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
\[ \int N(0, \sigma^2 I_d) \nu(\sigma) \, d\sigma \]
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:

![Diagram showing distribution](image)

1001 points drawn from a Gaussian Projection of the simplex in $\mathbb{R}^3$

**Result (formally)**

Suppose $X \in \mathbb{R}^D$ has mean zero and finite second moments. Choose $d \ll D$. For any linear projection $\Theta : \mathbb{R}^D \rightarrow \mathbb{R}^d$, let $F_{\Theta}$ be the projected distribution, and let
\[ F = \int N(0, \sigma^2 I_d) \nu(\sigma) \, d\sigma \]
where $\nu(\cdot)$ is the distribution of $\|X\|^2/D$. Then: for any $0 < \epsilon < 1$,
\[ P_{\Theta} \left[ \sup \{ \text{balls } B \subseteq \mathbb{R}^d \} \, |F_{\Theta}(B) - F(B)| > \epsilon \right] \leq \exp \left\{ -\Omega \left( \frac{d^2}{\epsilon^2 \text{ecc}(X)} \right) \right\} \]
where $\text{ecc}(X)$ is a measure of the eccentricity of $X$’s distribution.

**Gaussians and dimensionality**

Low dimension ($D = 1, 2, 3$): many naturally occurring 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.

A concentration theorem for projections

Sanjoy Dasgupta, Daniel Hsu, and Nakul Verma

{dasgupta, djhsu, naverna}@cs.ucsd.edu

University of California, San Diego