The counter-intuitive behaviour of high-dimensional spaces

Mark Wildon

Slides are online (two equivalent links):

- http://www.ma.rhul.ac.uk/~uvah099/Talks/HighDimensionalSpaces.pdf
- https://tinyurl.com/y8mptbej

I will probably have to turn off incoming video and mute everyone except Stefanie to conserve bandwidth. Please type in the chat box if you have a question and Stefanie will alert me if I miss it.

Outline and the wild claim

§1 Euclidean spaces $\mathbb{R}^n$: spheres and balls

§2 Binary codes $\mathbb{F}_2^n$: the geometry of Hamming balls

Outline and the wild claim

§1 Euclidean spaces $\mathbb{R}^n$: spheres and balls

§2 Binary codes $\mathbb{F}_2^n$: the geometry of Hamming balls

Quiz. Please order the following numbers:

$$2^{2^{2^{100}}}, 2^{2^{3^{100}}}, 2^{3^{2^{100}}}, 3^{2^{2^{100}}}.$$

Outline and the wild claim

§1 Euclidean spaces $\mathbb{R}^n$: spheres and balls

§2 Binary codes $\mathbb{F}_2^n$: the geometry of Hamming balls

Quiz. Please order the following numbers:

$$2^{2^{2^{100}}}, 2^{2^{3^{100}}}, 2^{3^{2^{100}}}, 3^{2^{2^{100}}}.$$

Answer. $2^{2^{2^{100}}} < 3^{2^{2^{100}}} < 2^{3^{2^{100}}} < 2^{2^{3^{100}}}$.

Outline and the wild claim

§1 Euclidean spaces $\mathbb{R}^n$: spheres and balls

§2 Binary codes $\mathbb{F}_2^n$: the geometry of Hamming balls

Quiz. Please order the following numbers:

$$2^{2^{2^{100}}}, 2^{2^{3^{100}}}, 2^{3^{2^{100}}}, 3^{2^{2^{100}}}.$$

Answer. $2^{2^{2^{100}}} < 3^{2^{2^{100}}} < 2^{3^{2^{100}}} < 2^{2^{3^{100}}}$.

- Rule of thumb: all that matters is the number at the top.
- In this spirit:
  - $\mathbb{F}_2^{256}$ is a finite set and $\mathbb{R}^3$ is infinite.
  - But there is a sense in which $\mathbb{F}_2^{256}$ is still the 'larger' space.

§1 Euclidean space.

Flatland (1884) by Edwin Abbott is

(a) A stinging satire of Victorian society

- Are you an isosceles triangle with a smaller angle of 59.5°? Sorry, you are a upper-lower middle class tradesman. Maybe your children will be lucky enough to be equilateral and go to university.

§1 Euclidean space.

Flatland (1884) by Edwin Abbott is

(a) A stinging satire of Victorian society

- Are you an isosceles triangle with a smaller angle of 59.5°? Sorry, you are a upper-lower middle class tradesman. Maybe your children will be lucky enough to be equilateral and go to university.
- Are you a hexagon? Congratulations, you are upper-middle class man and have a life of privilege.

§1 Euclidean space.

Flatland (1884) by Edwin Abbott is

(a) A stinging satire of Victorian society

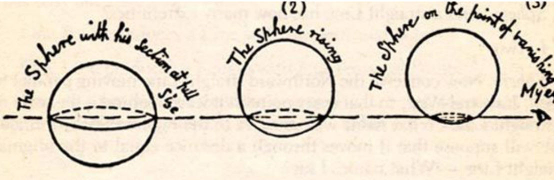
- Are you an isosceles triangle with a smaller angle of 59.5°? Sorry, you are a upper-lower middle class tradesman. Maybe your children will be lucky enough to be equilateral and go to university.
- Are you a hexagon? Congratulations, you are upper-middle class man and have a life of privilege.
- Are you a line segment? [The appalling truth this reveals about Victorian society will be revealed verbally.]

§1 Euclidean space.

Flatland (1884) by Edwin Abbott is

(a) A stinging satire of Victorian society
- Are you an isosceles triangle with a smaller angle of 59.5°? Sorry, you are a upper-lower middle class tradesman. Maybe your children will be lucky enough to be equilateral and go to university.
- Are you a hexagon? Congratulations, you are upper-middle class man and have a life of privilege.
- Are you a line segment? [The appalling truth this reveals about Victorian society will be revealed verbally.]

