High-Dimensional Phenomena and Convexity

Bo'az Klartag

Tel-Aviv University

MRC Lecture,

Stanford, September 2011

High-dimensional distributions

- We consider probability measures in high dimensions. Are there any general, interesting principles?

The classical central limit theorem

Suppose $X = (X_1, \ldots, X_n)$ is a random vector in $\mathbb{R}^n$, with independent components. Assume that n is large. Then, under mild assumptions, there exist coefficients $\theta_1, \ldots, \theta_n, b \in \mathbb{R}$ with

$$\mathbb{P}\left(\sum_{i=1}^{n} \theta_i X_i \leq t\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{(s-b)^2}{2}\right) ds \qquad (\forall t \in \mathbb{R})$$

- When X is properly normalized, i.e.,

$$\mathbb{E}X_i = 0, \qquad Var(X_i) = 1$$

we may select $\theta = (1, \ldots, 1)/\sqrt{n}$. Furthermore:
- *Most* choices of $\theta_1, \ldots, \theta_n \in \mathbb{R}$ with $\sum_i \theta_i^2 = 1$ will work.

Structure, symmetry or convexity?

- The CLT shows that measures composed of independent (or approx. indep.) random variables are quite regular.
- High-dimensional distributions with a clear *structure* or with *symmetries* might be easier to analyze.

We shall see that **convexity** conditions fit very well with the **high dimensionality**.

- Uniform measures on convex domains.
- Densities of the form $\exp(-H)$ on $\mathbb{R}^n$, with a convex H.

Convexity may sometimes substitute for structure and symmetries. The geometry of $\mathbb{R}^n$ forces regularity (usually, but not always, convexity is required).

An example: the sphere

- Consider the sphere $S^{n-1} = \{x \in \mathbb{R}^n ; |x| = 1\}$.

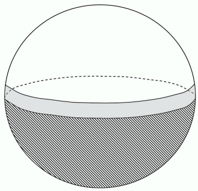
For a set $A \subseteq S^{n-1}$ and for $\varepsilon > 0$ denote

$$A_\varepsilon = \left\{ x \in S^{n-1} ; \exists y \in A, \, d(x,y) \leq \varepsilon \right\},$$

which is the ε-neighborhood of A. Write σ_{n-1} for the uniform probability measure on S^{n-1}.

- Consider the hemisphere $H = \{x \in S^{n-1} ; x_1 \leq 0\}$. Then,

$$\sigma_{n-1}(H_\varepsilon) = \mathbb{P}(Y_1 \leq \sin \varepsilon) \approx \mathbb{P}\left(\Gamma \leq \varepsilon\sqrt{n}\right)$$



where $Y = (Y_1, \ldots, Y_n)$ is distributed according to σ_{n-1}, and Γ is a standard normal random variable.

Concentration of measure

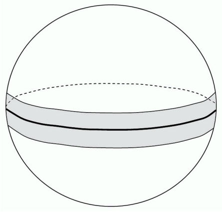
The amount of volume of a distance at least $1/10$ from the equator is at most

$$C\exp(-cn)$$

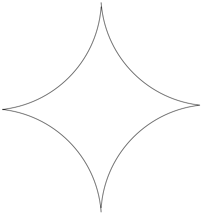
for universal constants $c, C > 0$.

- Most of the mass of the sphere S^{n-1} in high dimensions, is concentrated at a narrow strip near the equator $[x_1 = 0]$
- or any other equator.

“Concentration of Measure”



$dim \to \infty$



Use of the isoperimetric inequality

The isoperimetric inequality (Lévy, Schmidt, '50s).

For any Borel set $A \subset S^{n-1}$ and $\varepsilon > 0$,

$$\sigma_{n-1}(A) = 1/2 \quad \Rightarrow \quad \sigma_{n-1}(A_\varepsilon) \geq \sigma_{n-1}(H_\varepsilon),$$

where $H = \{x \in S^{n-1}; x_1 \leq 0\}$ is a hemisphere.

