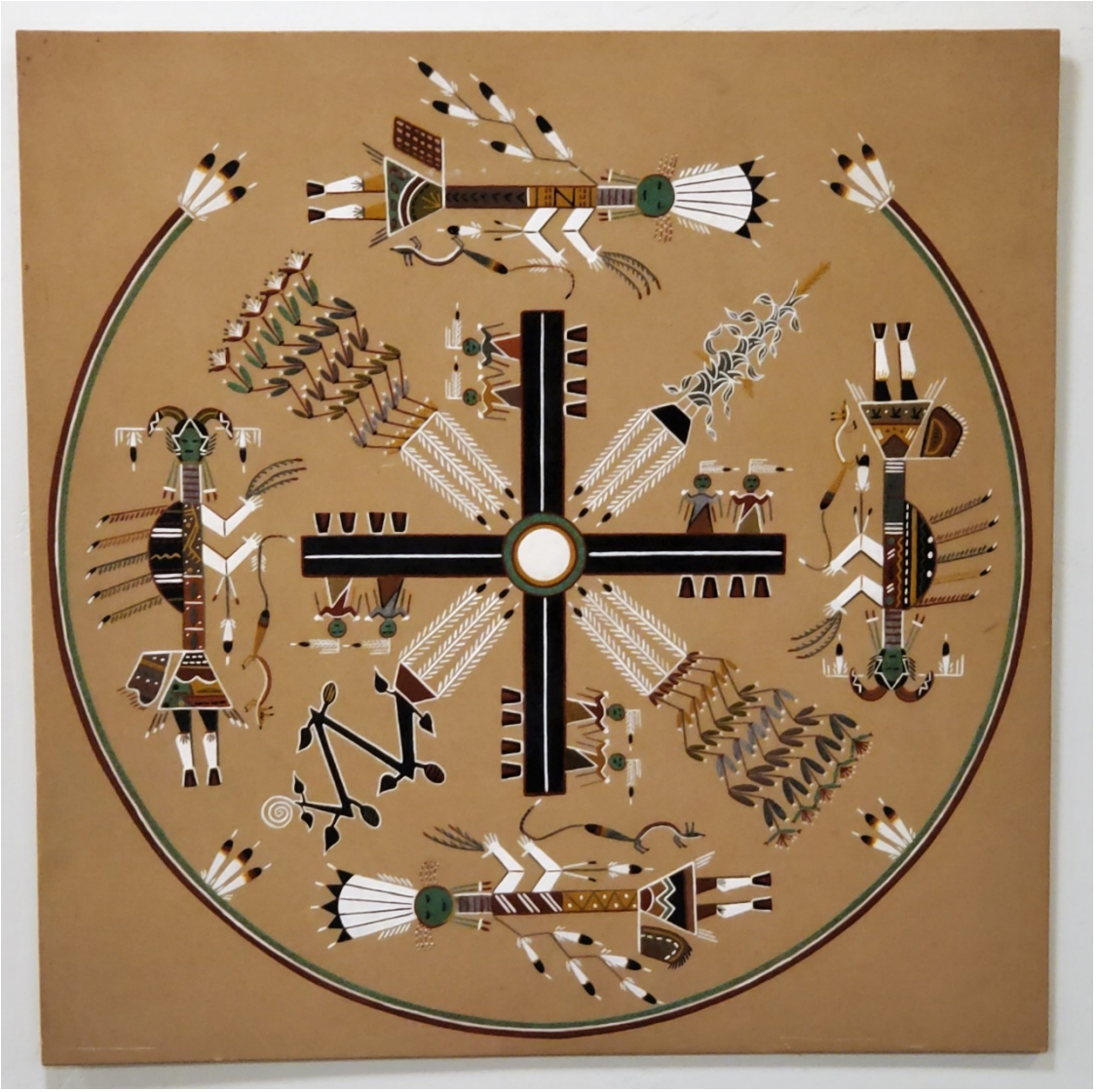