(b) A nice introduction to geometric reasoning by analogy

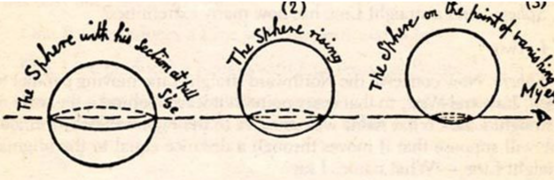


§1 Euclidean space.

Flatland (1884) by Edwin Abbott is

(a) A stinging satire of Victorian society
  - Are you an isosceles triangle with a smaller angle of 59.5°? Sorry, you are a upper-lower middle class tradesman. Maybe your children will be lucky enough to be equilateral and go to university.
  - Are you a hexagon? Congratulations, you are upper-middle class man and have a life of privilege.
  - Are you a line segment? [The appalling truth this reveals about Victorian society will be revealed verbally.]

(b) A nice introduction to geometric reasoning by analogy



(c) Highly recommended.

n-Sphereland

Let $B^n = \{x \in \mathbb{R}^n : ||x|| < 1\}$ be the solid n-dimensional unit ball and let

$$S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$$

be the n-dimensional sphere: it is the surface of B^{n+1}.

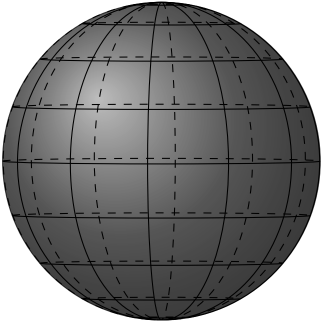
n-Sphereland

Let $B^n = \{x \in \mathbb{R}^n : ||x|| < 1\}$ be the solid n-dimensional unit ball and let

$$S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$$

be the n-dimensional sphere: it is the surface of B^{n+1}.

In 'n-Sphereland' the inhabitants are uniformly distributed on S^n.



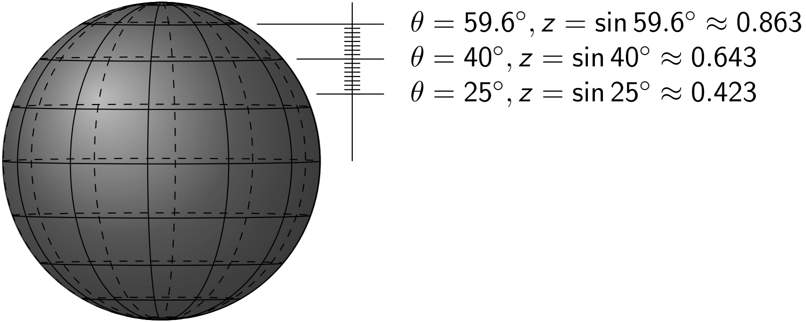
n-Sphereland

Let $B^n = \{x \in \mathbb{R}^n : ||x|| < 1\}$ be the solid n-dimensional unit ball and let

$$S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$$

be the n-dimensional sphere: it is the surface of B^{n+1}.

In 'n-Sphereland' the inhabitants are uniformly distributed on S^n.



Question: let $(X_1, \ldots, X_n, Z)$ be the coordinate of a randomly chosen n-Spherelander. Is Z uniformly distributed?

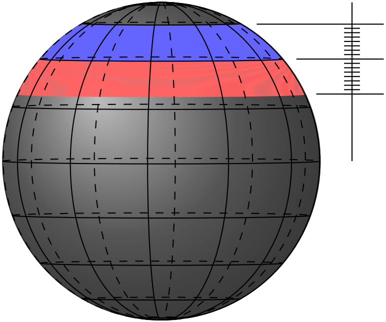
n-Sphereland

Let $B^n = \{x \in \mathbb{R}^n : ||x|| < 1\}$ be the solid n-dimensional unit ball and let

$$S^n = \{x \in \mathbb{R}^{n+1} : ||x|| = 1\}$$

be the n-dimensional sphere: it is the surface of B^{n+1}.

In 'n-Sphereland' the inhabitants are uniformly distributed on S^n.



