz.$$

The function $w(z, \alpha) := \alpha U(z/\alpha)$ is convex on $\mathbb{R}^n \times (0, +\infty)$ by Lemma 8, which means that the integrand above is log-concave. The log-concavity of G then follows from Prékopa's theorem [59], which implies that marginals of log-concave functions are log-concave. $\square$

We used the following lemma, which is a standard fact about the so-called perspective function in convex analysis. We give the short proof for completeness.

Lemma 8. *If $U : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function, then*

$$w(z, \alpha) := \alpha U(z/\alpha)$$

is a convex function on $\mathbb{R}^n \times (0, +\infty)$.

Proof. First note that by definition, $w(az, a\alpha) = aw(z, \alpha)$ for any $a > 0$ and any $(z, \alpha) \in \mathbb{R}^n \times (0, +\infty)$, which implies in particular that

$$\frac{1}{\alpha} w(z, \alpha) = w\left(\frac{z}{\alpha}, 1\right).$$

Hence

$$\begin{aligned}
& w(\lambda z_1 + (1-\lambda) z_2, \lambda\alpha_1 + (1-\lambda)\alpha_2) \\
& = [\lambda\alpha_1 + (1-\lambda)\alpha_2]\, U\left(\frac{\lambda\alpha_1 \frac{z_1}{\alpha_1} + (1-\lambda)\alpha_2 \frac{z_2}{\alpha_2}}{\lambda\alpha_1 + (1-\lambda)\alpha_2}\right) \\
& \leq \lambda\alpha_1 U\left(\frac{z_1}{\alpha_1}\right) + (1-\lambda)\alpha_2 U\left(\frac{z_2}{\alpha_2}\right) \\
& = \lambda w(z_1, \alpha_1) + (1-\lambda) w(z_2, \alpha_2),
\end{aligned}$$

for any $\lambda \in [0, 1]$, $z_1, z_2 \in \mathbb{R}^n$, and $\alpha_1, \alpha_2 \in (0, \infty)$. $\square$

3.3 Sharp varentropy bound

Before we proceed, we need to fix some definitions and notation.

Definition 3. *Let a random vector X taking values in $\mathbb{R}^n$ have probability density function f. The information content of X is the random variable $\tilde{h}(X) = -\log f(X)$. The entropy of X is defined as $h(X) = \mathbb{E}(\tilde{h}(X))$. The varentropy of a random vector X is defined as $V(X) = \mathrm{Var}(\tilde{h}(X))$.*

Note that the entropy and varentropy depend not on the realization of X but only on its density f, whereas the information content does indeed depend on the realization of X. For instance, one can write $h(X) = -\int_{\mathbb{R}^n} f \log f$ and

$$V(X) = \mathrm{Var}(\log f(X)) = \int_{\mathbb{R}^n} f(\log f)^2 - \left(\int_{\mathbb{R}^n} f \log f \right)^2.$$

Nonetheless, for reasons of convenience and in keeping with historical convention, we slightly abuse notation as above.

As observed in [8], the distribution of the difference $\tilde{h}(X) - h(X)$ is invariant under any affine transformation of $\mathbb{R}^n$ (i.e., $\tilde{h}(TX) - h(TX) = \tilde{h}(X) - h(X)$ for all invertible affine maps $T : \mathbb{R}^n \to \mathbb{R}^n$); hence the varentropy $V(X)$ is affine-invariant while the entropy $h(X)$ is not.

Another invariance for both $h(X)$ and $V(X)$ follows from the fact that they only depend on the distribution of $\log(f(X))$, so that they are unchanged if f is modified in such a way that its sublevel sets keep the same volume. This implies (see, e.g., [47, Theorem 1.13]) that if $\tilde{f}$ is the spherically symmetric, decreasing rearrangement of f, and $\tilde{X}$ is distributed according to the density $\tilde{f}$, then $h(X) = h(\tilde{X})$ and $V(X) = V(\tilde{X})$. The rearrangement-invariance of entropy was a key element in the development of refined entropy power inequalities in [66].

We can now state the optimal form of the inequality (10), first obtained by Nguyen [53] and Wang [65] as discussed in Section **??**.

Theorem 16. *[53, 65] Given a random vector X in $\mathbb{R}^n$ with log-concave density f,*

$$V(X) \le n$$

Remarks:

1. The probability bound *does not depend* on f– it is universal over the class of log-concave densities.

2. The bound is *sharp*. Indeed, let X have density $f = e^{-\varphi}$, with $\varphi : \mathbb{R}^n \to [0, \infty]$ being positively homogeneous of degree 1, i.e., such that $\varphi(tx) = t\varphi(x)$ for all $t > 0$ and all $x \in \mathbb{R}^n$. Then one can check that the random variable $Y = \varphi(X)$ has a gamma distribution with shape parameter n and scale parameter 1, i.e., it is distributed according to the density given by

$$f_Y(t) = \frac{t^{n-1} e^{-t}}{(n-1)!}.$$

To see this, note that

$$\begin{aligned} P(Y \ge t) = P(\varphi(X) \ge t) &= \int_{\{x; \varphi(x) \ge t\}} e^{-\varphi(x)} dx \\ &= \int_{\{x; \varphi(x) \ge t\}} \int_{\varphi(x)}^{+\infty} e^{-s} ds \\ &= \int_t^{+\infty} e^{-s} vol(\{x; t \le \varphi(x) \le s\}) ds. \end{aligned}$$

Using homogeneity of φ, we deduce that

$$P(Y \ge t) = \int_t^{+\infty} e^{-s} (s^n - t^n) \mathrm{vol}(\{x; \varphi(x) \le 1\}) ds.$$

Integrating by parts, we get

$$P(Y \geq t) = \int_t^{+\infty} e^{-s} n s^{n-1} ds \mathrm{vol}(\{x; \varphi(x) \leq 1\}).$$