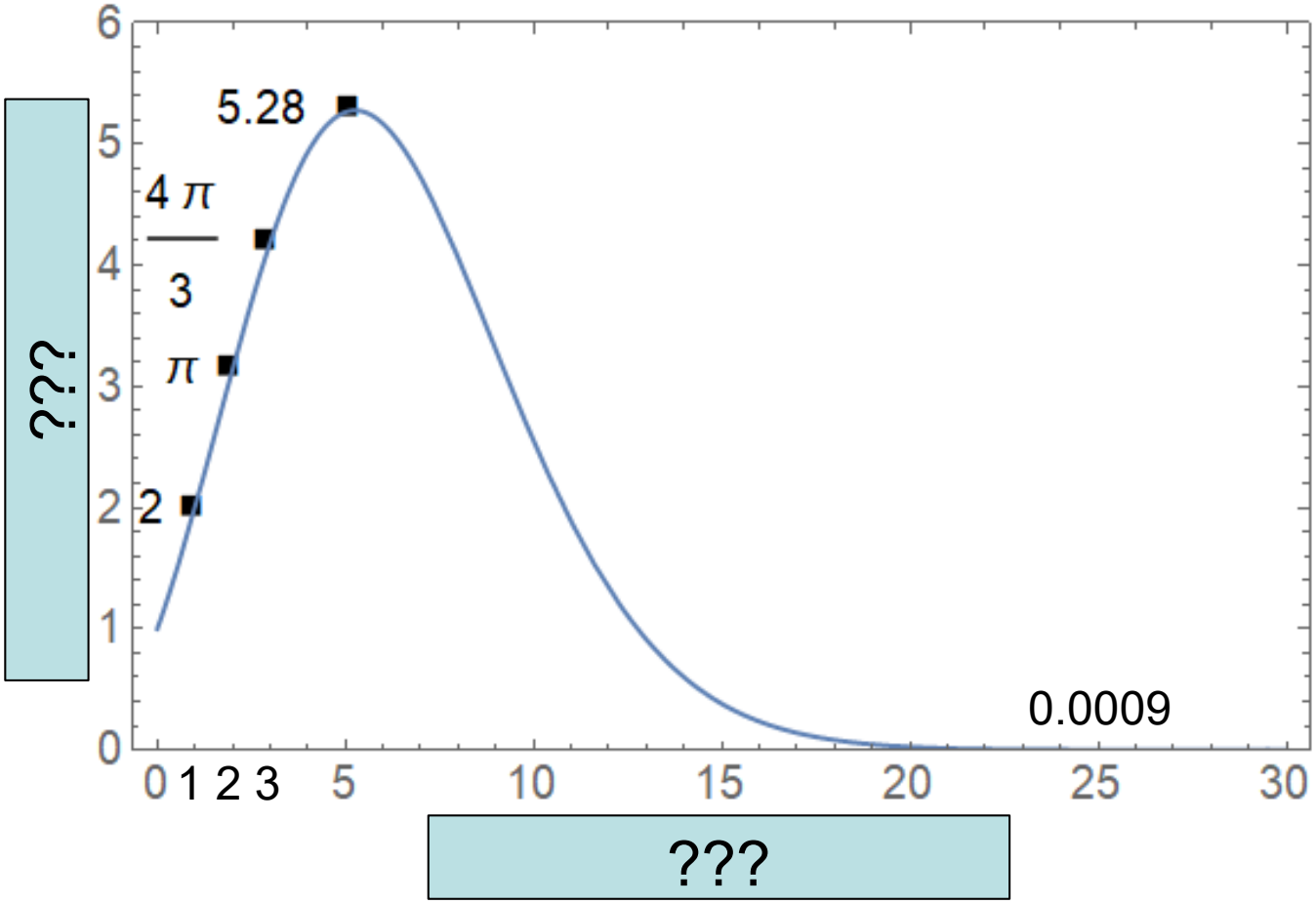