$\theta = 59.6^\circ, z = \sin 59.6^\circ \approx 0.863$

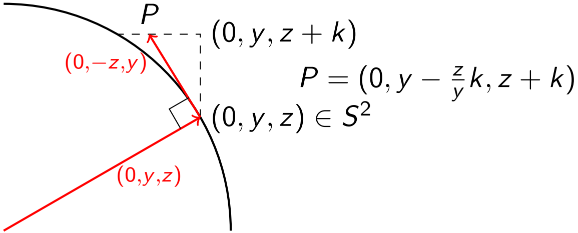
$\theta = 40^\circ, z = \sin 40^\circ \approx 0.643$

$\theta = 25^\circ, z = \sin 25^\circ \approx 0.423$

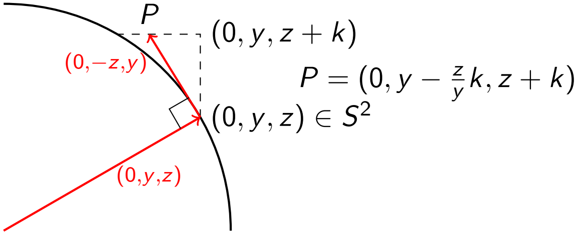
Question: let $(X_1, \ldots, X_n, Z)$ be the coordinate of a randomly chosen n-Spherelander. Is Z uniformly distributed?

Answer: Yes if $n = 2$

Answer: Yes if $n = 2$



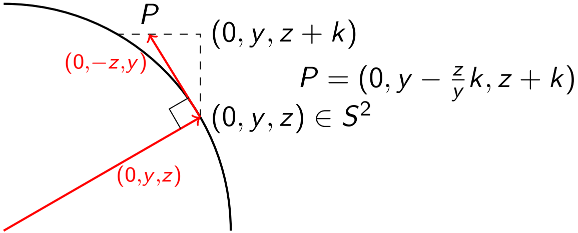
Answer: Yes if $n = 2$



- The length squared of the red line segment tangent to the circle is

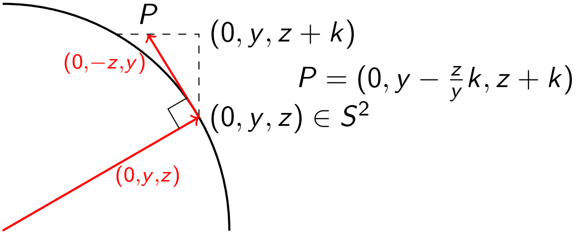
$$\left(\frac{z}{y}k\right)^2 + k^2 = k^2\left(\frac{z^2}{y^2} + 1\right) = k^2\left(\frac{z^2+y^2}{y^2}\right) = \frac{k^2}{1-z^2}$$

Answer: Yes if $n = 2$



- The length squared of the red line segment tangent to the circle is
$$\left(\frac{z}{y}k\right)^2 + k^2 = k^2\left(\frac{z^2}{y^2}+1\right) = k^2\left(\frac{z^2+y^2}{y^2}\right) = \frac{k^2}{1-z^2}$$
- Hence the surface area of the part of the sphere between heights z and $z+k$ is (to first order in k)
$$\frac{k}{\sqrt{1-z^2}} \times \text{circumference of latitude circle at height } z.$$

Answer: Yes if $n = 2$



- The length squared of the red line segment tangent to the circle is

$$\left(\frac{z}{y}k\right)^2 + k^2 = k^2\left(\frac{z^2}{y^2} + 1\right) = k^2\left(\frac{z^2 + y^2}{y^2}\right) = \frac{k^2}{1 - z^2}$$

- Hence the surface area of the part of the sphere between heights z and $z + k$ is (to first order in k)

$$\frac{k}{\sqrt{1 - z^2}} \times \text{circumference of latitude circle at height } z.$$

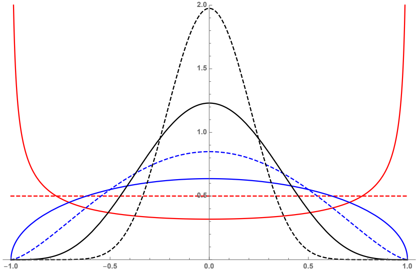
- This is $\frac{k}{\sqrt{1-z^2}}\sqrt{1 - z^2} = k$, independent of z.