Since $Y \geq 0$, taking $t = 0$, we get that

$$1 = P(Y \geq 0) = \int_0^{+\infty} e^{-s} n s^{n-1} ds \mathrm{vol}(\{x; \varphi(x) \leq 1\}) = n! \mathrm{vol}(\{x; \varphi(x) \leq 1\}).$$

Thus

$$P(Y \geq t) = \int_t^{+\infty} e^{-s} s^{n-1} ds/(n-1)!.$$

Consequently $\mathbb{E}(Y) = n$ and $\mathbb{E}(Y^2) = n(n+1)$, and therefore $V(X) = \mathrm{Var}(Y) = n$. Particular examples of equality include:

(a) The case where $\varphi(x) = \sum_{i=1}^n x_i$ on the cone of points with non-negative coordinates (which corresponds to X having i.i.d. coordinates with the standard exponential distribution), and

(b) The case where $\varphi(x) = \inf\{r > 0 : x \in rK\}$ for some compact convex set K containing the origin (which, by taking K to be a symmetric convex body, includes all norms on $\mathbb{R}^n$ suitably normalized so that $e^{-\varphi}$ is a density).

3. Bolley, Gentil and Guillin [12] in fact prove a stronger inequality, namely,

$$\frac{1}{V(X)} - \frac{1}{n} \geq \left[\mathbb{E}\{\nabla U(X) \cdot \mathrm{Hess}(U(X))^{-1} \nabla U(X)\} \right]^{-1}.$$

This gives a strict improvement of Theorem 16 when the density $f = e^{-U}$ of X is strictly log-concave, in the sense that $\mathrm{Hess}(U(X))$ is, almost surely, strictly positive definite. As noted by [12], one may give another alternative proof of Theorem 16 by applying a result of Hargé [39, Theorem 2].

Proof. Since f is a log-concave density, it necessarily holds that $f \in L^\alpha(\mathbb{R}^n)$ for every $\alpha > 0$; in particular, $G(\alpha) := \alpha^n \int f^\alpha$ is finite and infinitely differentiable on the domain $(0, \infty)$. By definition,

$$\log G(\alpha) = n \log \alpha + \log \int f^\alpha = n \log \alpha + F(\alpha).$$

Consequently,

$$\frac{d^2}{d\alpha^2}[\log G(\alpha)] = -\frac{n}{\alpha^2} + F''(\alpha).$$

By Theorem 15, $\log G(\alpha)$ is concave, and hence we must have that

$$-\frac{n}{\alpha^2} + F''(\alpha) \leq 0$$

for each $\alpha > 0$. However, Lemma 9 implies that $F''(\alpha) = V(X_\alpha)/\alpha^2$, so that we obtain the inequality

$$\frac{V(X_\alpha) - n}{\alpha^2} \leq 0.$$

For $\alpha = 1$, this implies that $V(X) \leq n$. □

Notice that if $f = e^{-U}$, where $U : \mathbb{R}^n \to [0,\infty]$ is positively homogeneous of degree 1, then the same change of variable as in the proof of Theorem 15 shows that

$$G(\alpha) = \int e^{-\alpha U(z/\alpha)} dz = \int e^{-U(z)} dz = \int f(z) dz = 1.$$

Hence the function G is constant. Then the proof above shows that $V(X) = n$, which establishes the equality case stated in Remark **??**.

Our proof of Theorem 16 used the following lemma, which is a standard computation (it is a special case of a well known fact about exponential families in statistics), but we write out a proof for completeness.

Lemma 9. *Let f be any probability density function on $\mathbb{R}^n$ such that $f \in L^\alpha(\mathbb{R}^n)$ for each $\alpha > 0$, and define*

$$F(\alpha) = \log \int_{\mathbb{R}^n} f^\alpha.$$

Let X_α be a random variable with density f_α on $\mathbb{R}^n$, where

$$f_\alpha := \frac{f^\alpha}{\int_{\mathbb{R}^n} f^\alpha}.$$

Then F is infinitely differentiable on $(0,\infty)$, and moreover, for any $\alpha > 0$,

$$F''(\alpha) = \frac{1}{\alpha^2} V(X_\alpha).$$

Proof. Note that the assumption that $f \in L^\alpha(\mathbb{R}^n)$ (or equivalently that $F(\alpha) < \infty$) for all $\alpha > 0$ guarantees that $F(\alpha)$ is infinitely differentiable for $\alpha > 0$ and that we can freely change the order of taking expectations and differentiation.

Now observe that

$$F'(\alpha) = \frac{\int f^\alpha \log f}{\int f^\alpha} = \int f_\alpha \log f;$$

if we wish, we may also massage this to write

$$F'(\alpha) = \frac{1}{\alpha}[F(\alpha) - h(X_\alpha)]. \tag{11}$$

Differentiating again, we get

$$\begin{aligned} F''(\alpha) &= \frac{\int f^\alpha (\log f)^2}{\int f^\alpha} - \left(\frac{\int f^\alpha \log f}{\int f^\alpha}\right)^2 \\ &= \int f_\alpha (\log f)^2 - \left(\int f_\alpha \log f\right)^2 \\ &= \mathrm{Var}[\log f(X_\alpha)] = \mathrm{Var}\left[\frac{1}{\alpha}\{\log f_\alpha(X_\alpha) + F(\alpha)\}\right] \\ &= \frac{1}{\alpha^2}\mathrm{Var}[\log f_\alpha(X_\alpha)] = \frac{V(X_\alpha)}{\alpha^2}, \end{aligned}$$

as desired. □

3.4 Sharp concentration of information

The first part of this section describes a strategy for obtaining exponential deviation inequalities when one has uniform control on variances of a family of random variables (log-concavity is not an assumption needed for this part).