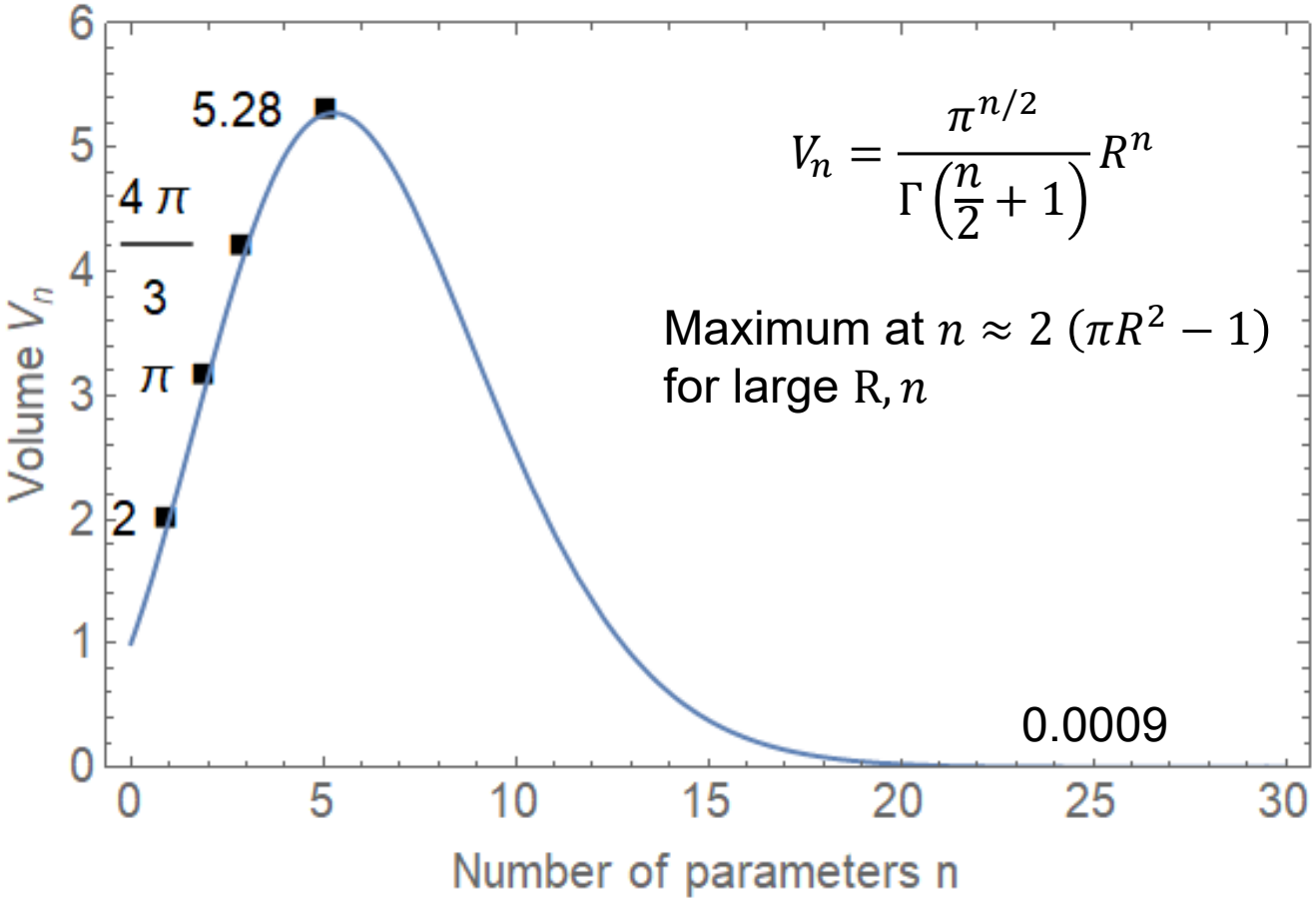
PAVEL NADOLSKY

THE CURSE OF WORLDVIEWS

Label the axes (15 seconds)