Answer: No if $n \neq 2$

To generalize, replace the circumference of the latitude circle at height z with the surface area of S^{n-1} of radius $\sqrt{1-z^2}$.
By dimensional analysis, the probability density function of Z is proportional to $\frac{1}{\sqrt{1-z^2}}(\sqrt{1-z^2})^{n-1} = \sqrt{1-z^2}^{n-2}$.

In[∘]:=
```
Plot[{f[1, z], f[2, z], f[3, z], f[5, z], f[10, z], f[25, z]}, {z, -1, 1},
  PlotRange → {0, 2},
  PlotStyle → {Red, {Red, Dashed}, Blue, {Blue, Dashed}, Black, {Black, Dashed}}]
```

Out[∘]=

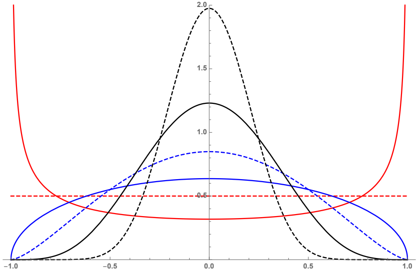


Answer: No if $n \neq 2$

To generalize, replace the circumference of the latitude circle at height z with the surface area of S^{n-1} of radius $\sqrt{1-z^2}$.
By dimensional analysis, the probability density function of Z is proportional to $\frac{1}{\sqrt{1-z^2}}(\sqrt{1-z^2})^{n-1} = \sqrt{1-z^2}^{\,n-2}$.

In[∘]:=
```
Plot[{f[1, z], f[2, z], f[3, z], f[5, z], f[10, z], f[25, z]}, {z, -1, 1},
 PlotRange → {0, 2},
 PlotStyle → {Red, {Red, Dashed}, Blue, {Blue, Dashed}, Black, {Black, Dashed}}]
```

Out[∘]=



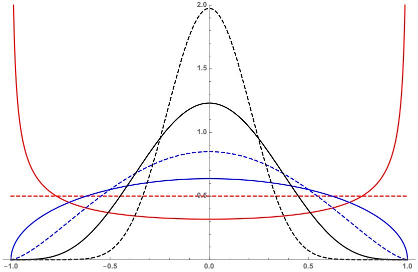
- For large n, by the Law of Large Numbers, $Z \approx \frac{1}{\sqrt{n}}$ with high probability.

Answer: No if $n \neq 2$

To generalize, replace the circumference of the latitude circle at height z with the surface area of S^{n-1} of radius $\sqrt{1-z^2}$.
By dimensional analysis, the probability density function of Z is proportional to $\frac{1}{\sqrt{1-z^2}}(\sqrt{1-z^2})^{n-1} = \sqrt{1-z^2}^{n-2}$.

```
In[∘]:= Plot[{f[1, z], f[2, z], f[3, z], f[5, z], f[10, z], f[25, z]}, {z, -1, 1},
  PlotRange → {0, 2},
  PlotStyle → {Red, {Red, Dashed}, Blue, {Blue, Dashed}, Black, {Black, Dashed}}]
```

Out[∘]=



- For large n, by the Law of Large Numbers, $Z \approx \frac{1}{\sqrt{n}}$ with high probability.
- In fact *all coordinates* are about $\frac{1}{\sqrt{n}}$ with high probability.

Volume of the unit ball

Question 1. What dimension maximizes the volume of the unit ball $B^n = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$?

Question 2. What proportion of the unit cube $[-1, 1]^n$ is occupied by B^n?

Volume of the unit ball

Question 1. What dimension maximizes the volume of the unit ball $B^n = \{x \in \mathbb{R}^n : ||x|| \le 1\}$?

Question 2. What proportion of the unit cube $[-1, 1]^n$ is occupied by B^n?

n	1	2	3	4	5	6	7
V_n	2	π	$\frac{4\pi}{3}$	$\frac{\pi^2}{2}$	$\frac{8\pi^2}{15}$	$\frac{16\pi^3}{15}$	$\frac{\pi^4}{3}$
$V_n \approx$	2	3.142	4.189	4.935	5.264	5.168	4.725

Volume of the unit ball

Question 1. What dimension maximizes the volume of the unit ball $B^n = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$?

