Hyperdimensional Sparse Space

Some Mathematical Properties

Simon Beaumont

From: *Pentti Kanerva - Sparse Distributed Memory, MIT Press 1988, ISBN 978-0-262-51469-9*

For a rigorous treatment see:

Fredrik Sandin and Blerim Emruli and Magnus Sahlgren, Incremental dimension reduction of tensors with random index, http://arxiv.org/abs/1103.3585

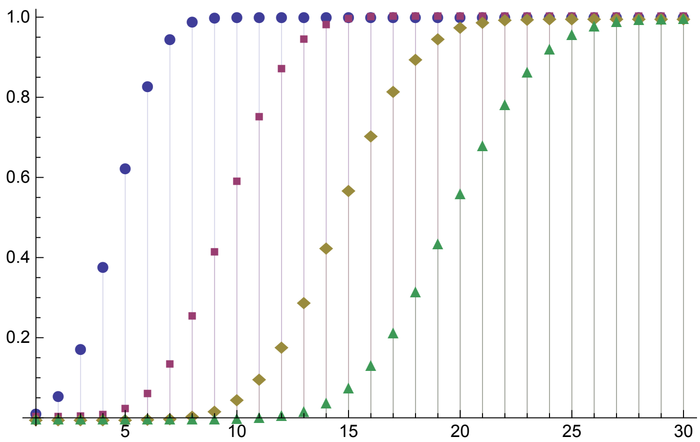
In the binary space $\{0, 1\}^n$ of vectors of length n with equal probability of the states 0 and 1 with a distance metric defined as the *Hamming-distance* or the number of non-zero bits in the *xor* of vectors, (equivalent to the square of the Euclidean distance), then the number of vectors in the space that are at a distance d from a specific vector is given by the binomial coefficient:

```
NumberVectorsAtDistance[n_, d_] := Binomial[n, d]
```

The number of vectors at a certain distance from a reference vector therefore is given by the binomial distribution with probability $p = \frac{1}{2}$, which has mean $\frac{n}{2}$ and variance $\frac{n}{4}$

```
DistanceFromGivenVector[n_, p_] := BinomialDistribution[n, p]
```

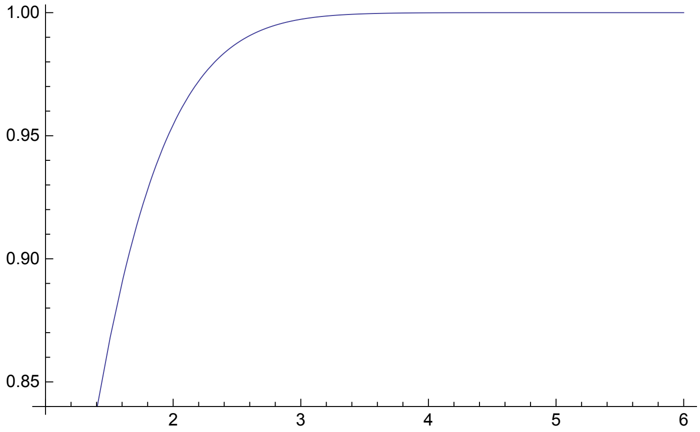
```
DiscretePlot[
 Evaluate@Table[CDF[DistanceFromGivenVector[n, 0.5], k], {n, {10, 20, 30, 40}}],
 {k, 30}, PlotRange → All, PlotMarkers → Automatic]
```



For large values of n the binomial distribution can be approximated with with a normal distribution; therefore the proportion of vectors within z standard deviations of the mean is $\text{erf}\left(\frac{z}{\sqrt{2}}\right)$. The distance distribution is highly concentrated around the mean as the error function quickly approaches unity for increasing *z:*

