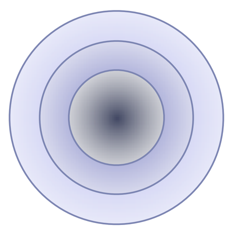
A concentration theorem for projections

Sanjoy Dasgupta, Daniel Hsu, and Nakul Verma
{dasgupta,djhsu,naverma}@cs.ucsd.edu
University of California, San Diego

Result (informally)

Suppose $X \in \mathbb{R}^D$ has mean zero and finite second moments. Choose $d \ll D$.
Then: under all but an $e^{-\Omega(D/d^2)}$ fraction of linear projections from $\mathbb{R}^D$, the projected distribution of X is very nearly the scale mixture of spherical Gaussians

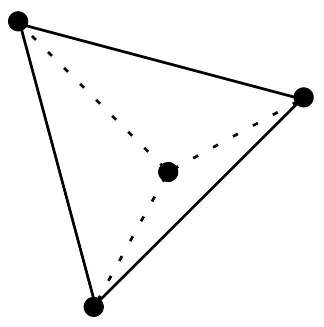
$$\int N(0, \sigma^2 I_d)\nu(d\sigma)$$



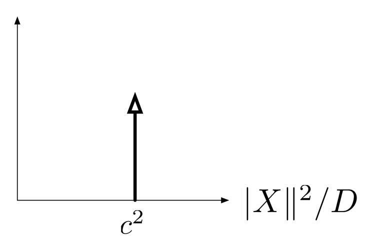
where $\nu(\cdot)$ is the distribution of $\|X\|^2/D$.
The extent of this effect depends on a coefficient of eccentricity of X's distribution.

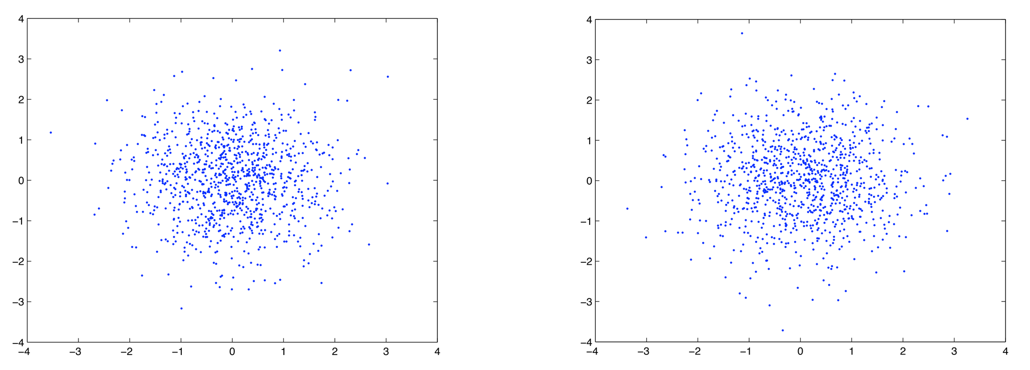
Example

Simplex in $\mathbb{R}^D$: uniform distribution over $D+1$ points.



Distribution of $\|X\|^2/D$ is:





1001 points drawn from a Gaussian — Projection of the simplex in $\mathbb{R}^{1000}$

Result (formally)

Suppose $X \in \mathbb{R}^D$ has mean zero and finite second moments. Choose $d \ll D$.

For any linear projection $\Theta : \mathbb{R}^D \to \mathbb{R}^d$, let F_Θ be the projected distribution, and let

$$F = \int N(0, \sigma^2 I_d)\nu(d\sigma)$$

where $\nu(\cdot)$ is the distribution of $\|X\|^2/D$.

Then: for any $0 < \epsilon < 1$,

$$\mathbb{P}_\Theta \left[\sup_{\text{balls } B \subseteq \mathbb{R}^d} |F_\Theta(B) - F(B)| > \epsilon \right] \leq \exp\left\{ -\tilde{\Omega}\left(\frac{\epsilon^4 D}{d^2} \cdot \frac{1}{\text{ecc}(X)} \right) \right\}$$

where $\text{ecc}(X)$ is a measure of the eccentricity of X's distribution.