Theorem 17. *Suppose $X \sim f$, where $f \in L^\alpha(\mathbb{R}^n)$ for each $\alpha > 0$. Let $X_\alpha \sim f_\alpha$, where*

$$f_\alpha(x) = \frac{f^\alpha(x)}{\int f^\alpha}.$$

If $K = K(f) := \sup_{\alpha>0} V(X_\alpha)$, then

$$\mathbb{E}\big[e^{\beta\{\tilde{h}(X)-h(X)\}}\big] \le e^{Kr(-\beta)}, \quad \beta \in \mathbb{R},$$

where

$$r(u) = \begin{cases} u - \log(1+u) & \text{for } u > -1 \\ +\infty & \text{for } u \le -1 \end{cases}.$$

Proof. Suppose X is a random vector drawn from a density f on $\mathbb{R}^n$, and define, for each $\alpha > 0$, $F(\alpha) = \log \int f^\alpha$. Set

$$K = \sup_{\alpha>0} V(X_\alpha) = \sup_{\alpha>0} \alpha^2 F''(\alpha);$$

the second equality follows from Lemma 9. Since $f \in L^\alpha(\mathbb{R}^n)$ for each $\alpha > 0$, $F(\alpha)$ is finite and moreover, infinitely differentiable for $\alpha > 0$, and we can freely change the order of integration and differentiation when differentiating $F(\alpha)$.

From Taylor-Lagrange formula, for every $\alpha > 0$, one has

$$F(\alpha) = F(1) + (\alpha-1)F'(1) + \int_1^\alpha (\alpha-u)F''(u)du.$$

Using that $F(1) = 0$, $F''(u) \le K/u^2$ for every $u > 0$ and the fact that for $0 < \alpha < u < 1$, one has $\alpha - u < 0$, we get

$$\begin{aligned} F(\alpha) &\le (\alpha-1)F'(1) + K\int_1^\alpha \frac{\alpha-u}{u^2}du \\ &= (\alpha-1)F'(1) + K\left[-\frac{\alpha}{u} - \log(u)\right]_1^\alpha. \end{aligned}$$

Thus, for $\alpha > 0$, we have proved that

$$F(\alpha) \le (\alpha-1)F'(1) + K(\alpha - 1 - \log\alpha).$$

Setting $\beta = 1 - \alpha$, we have for $\beta < 1$ that

$$e^{F(1-\beta)} \le e^{-\beta F'(1)} e^{K(-\beta-\log(1-\beta))}. \tag{12}$$

Observe that $e^{F(1-\beta)} = \int f^{1-\beta} = \mathbb{E}[f^{-\beta}(X)] = \mathbb{E}[e^{-\beta\log f(X)}] = \mathbb{E}\big[e^{\beta\tilde{h}(X)}\big]$ and $e^{-\beta F'(1)} = e^{\beta h(X)}$; the latter fact follows from the fact that $F'(1) = -h(X)$ as is clear from the identity (11). Hence the inequality (12) may be rewritten as

$$\mathbb{E}\big[e^{\beta\{\tilde{h}(X)-h(X)\}}\big] \le e^{Kr(-\beta)}, \quad \beta \in \mathbb{R}. \tag{13}$$

□

Note that the function r is convex on $\mathbb{R}$ and has a quadratic behavior in the neighborhood of 0 ($r(u) \sim_0 \frac{u^2}{2}$) and a linear behavior at $+\infty$ ($r(u) \sim_\infty u$).

Corollary 5. *With the assumptions and notation of Theorem 17, we have for any $t > 0$ that*

$$\mathbf{P}\{\tilde{h}(X) - h(X) \geq t\} \leq \exp\left\{ - Kr\left(\frac{t}{K}\right)\right\}$$
$$\mathbf{P}\{\tilde{h}(X) - h(X) \leq -t\} \leq \exp\left\{ - Kr\left(-\frac{t}{K}\right)\right\}$$

The proof is classical and often called the Cramér-Chernoff method (see for example section 2.2 in [15]). It uses the Legendre transform φ^* of a convex function $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined for $y \in \mathbb{R}$ by

$$\varphi^*(y) = \sup_x xy - \varphi(x).$$

Notice that if $\min \varphi = \varphi(0)$ then for every $y > 0$, the supremum is reached at a positive x, that is $\varphi^*(y) = \sup_{x>0} xy - \varphi(x)$. Similarly, for $y < 0$, the supremum is reached at a negative x.

Proof. The idea is simply to use Markov's inequality in conjunction with Theorem 17, and optimize the resulting bound.

For the lower tail, we have for $\beta > 0$ and $t > 0$,

$$\begin{aligned}\mathbb{P}[\tilde{h}(X) - h(X) \leq -t] &\leq \mathbf{E}\left[e^{-\beta\left(\tilde{h}(X)-h(X)\right)}\right]e^{-\beta t}\\ &\leq \exp\left\{K\left(r(\beta) - \frac{\beta t}{K}\right)\right\}.\end{aligned}$$

Thus minimizing on $\beta > 0$, and using the remark before the proof, we get

$$\mathbb{P}[\tilde{h}(X) - h(X) \leq -t] \leq \exp\left\{ - K \sup_{\beta>0}\left(\frac{\beta t}{K} - r(\beta)\right)\right\} = e^{-Kr^*\left(\frac{t}{K}\right)}. \tag{14}$$

Let us compute the Legendre transform r^* of r. For every t, one has

$$r^*(t) = \sup_u tu - r(u) = \sup_{u>-1}\left(tu - u + \log(1+u)\right).$$

One deduces that $r^*(t) = +\infty$ for $t \geq 1$. For $t < 1$, by differentiating, the supremum is reached at $u = t/(1-t)$ and replacing in the definition we get

$$r^*(t) = -t - \log(1-t) = r(-t).$$

Thus $r^*(t) = r(-t)$ for all $t \in \mathbb{R}$. Replacing, in the inequality (14), we get the result for the lower tail.