- For any set $A \subset S^{n-1}$ with $\sigma_{n-1}(A) = 1/2$,

$$\sigma_{n-1}(A_\varepsilon) \geq 1 - 2\exp(-\varepsilon^2 n/2).$$

Therefore, for any subset $A \subset S^{n-1}$ of measure $1/2$, its ε-neighborhood covers almost the entire sphere.

Concentration of Lipschitz functions

Corollary ("Lévy's lemma")

Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function. Denote

$$E = \int_{S^{n-1}} f(x) d\sigma_{n-1}(x).$$

Then, for any $\varepsilon > 0$,

$$\sigma_{n-1}\left(\left\{x \in S^{n-1}; |f(x) - E| \geq \varepsilon\right\}\right) \leq C \exp(-c\varepsilon^2 n),$$

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

- Lipschitz functions on the high-dimensional sphere are "effectively constant".

Approximately-Gaussian marginals

Maxwell's observation: The sphere's marginals are approximately Gaussian ($n \to \infty$).

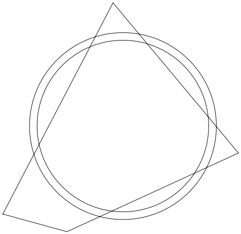
- What other distributions in high dimension have approximately-Gaussian marginals?

Normalization: A random vector $X = (X_1, \ldots, X_n)$ is "normalized" or "isotropic" or "balanced" if

$$\mathbb{E}X_i = 0, \quad \mathbb{E}X_iX_j = \delta_{i,j} \quad \forall i,j = 1, \ldots, n.$$

i.e., marginals have mean zero and var. one.

- **Diaconis-Freedman '84**:
  In order to have approx. Gaussian marginals, we need most of the mass of the random vector X to be contained in a **thin spherical shell**.



Thin-shell theorem

Theorem (Sudakov '76, Diaconis-Freedman '84,...)

Let X be a normalized random vector in $\mathbb{R}^n$, $\varepsilon > 0$. Assume

$$\mathbb{P}\left(\left| \frac{|X|}{\sqrt{n}} - 1 \right| \geq \varepsilon \right) \leq \varepsilon.$$

Then, there exists a subset $\Theta \subseteq S^{n-1}$, which is large, i.e., $\sigma_{n-1}(\Theta) \geq 1 - e^{-c\sqrt{n}}$, such that for any $\theta \in \Theta$,

$$|\mathbb{P}(X \cdot \theta \leq t) - \mathbb{P}(\Gamma \leq t)| \leq C\left(\varepsilon + \frac{1}{n^c}\right) \quad \forall t \in \mathbb{R}$$

where Γ is a standard Gaussian random variable.

- This "thin shell" assumption is also necessary.
- It is satisfied by i.i.d measures with bounded 4^{th} moments: $\mathbb{E}\left(|X|^2/n - 1\right)^2 = Var(|X|^2/n) = Var(X_1^2)/n \ll 1.$

Proof of thin-shell theorem

Main idea in proof: *The concentration phenomenon*.

- Fix $t \in \mathbb{R}$. Define

$$F_t(\theta) = \mathbb{P}(X \cdot \theta \leq t) \qquad (\theta \in S^{n-1}).$$

We need to prove that

For **most** unit vectors $\theta \in S^{n-1}$,

$$F_t(\theta) = \mathbb{P}(X \cdot \theta \leq t) \approx \mathbb{P}(\Gamma \leq t).$$

(a) Introduce a random vector Y, uniform on S^{n-1}, independent of X. Then,

$$\int_{S^{n-1}} F_t(\theta) d\sigma_{n-1}(\theta) = \mathbb{P}\left(|X|\, Y_1 \leq t\right) \approx \mathbb{P}(\Gamma \leq t).$$

(b) The function F_t typically deviates little from its mean (it has a Lipschitz approximation).