Previous results

Diaconis and Freedman (1984): Let $X = (X_1, \ldots, X_D)$ have a product distribution (i.e. independent coordinates). Then, under all but an $e^{-\Omega(D)}$ fraction of linear projections from $\mathbb{R}^D$ to $\mathbb{R}$, the projected distrubtion of X is nearly Gaussian.

Different approach: Projections of more general distributions to $\mathbb{R}^1$ studied by:

Sudakov (1978)
Von Weisacker (1997)
Bobkov (2002)

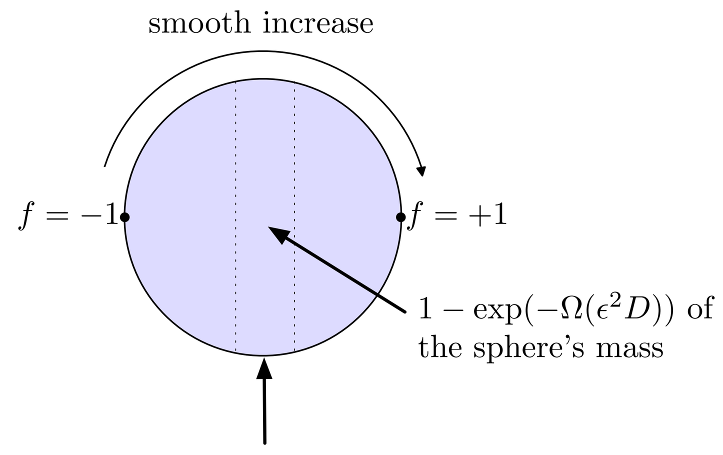
Analyses not based on CLT; use "concentration of measure" bounds. We build upon all these results.

Our contribution: how to handle (1) general distributions and (2) $d > 1$.

Concentration of measure

Contrast:

- Chernoff-bound application:
  Let $X_1, \ldots, X_D$ be independent tosses of a coin with bias p. Then
  $$\mathbb{P}\left(\left| \frac{X_1 + \ldots + X_D}{D} - p \right| > \epsilon \right) \leq 2e^{-\epsilon^2 D}.$$
- Modern concentration of measure result:
  Let f be a Lipschitz function on a sphere in $\mathbb{R}^D$. Then f is within ϵ of its median value on all but an $\exp(-\Omega(\epsilon^2 D))$ fraction of the sphere.



f is within ϵ of its median value.

What special feature makes the concentration property of the average hold in the Chernoff bound?
Answer: the average is Lipschitz!

In fact, for any Lipschitz f (under certain measures),

$$\mathbb{P}(|f - \mathbb{E}f| > \epsilon) \leq e^{-\Omega(\epsilon^2 D)}.$$

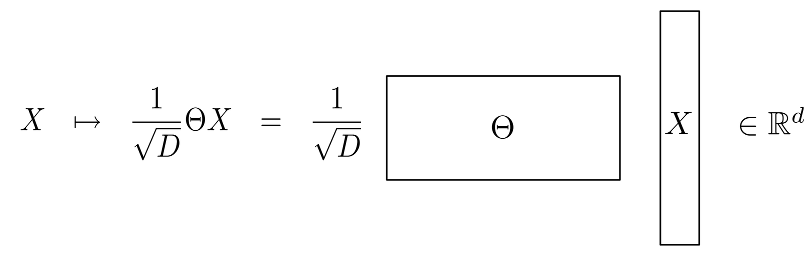
Proof outline

Sources of randomness

Two random objects:

- $X \in \mathbb{R}^D$, high dimensional distribution
- Projection $\Theta : \mathbb{R}^D \to \mathbb{R}^d$, $d \times D$ matrix

Random choice of Θ: pick each entry i.i.d. from $N(0, 1)$.