For the upper tail, we use the same argument: for $\beta > 0$ and $t > 0$,

$$\begin{aligned}\mathbb{P}[\tilde{h}(X) - h(X) \geq t] &\leq \mathbf{E}\left[e^{\beta\left(\tilde{h}(X)-h(X)\right)}\right]e^{-\beta t}\\ &\leq \exp\left\{K\left(r(-\beta) - \frac{\beta t}{K}\right)\right\}.\end{aligned}$$

Thus minimizing on $\beta > 0$, we get

$$\mathbb{P}[\tilde{h}(X) - h(X) \geq t] \leq \exp\left\{ - K \sup_{\beta > 0} \left(\frac{\beta t}{K} - r(-\beta) \right) \right\}. \tag{15}$$

Using the remark before the proof, in the right hand side term appears the Legendre transform of the function $\tilde{r}$ defined by $\tilde{r}(u) = r(-u)$. Using that $r^*(t) = r(-t) = \tilde{r}(t)$, we deduce that $(\tilde{r})^* = (r^*)^* = r$. Thus the inequality (15) gives the result for the upper tail. $\square$

Theorem 18. *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. For $\beta < 1$,*

$$\mathbb{E}\left[e^{\beta[\tilde{h}(X) - h(X)]} \right] \leq \mathbb{E}\left[e^{\beta[\tilde{h}(X^*) - h(X^*)]} \right],$$

where X^* has density $f^* = e^{-\sum_{i=1}^n x_i}$, restricted to the positive quadrant.

Proof. Taking $K = n$ in Theorem 17 (which we can do in the log-concave setting because of Theorem 16), we obtain:

$$\mathbb{E}\left[e^{\beta\{\tilde{h}(X) - h(X)\}} \right] \leq e^{nr(-\beta)}, \quad \beta \in \mathbb{R}.$$

Some easy computations will show:

$$\mathbb{E}\left[e^{\beta\{\tilde{h}(X^*) - h(X^*)\}} \right] = e^{nr(-\beta)}, \quad \beta \in \mathbb{R}.$$

This concludes the proof.

$\square$

As for the case of equality of Theorem 16, discussed in Remark **??**, notice that there is a broader class of densities for which one has equality in Theorem 18, including all those of the form $e^{-\|x\|_K}$, where K is a symmetric convex body.

Remarks:

1. The assumption $\beta < 1$ in Theorem 18 is strictly not required; however, for $\beta \geq 1$, the right side is equal to $+\infty$. Indeed, already for $\beta = 1$, one sees that for any random vector X with density f,

$$\begin{aligned} \mathbb{E}\left[e^{\tilde{h}(X) - h(X)} \right] &= e^{-h(X)} \mathbb{E}\left[\frac{1}{f(X)} \right] = e^{-h(X)} \int_{\text{supp(f)}} dx \\ &= e^{-h(X)} \text{Vol}_n(\text{supp(f)}), \end{aligned}$$

   where $\text{supp(f)} = \overline{\{x \in \mathbb{R}^n : f(x) > 0\}}$ is the support of the density f and Vol_n denotes Lebesgue measure on $\mathbb{R}^n$. In particular, this quantity for X^*, whose support has infinite Lebesgue measure, is $+\infty$.

2. Since

$$\lim_{\alpha \to 0} \frac{2}{\alpha^2} \mathbf{E}\left[e^{\alpha(\log f(X) - \mathbf{E}[\log f(X)])} \right] = V(X),$$

   we can recover Theorem 16 from Theorem 18.

Taking $K = n$ in Corollary 5 (again because of Theorem 16), we obtain:

Corollary 6. *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. For $t > 0$,*

$$\mathbb{P}[\tilde{h}(X) - h(X) \leq -nt] \leq e^{-nr(-t)},$$
$$\mathbb{P}[\tilde{h}(X) - h(X) \geq nt] \leq e^{-nr(t)},$$

where $r(u)$ is defined in Theorem 17.

The following inequality[9] is an immediate corollary of Corollary 6 since it merely expresses a bound on the support of the distribution of the information content.

Corollary 7. *Let X have a log-concave probability density function f on $\mathbb{R}^n$. Then:*

$$h(X) \leq -\log \|f\|_\infty + n.$$

Proof. By Corollary 6, almost surely,

$$\log f(X) \leq \mathbf{E}[\log f(X)] + n,$$

since when $t \geq 1$, $\mathbb{P}[\log f(X) - \mathbf{E}[\log f(X)] \geq nt] = 0$. Taking the supremum over all realizable values of X yields

$$\log \|f\|_\infty \leq \mathbf{E}[\log f(X)] + n,$$

which is equivalent to the desired statement. □

An immediate consequence of Corollary 7, unmentioned in [10], is a result due to [27]:

Corollary 8. *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. Then*

$$\|f\|_\infty \leq e^n f(\mathbf{E}[X]).$$

Proof. By Jensen's inequality,

$$\log f(\mathbb{E}X) \geq \mathbb{E}[\log f(X)].$$

By Corollary 7,

$$\mathbb{E}[\log f(X)] \geq \log \|f\|_\infty - n.$$

Hence,

$$\log f(\mathbb{E}X) \geq \log \|f\|_\infty - n.$$

Exponentiating concludes the proof. □

[9] Corollary 7 was obtained by Ball around 2003, and is also implicitly contained in [31] (see the proof of Theorem 7 there). However, it was first explicitly stated and proved in [10], who independently obtained it and developed several applications.