Question 2. What proportion of the unit cube $[-1, 1]^n$ is occupied by B^n?

n	1	2	3	4	5	6	7
V_n	2	π	$\frac{4\pi}{3}$	$\frac{\pi^2}{2}$	$\frac{8\pi^2}{15}$	$\frac{16\pi^3}{15}$	$\frac{\pi^4}{3}$
$V_n \approx$	2	3.142	4.189	4.935	5.264	5.168	4.725
$V_n/2^n$	1	$\frac{\pi}{4}$	$\frac{\pi}{6}$	$\frac{\pi^2}{32}$	$\frac{\pi^2}{60}$	$\frac{\pi^3}{384}$	$\frac{\pi^3}{840}$
$V_n/2^n \approx$	1	0.785	0.524	0.308	0.164	0.081	0.037

Volume of the unit ball

Question 1. What dimension maximizes the volume of the unit ball $B^n = \{x \in \mathbb{R}^n : ||x|| \leq 1\}$?

Question 2. What proportion of the unit cube $[-1,1]^n$ is occupied by B^n?

n	1	2	3	4	5	6	7
V_n	2	π	$\frac{4\pi}{3}$	$\frac{\pi^2}{2}$	$\frac{8\pi^2}{15}$	$\frac{16\pi^3}{15}$	$\frac{\pi^4}{3}$
$V_n \approx$	2	3.142	4.189	4.935	5.264	5.168	4.725
$V_n/2^n$	1	$\frac{\pi}{4}$	$\frac{\pi}{6}$	$\frac{\pi^2}{32}$	$\frac{\pi^2}{60}$	$\frac{\pi^3}{384}$	$\frac{\pi^3}{840}$
$V_n/2^n \approx$	1	0.785	0.524	0.308	0.164	0.081	0.037

In particular

$$\frac{V_{2m}}{2^{2m}} = \left(\frac{\pi}{4}\right)^m \frac{1}{m!}$$

which tends to 0 faster than any exponential. So high-dimensional balls are tiny ...

§2 Binary codes: $\mathbb{F}_2^n$ and Hamming balls

Let $C \subseteq \mathbb{F}_2^n$ be a binary code.

In **nearest neighbour decoding**, a received word $v \in \mathbb{F}_2^n$ is decoded as the codeword $u \in C$ nearest to v with respect to Hamming distance:

$$d(u, v) = \left|\{i \in \{1, \ldots, n\} : u_i \neq v_i\}\right|.$$

(If there are several, pick one at random, and fear the worst.)

For instance let $n = 4$ and $C = \{0000, 1110\}$.

- Suppose 0000 is sent and, because of noise in the channel, 0011 is received. Since

  $$d(0000, 0011) = 2 < d(1110, 0011) = 3,$$

  nearest neighbour decoding succeeds,
- If instead 1100 is received, then nearest neighbour decoding fails.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p. So typically pn bits flip.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p. So typically pn bits flip.
- The amount of information in a received bit is $1 - h(p)$, where

$$h(p) = -p \log_2 p - (1-p)\log_2(1-p)$$

  is the entropy (uncertainty) in each flipped bit. E.g. $h(\frac{1}{4}) \approx 0.811$ and $1 - h(\frac{1}{4}) \approx 0.189$. So a $\frac{1}{4}$-noisy bit conveys 0.189 bits of information.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p. So typically pn bits flip.
- The amount of information in a received bit is $1 - h(p)$, where

  $$h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

  is the entropy (uncertainty) in each flipped bit. E.g. $h(\frac{1}{4}) \approx 0.811$ and $1 - h(\frac{1}{4}) \approx 0.189$. So a $\frac{1}{4}$-noisy bit conveys 0.189 bits of information.
- Shannon's Noisy Coding Theorem says that if $\rho < 1 - h(p)$ then in a randomly chosen code of size $2^{\rho n}$, nearest neighbour decoding almost always succeeds.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p. So typically pn bits flip.
- The amount of information in a received bit is $1 - h(p)$, where
  $$h(p) = -p\log_2 p - (1-p)\log_2(1-p)$$
  is the entropy (uncertainty) in each flipped bit. E.g. $h(\frac{1}{4}) \approx 0.811$ and $1 - h(\frac{1}{4}) \approx 0.189$. So a $\frac{1}{4}$-noisy bit conveys 0.189 bits of information.
- Shannon's Noisy Coding Theorem says that if $\rho < 1 - h(p)$ then in a randomly chosen code of size $2^{\rho n}$, nearest neighbour decoding almost always succeeds.
- Thus we can send up to $1 - h(p)$ bits of (reliable) information for each bit sent through the channel. For instance,
  - The maximum 4G data rate is 100 million bits per second.