Volume of a unit ball in n dimensions



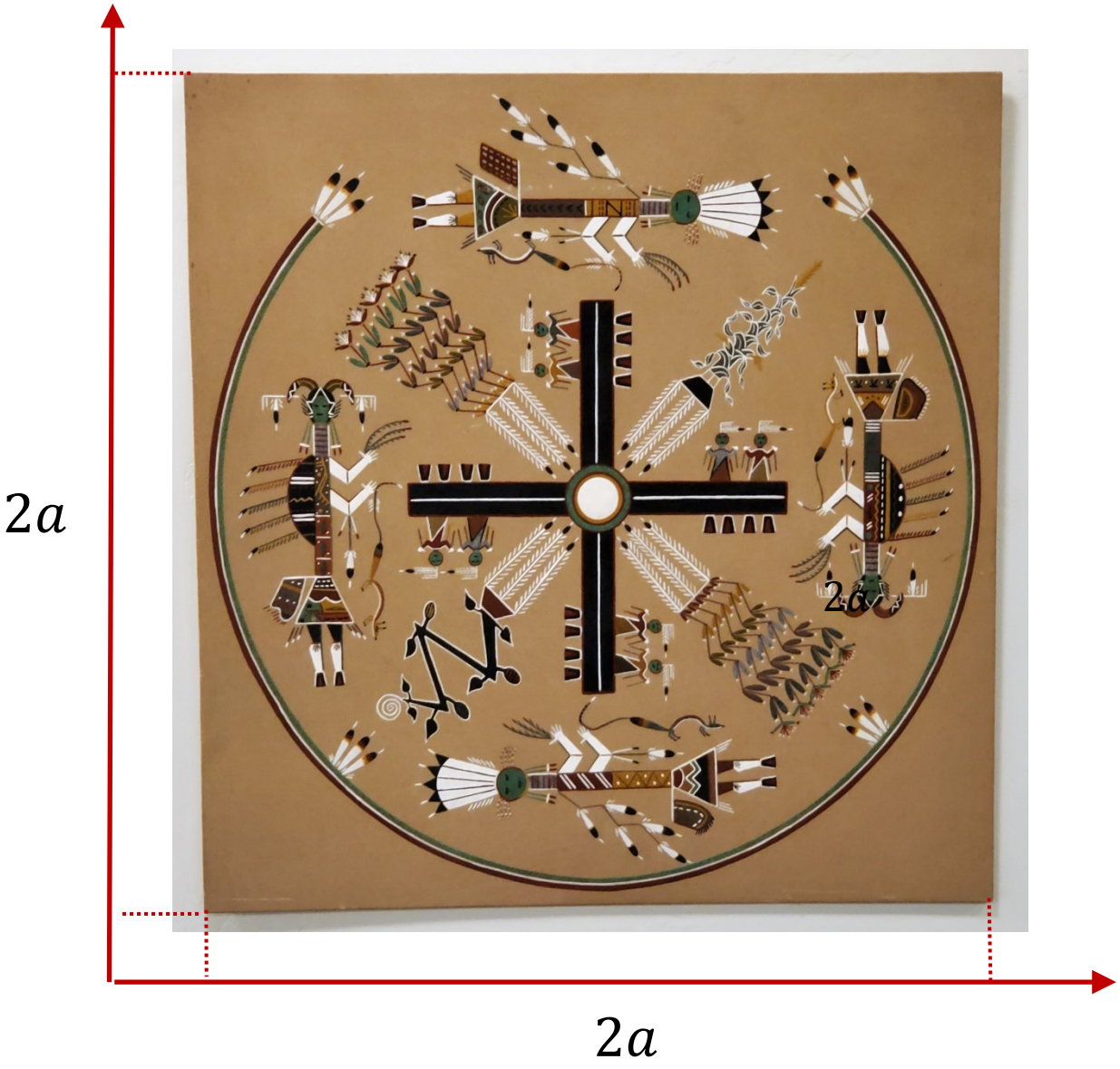
$$V_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} R^n$$

Maximum at $n \approx 2\,(\pi R^2 - 1)$
for large R, n

The Curse of Dimensionality!

Consequences of the curse of dimensionality for n ›10

1. Multivariate geometry is weird
2. Sampling of large-n distributions is highly inefficient
3. The Euclidean L_2 norm does not properly quantify the distances between neighbors for large n



Compare:

- the volume of a cube with side $2a$
- the volume of a sphere with radius a
- n=2

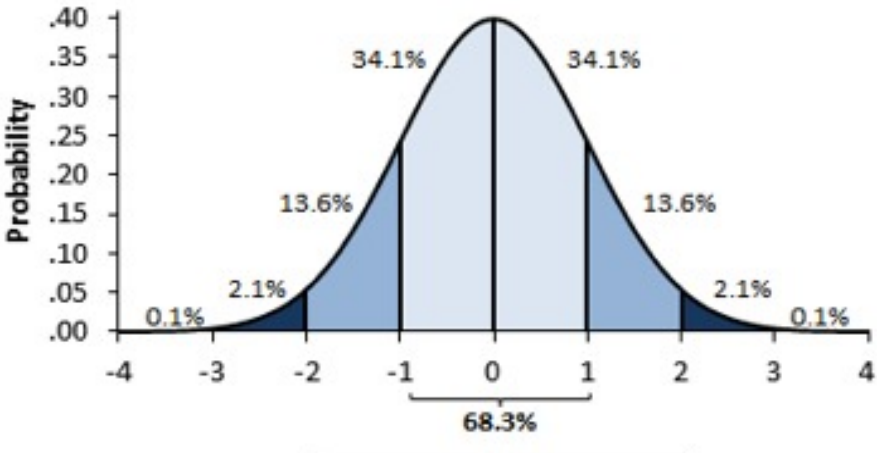
$$\frac{V_{sphere}}{V_{cube}} = \frac{\pi}{4} \approx \mathbf{0.8}$$

- n=25

$$\frac{V_{sphere}}{V_{cube}} \approx \frac{0.0009}{2^{25}} \approx \mathbf{3 \cdot 10^{-11}}$$