Approach

We'll show: for almost all Θ, the projected distribution

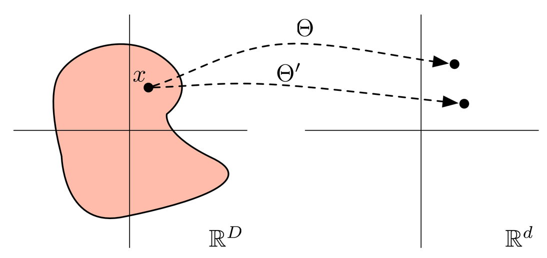
$$\frac{1}{\sqrt{D}}\Theta X \overset{\text{dist}}{\approx} F = \int N(0, \sigma^2 I_d)\nu(d\sigma).$$

In particular, for all balls $B \subseteq \mathbb{R}^d$: the projected distribution and F assign roughly the same mass to B.

Distribution of projected points

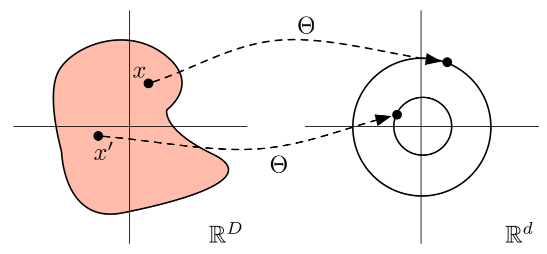
Fixed $x \in \mathbb{R}^D$. As Θ varies, the distribution of the projection of x is Gaussian, since Θ is Gaussian.

$$x \mapsto \frac{1}{\sqrt{D}}\Theta x \overset{\text{dist}}{=} N\left(0, \frac{\|x\|^2}{D} I_d\right)$$



Now pick both X and Θ at random. Resulting distribution is

$$F = \int N(0, \sigma^2 I_d)\nu(d\sigma).$$



Part I: a single ball

Fix a ball $B \subseteq \mathbb{R}^d$. Let $F_B(\Theta)$ be the probability mass that falls in B under projection Θ.
Claim: For this ball B, for almost all Θ, $F_B(\Theta) \approx F(B)$.

1. $\mathbb{E}_\Theta[F_B(\Theta)] = F(B)$.
2. Want to conclude that $F_B(\Theta) \approx F(B)$ for almost all Θ, but doesn't following from concentration of measure bounds since $F_B(\cdot)$ is not Lipschitz.
3. Consider a Lipschitz approximation G_B of F_B. We show:
   a. $G_B(\Theta) \approx \mathbb{E}_\Theta[G_B(\Theta)]$ for almost all Θ (by concentration of measure).
   b. G_B is not too different from F_B.

Part II: all balls

Determine a collection of balls $B_1, \ldots, B_M \subseteq \mathbb{R}^d$ with the following property: if the projected distribution assigns mass $F(B_i) \pm \varepsilon$ to every B_i, then it assigns mass $F(B) \pm 2\varepsilon$ to every ball $B \subseteq \mathbb{R}^d$. (M depends only on d, not D.)
Then take a union bound over B_i, and we're done.

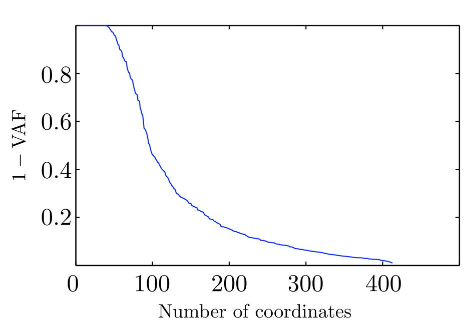
For any $0 < \epsilon < 1$

$$\mathbb{P}_\Theta \left[\sup_{\text{balls } B \subseteq \mathbb{R}^d} |F_\Theta(B) - F(B)| > \epsilon \right] \leq \exp\left\{ -\tilde{\Omega}\left(\frac{\epsilon^4 D}{d^2} \cdot \frac{1}{\text{ecc}(X)} \right) \right\}$$

Implications

Gaussians and dimensionality

Low dimension ($D = 1, 2, 3$): many naturally occuring data look Gaussian.
High dimension: too much independence required to be true!



For each coordinate in the MNIST 1's dataset, the plot shows the fraction of variance unaccounted for by the best affine combination of preceding coordinates. The ordering of coordinates is chosen greedily, by least VAF.

Nevertheless, most low-dimensional projections appear Gaussian, at least in terms of low-order statistics.

Mixture modeling

A randomized reduction to "well-behaved" data.