Violation of thin shell condition

- Consider the isotropic probability measure

$$\frac{1}{2}\left[\sigma_{n-1}^{r_1} + \sigma_{n-1}^{r_2}\right]$$

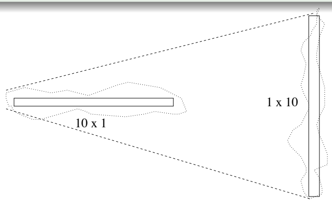
for appropriate $r_1, r_2 > 0$, where σ_{n-1}^{r} is the uniform probability on rS^{n-1}. It violates the **thin shell assumption.**

The main problem: “mixture of different scales”.

Anttila, Ball and Perissinaki ’03, Brehm and Voigt ’00:

Perhaps **convexity** conditions may rule out such examples.

Maybe convex bodies are inherently of a single scale?



What's special about convex sets?

Consider the classical Brunn-Minkowski inequality (1887):

$$Vol\left(\frac{A+B}{2}\right) \geq \sqrt{Vol(A)Vol(B)}$$

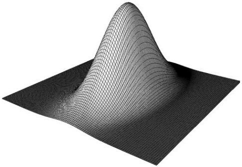
for any Borel sets $A, B \subset \mathbb{R}^n$.
Here $(A+B)/2 = \{(a+b)/2\,;\; a \in A, b \in B\}$.

- This inequality says a lot about convex sets.

A density function in $\mathbb{R}^n$ is **log-concave** if it takes the form e^{-H} with $H : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, a convex function.

The Gaussian density is log-concave, as is the characteristic function of a convex set.

- Any marginal, of any dimension, of a log-concave measure (e.g., the uniform measure on a convex set) is itself log-concave.



Back to thin shell bounds

Let μ be an isotropic probability measure on $\mathbb{R}^n$. To get approx. normal marginals, we need $|x|$ to be μ-concentrated near $\sqrt{n}$, i.e.,

$$\int_{\mathbb{R}^n} \left(\frac{|x|^2}{n} - 1 \right)^2 d\mu(x) \ll 1. \qquad (1)$$

- A possible attack on (1): Try to prove

$$\int_{\mathbb{R}^n} \varphi^2 d\mu \leq C \int_{\mathbb{R}^n} |\nabla \varphi|^2 d\mu \qquad (2)$$

  for **all** functions φ with $\int \varphi d\mu = 0$. We only need $\varphi(x) = |x|^2/n - 1$.
- This is a Poincaré-type inequality.

Spectral gap problem

Conjecture (Kannan, Lovász and Simonovits '95)

When μ is log-concave and isotropic in $\mathbb{R}^n$ and $\int \varphi d\mu = 0$,

$$\int_{\mathbb{R}^n} \varphi^2 d\mu \leq C \int_{\mathbb{R}^n} |\nabla \varphi|^2 d\mu.$$

- This is a spectral gap problem, for the operator
$$\triangle_\mu \varphi = \triangle \varphi - \nabla H \cdot \nabla \varphi$$
where $\exp(-H)$ is the density of μ.
- Equivalent to an isoperimetric inequality on convex bodies.

Why convexity? why log-concavity?

Because $\nabla^2 H \geq 0$. The **Bochner-Weitzenböck identity:**

$$\int_{\mathbb{R}^n} (\triangle_\mu \varphi)^2 d\mu = \int_{\mathbb{R}^n} |\nabla^2 \varphi|_{HS}^2 d\mu + \int_{\mathbb{R}^n} (\nabla^2 H)(\nabla \varphi) \cdot \nabla \varphi d\mu$$

Strong convexity assumptions

- It follows easily that when μ is a probability measure on $\mathbb{R}^n$ with density $\exp(-H)$ such that $\nabla^2 H \geq \delta$,

$$\delta \int_{\mathbb{R}^n} \varphi^2 d\mu \leq \int_{\mathbb{R}^n} |\nabla \varphi|^2 d\mu$$

for any φ with $\int \varphi d\mu = 0$.