An n-dimensional **standard normal** distribution

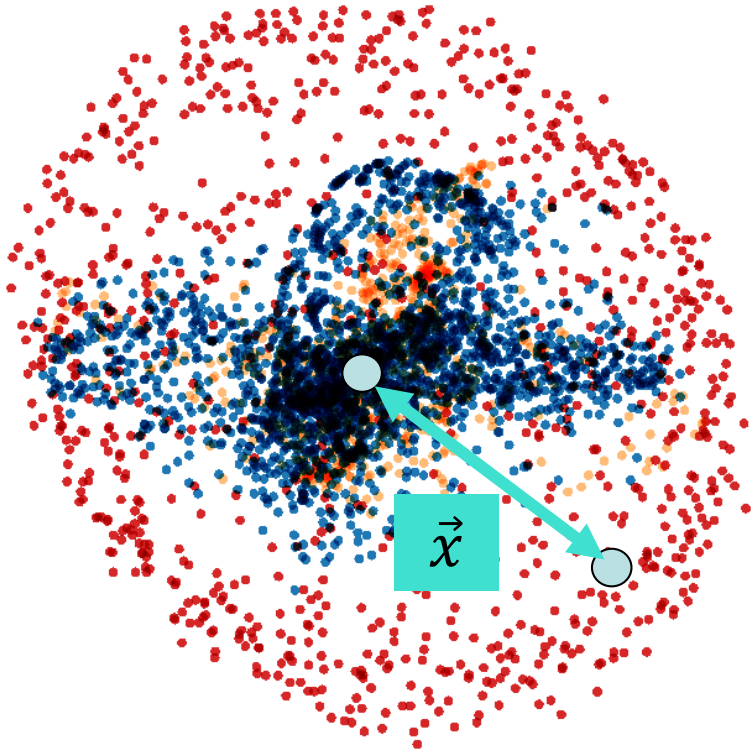
$$P(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\vec{x}^2}{2}\right)$$



Any 1-dim. projection contains 68% of the elements in the interval
$-1 < x_i < 1$

An n-dimensional **standard normal** distribution

$$P(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\vec{x}^2}{2}\right)$$

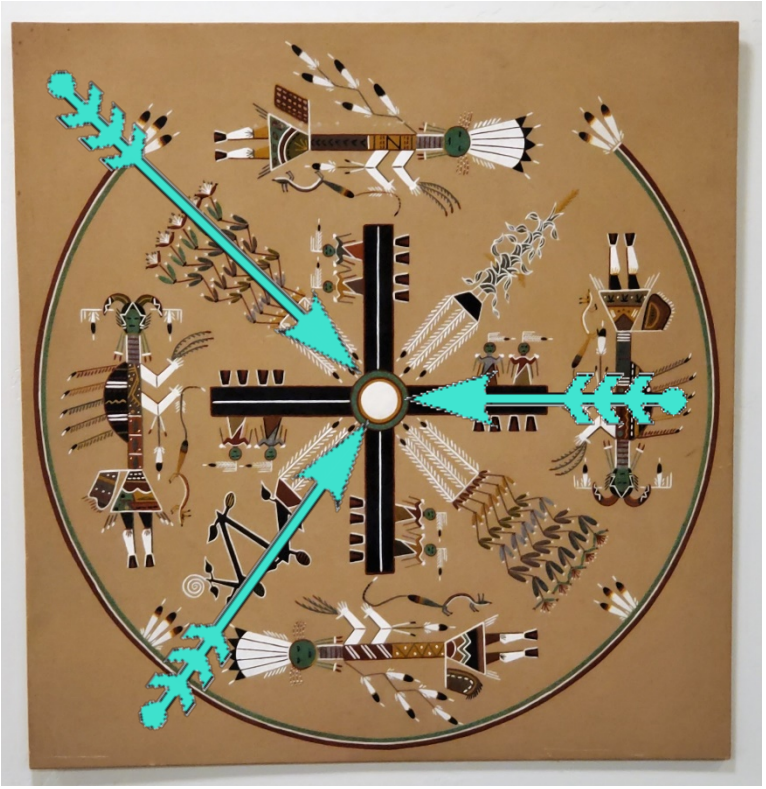


The mean distance of an element from the center ("truth") at $\vec{x} = 0$ is

$$\langle|\vec{x}|\rangle \approx \sqrt{n}$$

$\sqrt{n} \approx 5$ for $n = 25$

In a large-n **normal** distribution, a single element is likely to be very **abnormal** (be $\sim \sqrt{n}\,\sigma$ away from the "truth") in some direction(s)



Probability distribution of a "worldview"

When we reconstruct a **"worldview"** (a model with many parameters) from its low-dimensional projections:

- for a normal and other common distributions, a randomly selected reconstruction is highly likely to be far from the "truth" (the mean) in some way
- The "truth" can be reconstructed precisely by finding the average of individual reconstructions

The “truth” is an imperfect descriptor of a large-n probability distribution.

Be on the lookout for hidden dimensions! The inference of an n-dim distribution can transmogrify with the number of parameters or when using a different distance metric