High-dimensional distributions with convexity properties

Bo'az Klartag

Tel-Aviv University

A conference in honor of Charles Fefferman,
Princeton, May 2009

High-Dimensional Distributions

We are concerned with probability measures in high dimensions that satisfy certain geometric characteristics.

- 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. Then,

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

for appropriate coefficients $b, \theta_1, \ldots, \theta_n \in \mathbb{R}$.

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

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

we may select $\theta = (1, \ldots, 1)/\sqrt{n}$.

In this case, the gaussian approx. holds for "most" choices of $\theta_1, \ldots, \theta_n \in \mathbb{R}$ with $\sum_i \theta_i^2 = 1$.

Structure, Symmetry or Convexity?

The central limit theorem 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 take a more geometric point of view. We shall see that *convexity* conditions fit very well with the high dimensionality.

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

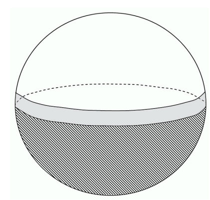
Convexity may sometimes substitute for structure and symmetries. The geometry of $\mathbb{R}^n$ forces regularity (usually, but not always, convexity is required).

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 $\varepsilon > 0$ denote

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

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



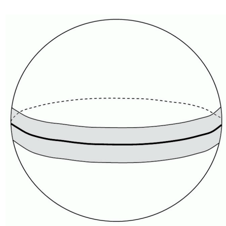
- Consider the hemisphere $H = \{x \in S^{n-1}; x_1 \le 0\}$. Then,

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

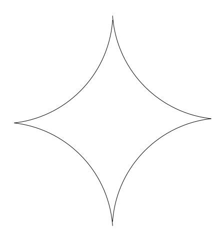
where $Y = (Y_1, \ldots, Y_n)$ is dist. according to σ_{n-1}, and Γ is a standard normal r.v.

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

"Concentration of Measure"



$dim \to \infty$



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 $\sigma_{n-1}(A) = 1/2$,

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

Corollary ("Lévy's lemma") Let $f : S^{n-1} \to \mathbb{R}$ be a 1-Lipschitz function (i.e., $f(x) - f(y) \leq d(x,y)$). Denote

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

Then, for any $\varepsilon > 0$,

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

for $c, C > 0$ universal constants.

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

Sudakov's Theorem

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 $X = (X_1, \dots, X_n)$ is "normalized" or "isotropic" if

$$\mathbb{E}X_i = 0, \quad \mathbb{E}X_i X_j = \delta_{i,j} \qquad \forall i, j = 1, \dots, n.$$

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

Theorem (Sudakov '76, Diaconis-Freedman '84,...) Let X be an isotropic r.v. in $\mathbb{R}^n$, $\varepsilon > 0$. Assume

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

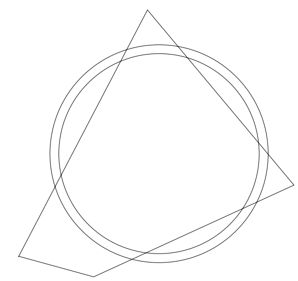
Then, there exists a subset $\Theta \subseteq S^{n-1}$ with $\sigma_{n-1}(\Theta) \geq 1 - e^{-c\sqrt{n}}$, such that for any $\theta \in \Theta$,

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

for $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-s^2/2\right) ds$.

Main assumption: Most of the mass of the random vector X is contained in a **thin spherical shell**, whose width is only ε times its radius.

This "thin shell" assumption in Sudakov's theorem is also necessary.



• Main idea in proof: *The concentration phenomenon*. Fix $t \in \mathbb{R}$. Define

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

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

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

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

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

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

Violation of Thin Shell Condition

There are isotropic distributions that violate the *thin shell assumption* and hence don't have many gaussian marginals. e.g.,