Shannon's probabilistic model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, each bit is flipped independently with probability p. So typically pn bits flip.
- The amount of information in a received bit is $1 - h(p)$, where

  $$h(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

  is the entropy (uncertainty) in each flipped bit. E.g. $h(\frac{1}{4}) \approx 0.811$ and $1 - h(\frac{1}{4}) \approx 0.189$. So a $\frac{1}{4}$-noisy bit conveys 0.189 bits of information.
- Shannon's Noisy Coding Theorem says that if $\rho < 1 - h(p)$ then in a randomly chosen code of size $2^{\rho n}$, nearest neighbour decoding almost always succeeds.
- Thus we can send up to $1 - h(p)$ bits of (reliable) information for each bit sent through the channel. For instance,
  - The maximum 4G data rate is 100 million bits per second.
  - $\big[$I should know, I have tried all four networks.$\big]$
  - If $p = \frac{1}{4}$ then since $1 - h(\frac{1}{4}) \approx 0.189$, we can reliably send 18.8 million bits per second.

Hamming's (simplified) adversarial model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

- When $u \in C$ is sent, exactly pn bits flip, chosen adversarially.
- Nearest neighbour decoding always succeeds if and only if the Hamming balls of radius pn about codewords are disjoint.

Hamming's (simplified) adversarial model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

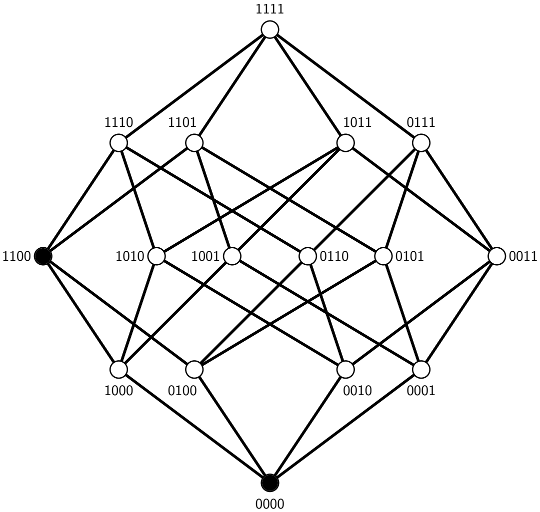
- When $u \in C$ is sent, exactly pn bits flip, chosen adversarially.
- Nearest neighbour decoding always succeeds if and only if the Hamming balls of radius pn about codewords are disjoint.
- The Plotkin bound implies that if $p \geq \frac{1}{4}$ and the Hamming balls of radius $\frac{n}{2}$ are disjoint then $|C| \leq 4n$. Hence $(\log_2 |C|)/n \to 0$ as $n \to \infty$.

Hamming's (simplified) adversarial model

Let $C \subseteq \mathbb{F}_2^n$ be a binary code. Let $p < \frac{1}{2}$.