3.5 What happens for s-concave case

Recall that for $x > 0$, the gamma function $\Gamma(x)$ is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt.$$

For $x, y > 0$, the beta function $B(x, y)$ is defined by

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt.$$

The following result is proved by Borell [13] for $s > 0$, except that the function φ is assumed to be decreasing. It was then noticed by some people and available for example in Guédon, Nayar and Tkocz [36] that the result remains true without any monotonicity hypothesis. For $s < 0$, it is proved by Fradelizi, Guédon and Pajor [28], and the case $s = 0$ follows by taking the limits (or reproducing the mechanics of the proof).

Proposition 1. *Let $s \in \mathbb{R}$ and let $\varphi : [0, +\infty) \to [0, +\infty)$ be an s-concave integrable function.*
1) If $s > 0$, then $p \mapsto B(p, s^{-1} + 1)^{-1} \int_0^{+\infty} t^{p-1}\varphi(t)dt$ is log-concave for $p > 0$.
2) If $s = 0$, then $p \mapsto \Gamma(p)^{-1} \int_0^{+\infty} t^{p-1}\varphi(t)dt$ is log-concave for $p > 0$.
3) If $s < 0$, then $p \mapsto B(p, -s^{-1} - p)^{-1} \int_0^{+\infty} t^{p-1}\varphi(t)dt$ is log-concave for $0 < p < -1/s$.

Let us define the function $\varphi_s(t) = (1 - st)_+^{1/s}\mathbf{1}_{\mathbb{R}_+}$ for $s \neq 0$, and $\varphi_0(t) = e^{-t}\mathbf{1}_{\mathbb{R}_+}$. Then the preceding proposition may be expressed in the following way: if $\varphi : [0, +\infty) \to [0, +\infty)$ is s-concave, then the function

$$p \mapsto \frac{\int_0^{+\infty} t^{p-1}\varphi(t)dt}{\int_0^{+\infty} t^{p-1}\varphi_s(t)dt}$$

is log-concave for p such that $1/p > \max(0, -s)$. Using the preceding proposition, we can prove the following theorem which unifies and partially extends previous results of Borell [13], Bobkov and Madiman [10], and Fradelizi, Madiman and Wang [30].

Theorem 19. *Let $s \in \mathbb{R}$ and let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable s-concave function. Then the function*

$$p \mapsto (p + s) \cdots (p + ns) \int_{\mathbb{R}^n} f(x)^p dx \tag{16}$$

is log-concave for $p > \max(0, -ns)$.

Proof. The case $s = 1$ is due to Borell [13] and the case $s > 0$ deduces directly by applying Borell's result to f^s. The case $s = 0$ was proved by Fradelizi, Madiman and Wang [30]. The case $s = -1$ is due to Bobkov and Madiman [10][10], except that the range was $p > n + 1$. In the same way, the case $s < 0$ deduces from the case $s = -1$ by applying it to $f^{|s|}$. So we only need to prove the extension of the range for $s = -1$. Let us assume that $s = -1$. Thus f is -1-concave, which means that $g = f^{-1}$ is convex on its support. As done by Bobkov and Madiman [10], we write

$$\int_{\mathbb{R}^n} f(x)^p dx = \int_{\mathbb{R}^n} g(x)^{-p} dx = \int_0^{+\infty} pt^{p-1}\psi(1/t)dt,$$

[10] The details of this proof were omitted from [10] because of space considerations, and are being presented here. A complete presentation will appear in [7].

where $\psi(t) = |\{x \in \mathbb{R}^n : g(x) \le t\}|_n$ is the Lebesgue measure of the sub-level set $\{x \in \mathbb{R}^n : g(x) \le t\}$. Using Brunn-Minkowski theorem, we can see that ψ is a $1/n$-concave function. Using the properties of the perspective function, we can deduce that the function $\varphi(t) = t^n\psi(1/t)$ is also a $1/n$-concave function. Thus it follows that

$$\int_{\mathbb{R}^n} f(x)^p dx = p\int_0^{+\infty} t^{p-n-1}\varphi(t)dt.$$

Applying Proposition 1 to $s = 1/n$ and p replaced by $p-n$ we get that

$$B(p-n, n+1)^{-1}\int_0^{+\infty} t^{p-1-n}\varphi(t)dt$$

is log-concave on $(n, +\infty)$. Then we can conclude the proof using the following identity

$$B(p-n, n+1)^{-1} = \frac{p(p-1)\cdots(p-n)}{\Gamma(n+1)}.$$

□

The fact that Theorem 19 is optimal can be seen from the following example. Let $U : \mathbb{R}^n \to [0, \infty]$ be a positively homogeneous convex function of degree 1, i.e. that $U(tx) = tU(x)$ for all $x \in \mathbb{R}^n$ and all $t > 0$. We define $f_{s,U} = (1 - sU)_+^{1/s}$ for $s \neq 0$ and $f_{0,U} = e^{-U}$ for $s = 0$. Then we have

$$\int_{\mathbb{R}^n} f_{s,U}(x)^p dx = \frac{C_U n!}{(p+s)\cdots(p+ns)},$$

where C_U is the Lebesgue measure of the sub-level set $\{x \in \mathbb{R}^n : U(x) \le 1\}$. We only check the identity for $s > 0$, and the other two cases can be proved similarly.

$$\begin{aligned}
\int_{\mathbb{R}^n} f_{s,U}(x)^p dx &= p\int_0^1 t^{p-1}|\{x \in \mathbb{R}^n : (1 - sU(x))_+^{1/s} > t\}|dt \\
&= p\int_0^1 t^{p-1}|\{x \in \mathbb{R}^n : U(x) < (1-t^s)/s\}|dt \\
&= C_U p\int_0^1 t^{p-1}((1-t^s)/s)^n dt \\
&= C_U s^{-n-1} pB(p/s, n+1)
\end{aligned}$$

In the third equation, we use the homogeneity of U and the property of Lebesgue measure. Then we can prove the identity using the following fact

$$B(p/s, n+1) = \frac{n!}{(p/s+n)\cdots p/s}.$$

Thus the preceding theorem can be written in the following way: if $f : \mathbb{R}^n \to \mathbb{R}_+$ is an integrable s-concave function, then

$$p \mapsto \frac{\int_{\mathbb{R}^n} f(x)^p dx}{\int_{\mathbb{R}^n} f_{s,U}(x)^p dx}$$

is log-concave for $p > \max(0, -ns)$.