Assume that μ is isotropic. To get approx. Gaussian marginals we need

$$\int_{\mathbb{R}^n} \left(\frac{|x|^2}{n} - 1 \right)^2 d\mu(x) \ll 1.$$

- Thus, we get a non-trivial **thin shell bound** and approximately Gaussian marginals as long as $\delta \gg 1/n$. This is more than log-concavity, but not so bad.

Central limit theorem for convex sets

- What can we do without making **strong** uniform convexity assumptions?

Theorem (K. '07)

Let X be an isotropic random vector in $\mathbb{R}^n$, with a log-concave density.

Then there exists $\Theta \subseteq S^{n-1}$ with $\sigma_{n-1}(\Theta) \geq 1 - \exp(-\sqrt{n})$, such that for $\theta \in \Theta$, and a measurable set $A \subseteq \mathbb{R}$,

$$\left| \mathbb{P}(X \cdot \theta \in A) - \frac{1}{\sqrt{2\pi}} \int_A e^{-s^2/2} ds \right| \leq \frac{C}{n^\alpha},$$

where $C, \alpha > 0$ are universal constants.

- Without assuming that X is isotropic, there is still at least one approx. gaussian marginal, for **any** log-concave density in $\mathbb{R}^n$.   (due to linear invariance)

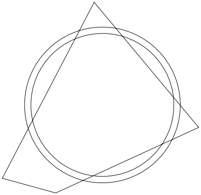
Proof ideas: a geometric approach

- Of course, a key ingredient in the proof of the central limit theorem for convex bodies is the thin-shell bound

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \le \frac{C}{n^\alpha}, \qquad (3)$$

for universal constants $C, \alpha > 0$.

Most of the volume of a convex body in high dimensions, with the isotropic normalization, is located **near a sphere.**



How can we prove (3) for a general log-concave density?

Observation

Suppose X is log-concave, isotropic and **radial**. *Then,*

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \le \frac{C}{n}.$$

Proof ideas: one-dimensional log-concavity

Explanation for the observation: The density of X is

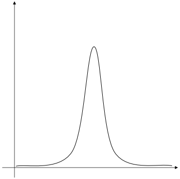
$$e^{-H(x)} = e^{-H(|x|)}.$$

Then the density of the (real-valued) random variable $|X|$ is

$$t \mapsto C_n t^{n-1} e^{-H(t)} \qquad (t > 0)$$

with H being **convex**, and $C_n = Vol_{n-1}(S^{n-1})$.

Laplace method:
Such densities are necessarily very peaked (like $t \mapsto t^{n-1} e^{-t}$).



- A straightforward computation shows that

$$\mathbb{E}\left(|X| - \sqrt{n}\right)^2 \leq C.$$

- **Our problem:** The density of our random vector X is *log-concave* and *isotropic*, but not at all *radial*.

Proof ideas: concentration of measure

We will reduce matters to the (approximately) radial case, by projecting to a random lower-dimensional subspace!

- The **Grassmannian** $G_{n,\ell}$ of all ℓ-dimensional subspaces carries a uniform probability $\sigma_{n,\ell}$. It enjoys concentration properties, as in S^{n-1} (Gromov-Milman, 1980s).

For a subspace $E \subset \mathbb{R}^n$, denote by $f_E : E \to [0, \infty)$ the log-concave density of $Proj_E(X)$.

- Fix $r > 0$, a dimension ℓ. Using the log-concavity of f, one may show that the map

$$(E, \theta) \quad \mapsto \quad \log f_E(r\theta) \qquad (E \in G_{n,\ell}, \theta \in S^{n-1} \cap E)$$

may be approximated by a Lipschitz function. Therefore, by concentration phenomenon, it is "**effectively constant**".

Completing the proof

Recall:

- X is an isotropic, log-concave random vector in $\mathbb{R}^n$.
- For a subspace $E \subset \mathbb{R}^n$, denote by $f_E : E \to [0, \infty)$ the log-concave density of $Proj_E(X)$.
- The map $(E, \theta) \quad \mapsto \quad \log f_E(r\theta)$ is "effectively constant".

Hence for most subspaces $E \in G_{n,\ell}$, the function f_E is approximately radial.