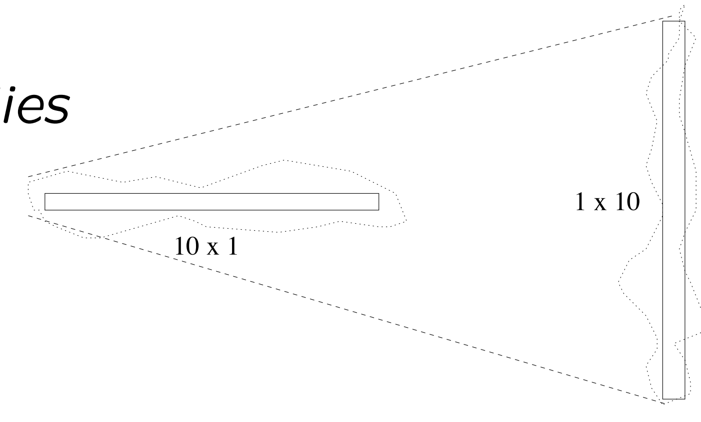
$$\frac{1}{2}\left[\sigma_{n-1}^{r_1} + \sigma_{n-1}^{r_2}\right]$$

for $r_1 = \sqrt{n}/2$ and $r_2 = \sqrt{7n}/2$, where σ_{n-1}^{r} is the uniform probability on rS^{n-1}.

• The main problem: "mixture of different scales".

It was suggested by Anttila, Ball and Perissinaki '03, and by Brehm and Voigt '00 that perhaps **<u>convexity</u>** conditions may rule out such examples.

Perhaps convex bodies are inherently of a single scale?



What's special about convex sets?

Consider the classical Brunn-Minkowski inequality (1887):

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

for any non-empty Borel sets $A, B \subset \mathbb{R}^n$.
Here $A + B = \{a + b; a \in A, b \in B\}$.

- 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 $H : \mathbb{R}^n \to (-\infty, \infty]$, a convex function.

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

The B-M inequality implies that marginals of the uniform measure on convex bodies, of all dimensions, have **log-concave** densities.

- Any marginal, of any dimension, of a log-concave density 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 common line of attack on (1): Try to prove

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

for **all** φ with $\int \varphi d\mu = 0$, with $\alpha \gg 1/n$. Our case is $\varphi(x) = |x|^2/n - 1$.

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.

• Kannan, Lovász and Simonovits conjecture: When H is convex, (2) holds with $\alpha = c$.
It is equivalent to an isoperimetric problem.

Strong Convexity Assumptions

Assume that μ is isotropic, log-concave, with density $\exp(-H)$. Then $\nabla^2 H \geq 0$. Suppose that the strong convexity assumption holds:

$$\nabla^2 H(x) \geq \delta \qquad \text{for all } x \in \mathbb{R}^n$$

for some $\delta > 0$.

Then the desired spectral-gap inequality holds with $\alpha = \delta$. We get a non-trivial *thin shell* bound as long as $\delta \gg 1/n$.

- This fact (due to Brascamp-Lieb '76) follows from Bochner-type integration by parts:

$$\begin{aligned}
&\int_{\mathbb{R}^n} (\triangle_\mu \varphi)^2 d\mu \\
&\quad = \int_{\mathbb{R}^n} |\nabla^2 \varphi|_{HS}^2 d\mu + \int_{\mathbb{R}^n} (\nabla^2 H)(\nabla \varphi) \cdot \nabla \varphi d\mu \\
&\quad \geq \delta \int_{\mathbb{R}^n} |\nabla \varphi|^2 d\mu = -\delta \int_{\mathbb{R}^n} \varphi \triangle_\mu \varphi d\mu,
\end{aligned}$$

hence $\triangle_\mu^2 \geq -\delta \triangle_\mu$ and the second eigenvalue of $-\triangle_\mu$ is at least δ.

Central Limit Theorem for Convex Sets

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

Theorem 1. Let X be an isotropic random vector in $\mathbb{R}^n$, with a log-concave density.

Then $\exists \Theta \subseteq S^{n-1}$ 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 $\mathbb{R}^n$.