We say that a $\mathbb{R}^n$-valued random variable X is κ-concave if the probability measure induced by X is κ-concave. In this section, we let X be a κ-concave random variable with density f and $\kappa < 0$. Then Borell's characterization implies that there is a convex function V such that $f = V^{-\beta}$. In the following, we will study the deviation of $\widetilde{h}(X)$ from its mean $h(X)$, that is corresponding to taking $\varphi = -\log f$ in Section 2. Then the moment generating function is

$$\mathbb{E}f^{-\alpha}(X) = \int_{\mathbb{R}^n} f(x)^{1-\alpha} dx.$$

The integral is finite as long as $(1-\alpha)\beta > n$, i.e. that $\alpha < 1 - n/\beta$.

Proposition 2. *Let $\beta > n$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. Then the function*

$$\alpha \mapsto \prod_{i=1}^{n} ((1-\alpha)\beta - i)\mathbb{E}f^{-\alpha}(X) \tag{17}$$

is log-concave for $\alpha < 1 - n/\beta$.

Proof. It easily follows from Theorem 19 with p replaced by $1-\alpha$ and s replaced by $-1/\beta$. □

Following Lemma **??**, we can set

$$c(\alpha) = -\sum_{i=1}^{n} \log((1-\alpha)\beta - i). \tag{18}$$

Corollary 9. *Under the conditions and notations of Proposition 2, we have*

$$Var(\widetilde{h}(X)) \leq \beta^2 \sum_{i=1}^{n} (\beta - i)^{-2}. \tag{19}$$

Proof. By Lemma **??**, we know that $\mathrm{Var}(\widetilde{h}(X)) = L''(\alpha)$, where X_α is a random variable with density proportional to $f^{1-\alpha}$ and $L(\alpha) = \log \mathbb{E}f^{-\alpha}(X)$ is the logarithmic moment generating function. By Proposition 2, we know that $L''(\alpha) \leq c''(\alpha)$, where $c(\alpha)$ is defined in (18). Then the variance bound (19) follows by differentiating $c(\alpha)$ twice and setting $\alpha = 0$. □

The variance bound is sharp. Suppose X has density $f = (1 + U/\beta)_+^{-\beta}$ with U being a positively homogeneous convex function of degree 1. In this case, the function in Proposition 2 is log-affine, i.e. $L''(\alpha) = c''(\alpha)$. Then we have equality in the above variance bound. In particular, it includes the Pareto distribution with density

$$f(x) = \frac{1}{Z_n(a, \beta)} (a + x_1 + \cdots + x_n)^{-\beta}, \ x_i > 0, \tag{20}$$

where $a > 0$ and $Z_n(a, \beta)$ is a normalizing constant.

Theorem 20. *Let $\beta > n$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. Then we have*

$$\mathbb{E}e^{\alpha(\widetilde{h}(X) - h(X))} \leq e^{\psi_c(\alpha)} \tag{21}$$

for $\alpha < 1 - n/\beta$, where

$$\psi_c(\alpha) = -\alpha\beta \sum_{i=1}^{n} (\beta - i)^{-1} - \sum_{i=1}^{n} \log \frac{(1-\alpha)\beta - i}{\beta - i}. \tag{22}$$

Particularly, we have equality for Pareto distributions.

Proof. The moment generating function bound (21) easily follows from Lemma **??** and Proposition 2. Some easy calculations will show the equality case for Pareto distributions. Essentially that is due to the identity $L''(\alpha) = c''(\alpha)$, where $c(\alpha)$ is defined in (18).

□

Corollary 10. *Under the conditions and notations of Theorem 20, we have for $t > 0$ that*

$$\mathbf{P}(\widetilde{h}(X) - h(X) > t) \le e^{-\psi^*_{c,+}(t)}, \tag{23}$$

$$\mathbf{P}(\widetilde{h}(X) - h(X) < -t) \le e^{-\psi^*_{c,-}(-t)}, \tag{24}$$

where $\psi^*_{c,+}$ and $\psi^*_{c,-}$ are Fenchel-Legendre dual functions of $\psi_{c,+}$ and $\psi_{c,-}$, respectively.

In general we do not have explicit expressions for $\psi^*_{c,+}$ or $\psi^*_{c,-}$. The following result was obtained by Bobkov and Madiman [10] with the assumption $\beta \ge n+1$, which can be relaxed to $\beta > n$. It basically says that the entropy of an κ-concave distribution can not exceed that of the Pareto distribution with the same maximal density value.

Corollary 11. *Under the conditions and notations of Theorem 20, we have*

$$h(X) \le -\log \|f\|_\infty + \beta \sum_{i=1}^{n} (\beta - i)^{-1}, \tag{25}$$

where we denote by $\|f\|_\infty$ the essential supremum. We have equality for Pareto distributions.