- From the radial case, for most subspaces E,

$$\mathbb{E}_X \left(\frac{|Proj_E(X)|}{\sqrt{\ell}} - 1 \right)^2 \leq \frac{C}{\ell}.$$

- Since usually $|Proj_E(X)| \approx \sqrt{\ell/n}|X|$, then

$$\mathbb{E} \left(\frac{|X|}{\sqrt{n}} - 1 \right)^2 \leq \frac{C}{\ell} \leq \frac{C}{n^\alpha}.$$

Rate of convergence

We are still lacking optimal rate of convergence results, or equivalently, optimal thin shell bounds.

- Suppose X is isotropic and log-concave in $\mathbb{R}^n$. The best available thin shell bound is

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}}-1\right)^2 \leq \frac{C}{n^\alpha}$$

  with $\alpha = 1/3$ due to Guédon-Milman '11, improving upon $\alpha = 1/4$ (Fleury '10) and $\alpha = 1/6$ (K., '07).
- Perhaps $\alpha = 1$? More reasons to care about α later on.
- In the large deviations regime, there is a **sharp** result due to Paouris '06:

$$\mathbb{P}\left(|X| \geq t\right) \leq C\exp\left(-ct\right) \qquad \text{for } t \geq C\sqrt{n},$$

All known proofs are based on the idea that “projections to a random lower dimensional subspace are roughly radial”.

Multi-dimensional CLT

Theorem (joint with R. Eldan, '08)

Let X be an isotropic random vector with a log-concave density in $\mathbb{R}^n$. Let $\ell \leq n^\alpha$. Then $\exists \mathcal{E} \subseteq G_{n,\ell}$ with $\sigma_{n,\ell}(\mathcal{E}) \geq 1 - \exp(-\sqrt{n})$, such that for all $E \in \mathcal{E}$,

1. $\displaystyle\int_E |f_E(x) - \gamma_E(x)|\, dx \leq \frac{C}{n^\alpha}.$
2. For any $x \in E$ with $|x| \leq cn^\alpha$,

$$\left| \frac{f_E(x)}{\gamma_E(X)} - 1 \right| \leq \frac{C}{n^\alpha}.$$

with f_E being the density of $Proj_E(X)$, where $\gamma_E(x) = (2\pi)^{\ell/2} \exp(-|x|^2/2)$ and $C, \alpha > 0$ are constants.

- **Milman's form of Dvoretzky's Theorem**: The geometric projection of a convex body K onto an ℓ-dimensional subspace is close to a ball, only when $\ell \leq c \log n$.

Is this really a famous open problem?

Some of the theorems mentioned so far are byproducts of failed attempts to answer the following innocent question:

Question (Bourgain, 1980s)

Suppose $K \subset \mathbb{R}^n$ is a convex body of volume one. Does there exist an $(n-1)$-dimensional hyperplane $H \subset \mathbb{R}^n$ such that

$$Vol_{n-1}(K \cap H) > c$$

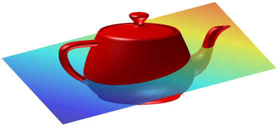
where $c > 0$ is a universal constant?

- Known: $Vol_{n-1}(K \cap H) > cn^{-1/4}$ (Bourgain '91, K. '06).
- Affirmative answer for: unconditional convex bodies, zonoids, their duals, certain random models of convex bodies, outer finite volume ratio, few vertices/facets, subspaces/quotients of L^p, Schatten class, ...

The slicing problem

This question has many equivalent formulations in terms of:

1. The minimal volume of an **ellipsoid** that captures half of the volume of a convex body.
2. The determinant of the **covariance matrix** of a convex body $K \subset \mathbb{R}^n$.
3. The probability that $n+2$ **random points** in a convex body are the vertices of a convex polytope.



Conditional Theorem (Eldan, K., '10)

Assume that for any isotropic random vector X in $\mathbb{R}^n$, distributed uniformly in a convex body,

$$\mathbb{E}\left(\frac{|X|^2}{n} - 1\right)^2 \leq \frac{C}{n}$$

(i.e., **optimal** thin-shell bound). Then, the **hyperplane conjecture** is correct.