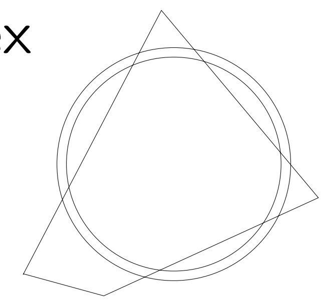
(due to linear invariance)

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

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

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

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



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

Observation: Suppose X is an isotropic random vector, whose density f is log-concave and radial. Then,

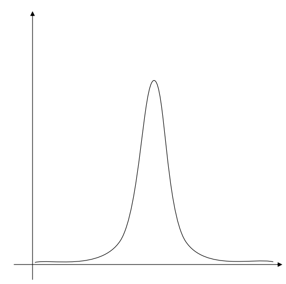
$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \leq \frac{C}{n}.$$

Explanation of the observation: Write $f(x) = f(|x|)$ for the density of X. Then the density of the (real-valued) r.v. $|X|$ is

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

with f log-concave, and $C_n = Vol_{n-1}(S^{n-1})$.

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



<u>Problem:</u> The density of our r. v. X is assumed to be log-concave, but not at all radial.

- 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 $Proj_E(X)$.

The General, Log-Concave Case

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

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

may be approximated by a Lipschitz function.

Using **concentration phenomenon** we see that the map $(E, \theta) \ \mapsto \ \log f_E(r\theta)$ is "effectively constant".

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

From the already-established radial, log-concave case, for most subspaces E,

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

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

$$\mathbb{E}\left(\frac{|X|}{\sqrt{n}} - 1\right)^2 \leq \frac{C}{\ell} \leq \frac{C}{n^\alpha} \qquad (\alpha \approx 1/5).$$

Rate of Convergence

We are still lacking optimal rate of convergence results and optimal thin shell bounds. The best available thin shell bound is

$$\mathbb{P}\left(\left|\frac{|X|}{\sqrt{n}} - 1\right| \geq t\right) \leq C \exp\left(-cn^{\alpha}t^{\beta}\right) \quad \text{for } 0 < t < 1,$$

with, say, $\alpha = 0.33$ and $\beta = 3.33$, where $c, C > 0$ are universal constants. Probably non-optimal.

• In the large deviations regime, there is a *sharp* result, with the right exponent.

Theorem 2. (Paouris '06) For an isotropic, log-concave random vector X in $\mathbb{R}^n$,

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

for $C > 0$ a universal constant.

• Paouris observed that the "effective support" of the density of $Proj_E(X)$ is typically approx. a Euclidean ball, for $\dim(E) \sim \sqrt{n}$.

Multi-Dimensional CLT

Theorem 3. (joint with R. Eldan)
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 $E \in \mathcal{E}$ and a set $A \subseteq E$,

$$\left| \mathbb{P}(Proj_E(X) \in A) - \int_A \varphi_E(x) dx \right| \leq \frac{C}{n^\alpha},$$

where $\varphi_E(x) = (2\pi)^{-\ell/2} \exp(-|x|^2/2)$.

Moreover, denote by f_E the density of $Proj_E(X)$. Then for any $x \in E$ with $|x| \leq cn^\alpha$,

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

Here, $C, c, \alpha > 0$ are universal constants.

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

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?

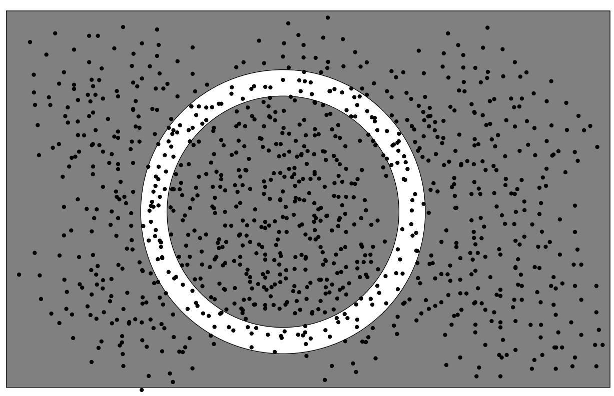
(suggested by Gromov '88, in analogy with Dvoretzky's Theorem)

- When is a probability measure μ on $\mathbb{R}^d$ approximately radial?
  - A prob. measure on the sphere S^{d-1} is approx. spherically-symmetric if it is close to σ_{d-1} in, say, the W_1 Monge-Kantatovich transportation metric.
  - A prob. measure on a spherical shell is approx. radial if its radial projection to the sphere is approx. spherically-symmetric.

No Convexity Assumptions

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.