Proof. As a function of α, we have

$$(-\alpha t - \psi_c(\alpha))' = -t + \beta \sum_{i=1}^{n} (\beta - i)^{-1} - \beta \sum_{i=1}^{n} ((1-\alpha)\beta - i)^{-1}.$$

For any $t > \beta \sum_{i=1}^{n} (\beta - i)^{-1}$, we can see that $-\alpha t - \psi_c(\alpha)$ is a decreasing function of $\alpha < 1 - n/\beta$. It is clear that $\lim_{\alpha \to -\infty} (-\alpha t - \psi_c(\alpha)) = \infty$. Therefore we have $\psi^*_{c,-}(-t) = \infty$ for $t > \beta \sum_{i=1}^{n} (\beta - i)^{-1}$. Using the lower tail estimate in Corollary 10, almost surely we have

$$\widetilde{h}(X) - h(X) \ge -\beta \sum_{i=1}^{n} (\beta - i)^{-1}.$$

Taking the supremum over all realizable values of X yields

$$-\log \|f\|_\infty - h(X) \ge -\beta \sum_{i=1}^{n} (\beta - i)^{-1}.$$

That is equivalent to the desired statement. □

4 Lecture 4: Convexity, curvature, and concentration

4.1 Weighted Prékopa-Leindler inequality

A uniform lower bound on the Hessian of V is a way to quantify a property stronger than log-concavity of e^{-V}. For such uniformly log-concave measures, one can obtain a local-global inequality that strengthens PLI.

First we need an easy lemma.

Lemma 10. *If $V : \mathbb{R}^n \to \mathbb{R}$ is C^2 and satisfies Hess$(V) \geq c \in \mathbb{R}$, then*

$$\lambda V(x) + (1-\lambda)V(y) - V(\lambda x + (1-\lambda)y) \geq c\lambda(1-\lambda)\frac{|x-y|^2}{2}.$$

Theorem 21. *Let μ be a probability measure on $\mathbb{R}^n$ with a density of form e^{-V}, where Hess$(V) \geq c \in \mathbb{R}$. Suppose $f, g, h : \mathbb{R}^n \to [0, \infty)$ are integrable functions satisfying, for a given $\lambda \in (0,1)$,*

$$h(\lambda x + (1-\lambda)y) \geq \exp\left\{ - c\lambda(1-\lambda)\frac{|x-y|^2}{2}\right\} f^{\lambda}(x) g^{1-\lambda}(y)$$

for every $x, y \in \mathbb{R}^d$, then

$$\int h d\mu \geq \left(\int f d\mu\right)^{\lambda}\left(\int g d\mu\right)^{1-\lambda}. \tag{26}$$

Proof. Follows from usual PLI. □

This observation was made by Cordero, McCann and Schmuckenschlager [**?**], where they in fact proved such an inequality on Riemannian manifolds. It is instructive to think in that broader context...

Draw pictures.

4.2 Properties of uniformly log-concave measures

TO BE ADDED: The PLI is also useful to get many other things, including LSI and concentration for a large subclass of LC measures.

Explain concentration using examples of sphere and Gaussian.

Compare with isoperimetry, and state Gaussian isoperimetric inequality.

4.3 Aside: Ways of capturing curvature

TO BE ADDED: a discussion about curvature-dimension conditions for metric measure spaces, and how this is related to convex measures.

5 Lecture 5: Klartag's CLT

5.1 Marginals and Gaussian mixtures

TO BE ADDED...

5.2 The thin shell phenomenon

TO BE ADDED...

References

[1] P.H. Algoet and T.M. Cover. A sandwich proof of the Shannon-Mcmillan-Breiman theorem. *Ann. Probab.*, 16:876–898, 1988.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.

[3] Z. Artstein and R. A. Vitale. A strong law of large numbers for random compact sets. *Ann. Probab.*, 3(5):879–882, 1975.

[4] F. Balabdaoui, K. Rufibach, and J. A. Wellner. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.*, 37(3):1299–1331, 2009.

[5] A.R. Barron. The strong ergodic theorem for densities: Generalized Shannon-Mcmillan-Breiman theorem. *Ann. Probab.*, 13:1292–1303, 1985.

[6] D. Bertsimas and S. Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556 (electronic), 2004.

[7] S. Bobkov, M. Fradelizi, J. Li, and M. Madiman. When can one invert Hölder's inequality? (and why one may want to). *Preprint*, 2016.

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

[9] S. Bobkov and M. Madiman. Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.*, 349:201–204, Février 2011.

[10] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, August 2011.

[11] S. G. Bobkov, M. Madiman, and L. Wang. Fractional generalizations of Young and Brunn-Minkowski inequalities. In C. Houdré, M. Ledoux, E. Milman, and M. Milman, editors, *Concentration, Functional Inequalities and Isoperimetry*, volume 545 of *Contemp. Math.*, pages 35–53. Amer. Math. Soc., 2011.

[12] F. Bolley, I. Gentil, and A. Guillin. Dimensional improvements of the logarithmic sobolev, talagrand and brascamp-lieb inequalities. *Preprint,* `arXiv:1507:01086`, July 2015.

[13] C. Borell. Complements of Lyapunov's inequality. *Math. Ann.*, 205:323–331, 1973.

[14] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12:239–252, 1974.

[15] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[16] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Functional Analysis*, 22(4):366–389, 1976.

[17] L. Breiman. The individual ergodic theorem for information theory. *Ann. Math. Stat.*, 28:809–811, 1957. (See also the correction: *Ann. Math. Stat.*, Vol. 31, pp. 809-810, 1960).

[18] S. Bubeck and R. Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. *Preprint,* `arXiv:1412.1587`, 2015.

[19] M. Cule and R. Samworth. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Preprint*, 2009.

[20] A. Dvoretzky. A theorem on convex bodies and applications to Banach spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 45:223–226; erratum, 1554, 1959.

[21] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, pages 123–160. Jerusalem Academic Press, Jerusalem; Pergamon, Oxford, 1961.

[22] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.*, 23(2):532–569, 2013.

[23] R. Eldan and B. Klartag. Approximately gaussian marginals and the hyperplane conjecture. In C. Houdré, M. Ledoux, E. Milman, and M. Milman, editors, *Concentration, Functional Inequalities and Isoperimetry*, volume 545 of *Contemp. Math.*, pages 55–68. Amer. Math. Soc., 2011.