Optimal thin shell bounds

- We do have the **optimal thin shell bound** in the case of i.i.d. random variables with finite fourth moments.

Is **convexity** as good as **independence** in the context of the quality of the Gaussian approximation?

- The relation between **thin shell bounds** and the **slicing problem** is direct. Any non-trivial bound on the thin shell leads to a non-trivial bound for the slicing problem.
- Optimal thin shell bounds are known in the presence of mild **symmetries**:
  Coordinate reflections (K. '09), symmetries of the simplex (Barthe, Cordero '11).
- There are CLT's for other forms of convexity. Say, unit balls of ℓ_p^n for $p < 1$, or densities of the form $\varphi(H)$ for convex H.

Universality beyond convexity?

- What can we say about 2D marginals of **general** probability measures on $\mathbb{R}^n$?

They can be far from Gaussian. But perhaps some marginals are approx. **spherically-symmetric**? This was suggested by Gromov '88, in analogy with Dvoretzky's Theorem.

When is a probability measure μ on $\mathbb{R}^d$ approx. radial?

1. A prob. measure μ on the sphere S^{d-1} is approx. spherically-symmetric if it is close to σ_{d-1} in, say, the W_1 Monge-Kantorovich-Wasserstein transportation metric.

$$W_1(\mu, \sigma_{d-1}) = \sup_{Lip(f) \leq 1} \int f d\mu - \int f d\sigma_{n-1}$$

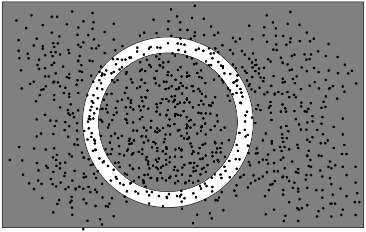
2. A prob. measure on a **spherical shell** is approx. radial if its radial projection to the sphere is approx. spherically-symmetric.

What is an approximately-radial density?

Definition (Gromov)

A probability measure μ on $\mathbb{R}^d$ is ε-radial, if for any spherical shell $S = \{a \leq |x| \leq b\} \subset \mathbb{R}^d$ with $\mu(S) \geq \varepsilon$,

- when we condition μ to the shell S, and project radially to the sphere, the resulting prob. measure is ε-close to the uniform measure on S^{d-1} in the W_1 metric.



No convexity assumptions

Theorem (K. '10)

Let μ be an absolutely continuous probability measure on $\mathbb{R}^n$, and assume that

$$n \geq \left(\frac{C}{\varepsilon}\right)^{Cd}.$$

Then, there exists a linear map that pushes μ forward to an ε-radial measure on $\mathbb{R}^d$.

- The case $d = 1$ means that the measure is approx. symmetrical on the real line.
- Gromov had a proof for the case $d = 1, 2$ which does not seem to generalize to higher dimensions.
- As opposed to all proofs discussed here, our proof of this Theorem doesn't rely so heavily on the **isoperimetric inequality**.

Super-Gaussian marginals

Most marginals are approximately spherically-symmetric, with almost no assumptions.

- In fact, we do not even have to assume that μ is absolutely continuous: It can be **discrete**, as long as "there is no low-dimensional subspace of large measure".

Corollary ("any measure has super-Gaussian marginals")

Let X be an absolutely-continuous random vector in $\mathbb{R}^n$. Then, there exists a non-zero linear functional φ on $\mathbb{R}^n$ with

$$\mathbb{P}\left(\varphi(X) \geq tM\right) \geq c\exp(-Ct^2) \quad \text{for } 0 \leq t \leq R_n,$$

$$\mathbb{P}\left(\varphi(X) \leq -tM\right) \geq c\exp(-Ct^2) \quad \text{for } 0 \leq t \leq R_n,$$

where M is a median of $|\varphi(X)|$, and $R_n = c(\log n)^{1/4}$.

Thank you!