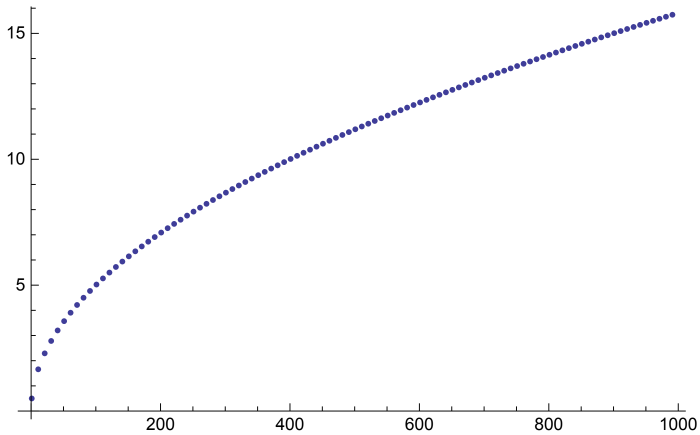
```
ProportionOfVectorsWithinStdDeviation[z_] := Erf[z / √2]
```

```
Plot[ProportionOfVectorsWithinStdDeviation[z], {z, 1, 6}]
```

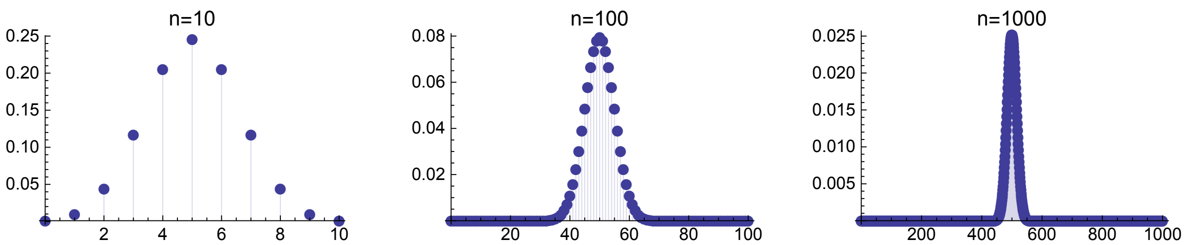


The mean distance is $n/2$ and the standard deviation of distance is $\frac{\sqrt{n}}{2}$. This implies that the mean distance is $n/2$ standard deviations, thus in a high-dimensional binary space all vectors are located at a distance ~$n/2$ from any other vector.

```
ListPlot[Table[{n, √n / 2 // N}, {n, 1, 1000, 10}]]
```




```
GraphicsRow[Table[DiscretePlot[PDF[BinomialDistribution[n, 0.5], k],
    {k, 0, n}, PlotRange → All, PlotMarkers → Automatic,
    PlotLabel → StringForm["n=``", n]], {n, {10, 100, 1000}}]]
```



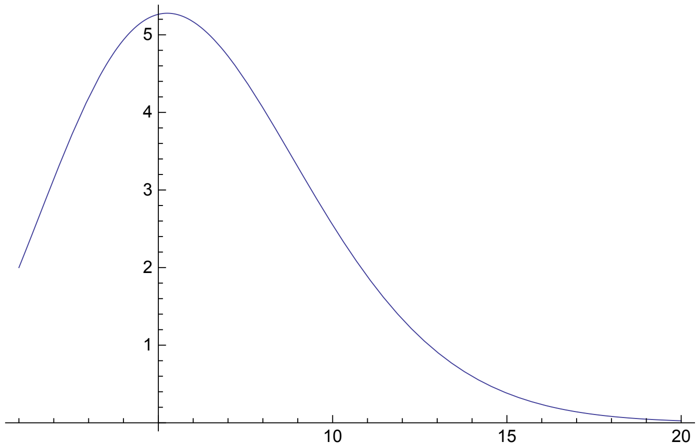
The volume (content) of the n-cube and n-sphere....

```
VolumeOfUnitNCube[n_] := 1^n
```

```
VolumeOfUnitNSphere[n_] := Pi^(n/2) / Gamma[n/2 + 1]
```

The volume of a hypersphere is interesting itself ... and actually -> 0 as n -> ∞ but is greatest at just over five dimensions!

```
Plot[VolumeOfUnitNSphere[n], {n, 1, 20}]
```



All the space is in the corners? Well there’s no space in the middle anyway...

```
ListPlot[Table[Log[VolumeOfUnitNCube[n] / VolumeOfUnitNSphere[n]], {n, 1, 80, 1}]]
```