[24] R. Eldan and J. Lehec. Bounding the norm of a log-concave vector via thin-shell estimates. In B. Klartag and E. Milman, editors, *Geometric Aspects of Functional Analysis*, volume 2116 of *Lecture Notes in Mathematics*, pages 107–122. Springer, 2014.

[25] W. R. Emerson and F. P. Greenleaf. Asymptotic behavior of products $C^p = C + \cdots + C$ in locally compact abelian groups. *Trans. Amer. Math. Soc.*, 145:171–204, 1969.

[26] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.

[27] M. Fradelizi. Sections of convex bodies through their centroid. *Arch. Math. (Basel)*, 69(6):515–522, 1997.

[28] M. Fradelizi, O. Guédon, and A. Pajor. Thin-shell concentration for convex measures. *Studia Math.*, 223(2):123–148, 2014.

[29] M. Fradelizi, M. Madiman, A. Marsiglietti, and A. Zvavitch. Do Minkowski averages get progressively more convex? *C. R. Acad. Sci. Paris Sér. I Math.*, 354(2):185–189, February 2016.

[30] M. Fradelizi, M. Madiman, and L. Wang. Optimal concentration of information content for log-concave densities. In C. Houdré, D. Mason, P. Reynaud-Bouret, and J. Rosinski, editors, *High Dimensional Probability VII: The Cargèse Volume*, Progress in Probability. Birkhäuser, Basel, 2016. Available online at `arXiv:1508.04093`.

[31] M. Fradelizi and M. Meyer. Increasing functions and inverse Santaló inequality for unconditional functions. *Positivity*, 12(3):407–420, 2008.

[32] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405 (electronic), 2002.

[33] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.*, 50(4):265–289, 1985.

[34] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.

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

[36] O. Guédon, P. Nayar, and T. Tkocz. Concentration inequalities and geometry of convex bodies. In *Analytical and Probabilistic Methods in the Geometry of Convex Bodies*, volume 2 of *IM PAN Lecture Notes*, pages 9–86. Polish Acad. Sci., Warsaw, 2014.

[37] L. G. Hačijan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244(5):1093–1096, 1979.

[38] H. Hadwiger and D. Ohmann. Brunn-Minkowskischer Satz und Isoperimetrie. *Math. Z.*, 66:1–8, 1956.

[39] G. Hargé. Reinforcement of an inequality due to Brascamp and Lieb. *J. Funct. Anal.*, 254(2):267–300, 2008.

[40] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

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

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

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

[44] R. Lang. A note on the measurability of convex sets. *Arch. Math. (Basel)*, 47(1):90–92, 1986.

[45] L. Leindler. On a certain converse of Hölder's inequality. II. *Acta Sci. Math. (Szeged)*, 33(3-4):217–223, 1972.

[46] W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland, Amsterdam, 2001.

[47] E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[48] L. A. Lusternik. Die Brunn-Minkowskische ungleichung fur beliebige messbare mengen. *C. R. (Doklady) Acad. Sci. URSS*, 8:55–58, 1935.

[49] B. McMillan. The basic theorems of information theory. *Ann. Math. Stat.*, 24:196–219, 1953.

[50] E. Milman. Sharp isoperimetric inequalities and model spaces for the curvature-dimension-diameter condition. *J. Eur. Math. Soc. (JEMS)*, 17(5):1041–1078, 2015.

[51] V. Milman. Dvoretzky's theorem—thirty years later. *Geom. Funct. Anal.*, 2(4):455–479, 1992.

[52] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[53] V. H. Nguyen. *Inégalités fonctionnelles et convexité*. PhD thesis, Université Pierre et Marie Curie (Paris VI), October 2013.

[54] V. H. Nguyen. Dimensional variance inequalities of Brascamp-Lieb type and a local approach to dimensional Prékopa's theorem. *J. Funct. Anal.*, 266(2):931–955, 2014.

[55] S.-i. Ohta and A. Takatsu. Displacement convexity of generalized relative entropies. *Adv. Math.*, 228(3):1742–1787, 2011.

[56] S.-I. Ohta and A. Takatsu. Displacement convexity of generalized relative entropies. II. *Comm. Anal. Geom.*, 21(4):687–785, 2013.

[57] S. Orey. On the Shannon-Perez-Moy theorem. In *Particle systems, random media and large deviations (Brunswick, Maine, 1984)*, pages 319–327. Amer. Math. Soc., Providence, R.I., 1985.

[58] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32:301–316, 1971.

[59] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.

[60] G. Schechtman. A remark concerning the dependence on ϵ in Dvoretzky's theorem. In *Geometric aspects of functional analysis (1987–88)*, volume 1376 of *Lecture Notes in Math.*, pages 274–277. Springer, Berlin, 1989.

[61] R. Schneider. A measure of convexity for compact sets. *Pacific J. Math.*, 58(2):617–625, 1975.

[62] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[63] R. M. Starr. Quasi-equilibria in markets with non-convex preferences. *Econometrica*, 37(1):25–38, January 1969.

[64] P. M. Vaidya. A new algorithm for minimizing convex functions over convex sets. *Math. Programming*, 73(3, Ser. A):291–341, 1996.

[65] L. Wang. *Heat capacity bound, energy fluctuations and convexity*. PhD thesis, Yale University, May 2014.

[66] L. Wang and M. Madiman. Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory*, 60(9):5116–5137, September 2014.

[67] D. B. Yudin and A. S. Nemirovski. Evaluation of the information complexity of mathematical programming problems (in Russian). *Ekonomika i Matematicheskie Metody*, 12:128–142, 1976. (English translation: Matekon 13, 2, 3–45, 1976).