Theorem 4. Let μ be an absolutely continuous probability measure on $\mathbb{R}^n$, and let

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

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 $d = 1, 2$. As opposed to all proofs discussed here, our proof of Theorem 4 doesn't rely so heavily on the *isoperimetric inequality*.

Do we have to assume that μ is absolutely continuous?

Example. Take μ to be a combination of a gaussian measure and several atoms. None of the marginals are approx. radial.

Definition. A prob. measure μ on $\mathbb{R}^n$ is "decently high-dimensional with accuracy δ", or $1/\delta$-*dimensional* in short, if

$$\mu(E) \leq \delta \dim(E)$$

for any subspace $E \subseteq \mathbb{R}^n$.

We say that μ is *decent* if it is n-dimensional. Of course, all absolutely-continuous measures are decent, as well as many discrete measures.

Theorem 4'. Let μ be a decent probability measure on $\mathbb{R}^n$, and let

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

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

(If μ is $1/\delta$-dim., then we can take $\varepsilon = c\delta^{c/d}$).

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

Corollary ("any high-dim. measure has super-Gaussian marginals"). Let X be a decent random vector in $\mathbb{R}^n$. Then, there exists a non-zero linear functional φ on $\mathbb{R}^n$ with

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

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

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

(perhaps $R_n = c(\log n)^{1/2}$, but no better).

Almost Sub-Gaussian Estimates

We can't have upper bounds, without convexity assumptions. Suppose X is uniform in a convex body in $\mathbb{R}^n$.
A classical fact (follows from Brunn-Minkowski):

<u>Theorem</u> (Borell '74): For any linear functional $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $t \geq 0$,

$$\mathbb{P}\{|\varphi(X)| \geq t\,\mathbb{E}|\varphi(X)|\} \leq C \exp(\,-ct\,)$$

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

A uniformly subexponential tail. This is sharp, as shown by the example of a truncated cone.

Suppose X is uniform in a centered ellipsoid. Then, for all linear functionals φ,

$$\mathbb{P}\{|\varphi(X)| \geq t\,\mathbb{E}|\varphi(X)|\} \leq C \exp\left(-ct^2\right).$$

Moreover, the tail is very close to being gaussian.

Question: Is it true that for any convex body there is a linear functional with a uniformly sub-gaussian tail?

(if true, a convex body cannot display "cone-type" behavior in all directions)

True for unconditional convex bodies (Bobkov-Nazarov '03) and for zonoids (Paouris '03). For arbitrary convex bodies:

Theorem (K. '05, Giannopoulos, Pajor, Paouris '06):

Suppose X is uniform in a convex set. Then there exists a non-zero linear functional $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that for any $t \geq 1$,

$$\mathbb{P}\{|\varphi(X)| \geq t\, \mathbb{E}|\varphi(X)|\} \leq C \exp\left(-c\frac{t^2}{\log^2(t+1)}\right)$$

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

Unconditional Convex Bodies

Suppose that our log-concave density $f : \mathbb{R}^n \to [0, \infty)$ is "unconditional":

$$f(x_1, ..., x_n) = f(|x_1|, ..., |x_n|) \qquad \forall x \in \mathbb{R}^n.$$

- In the unconditional case, we can identify some approx. gaussian marginals, and also prove a <u>sharp</u> thin shell estimate.

<u>Theorem:</u>

Suppose X is an isotropic random vector in $\mathbb{R}^n$, with an unconditional, log-concave density. Then, for any $t \in \mathbb{R}$,

$$\left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \le t \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-s^2/2} ds \right| \le \frac{C}{n}$$

and more generally, for any $(\theta_1, \ldots, \theta_n) \in S^{n-1}$,

$$\left| \mathbb{P}\left(\sum_{i=1}^n \theta_i X_i \le t \right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-s^2/2} ds \right| \le C \sum_{i=1}^n \theta_i^4.$$

Additionally, for $t \in [0,1]$ let us define

$$Y_t = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor tn \rfloor} X_j.$$

The stochastic process $(Y_t)_{0 \le t \le 1}$ converges to the standard Brownian motion.

The proof of the optimal bounds in the unconditional case relies, of course, on an optimal thin shell bound:

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

It is proven using a Bochner type formula and some L^2 technique.