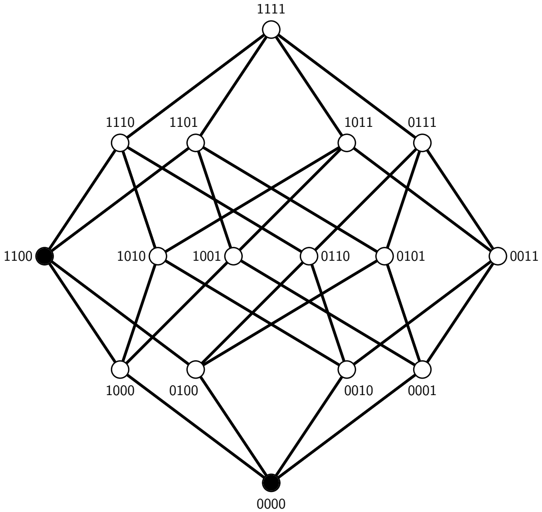
- When $u \in C$ is sent, exactly pn bits flip, chosen adversarially.
- Nearest neighbour decoding always succeeds if and only if the Hamming balls of radius pn about codewords are disjoint.
- The Plotkin bound implies that if $p \geq \frac{1}{4}$ and the Hamming balls of radius $\frac{n}{2}$ are disjoint then $|C| \leq 4n$. Hence $(\log_2 |C|)/n \to 0$ as $n \to \infty$.
- For instance, if $p = \frac{1}{4}$, only 28.6 bits can be sent per second on the 4G network.

Difference between probabilistic and adversarial errors



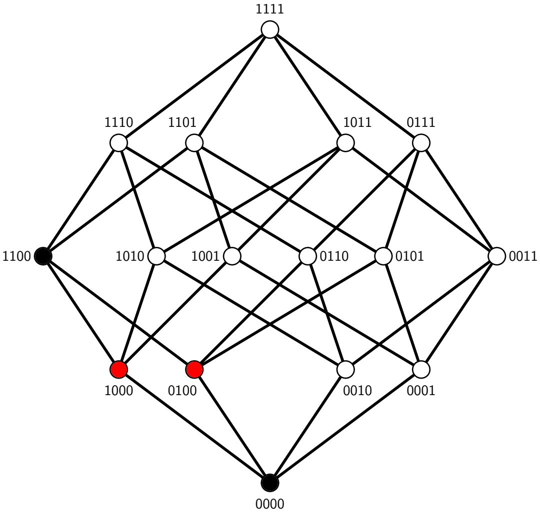
Question: why the huge difference between the two models?

Difference between probabilistic and adversarial errors



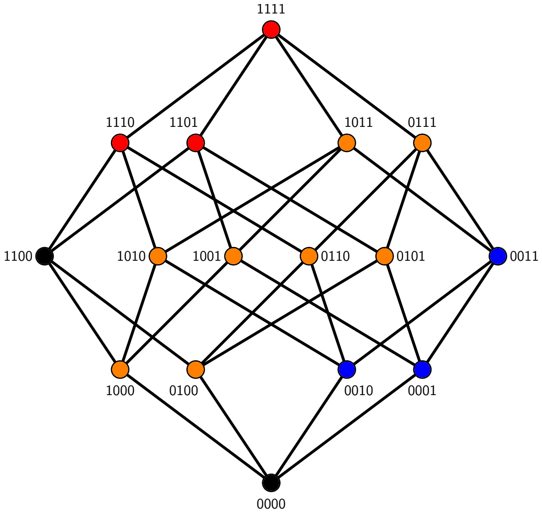
My answer: because the traditional picture (which I keep on drawing in my coding theory courses) is completely misleading.

Difference between probabilistic and adversarial errors



One adversarial error: The sent codeword 0000 heads for 1100 like a homing missile, and we assume nearest neighbour decoding makes the wrong choice.

Difference between probabilistic and adversarial errors



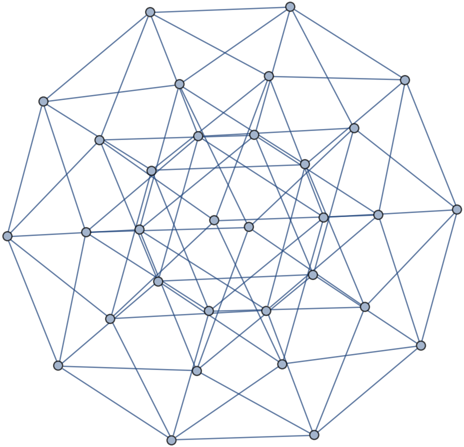
Probabilistic errors: Even if up to 2 errors occur (see middle of diagram and below) still more likely than not to decode correctly.

The effect is greater for larger n

Why: because $\mathbb{F}_2^n$ is really, really highly connected. In this sense $\mathbb{F}_2^{256}$ is 'larger' than $\mathbb{R}^4$.

In[∘]:= `HypercubeGraph[5]`

Out[∘]=

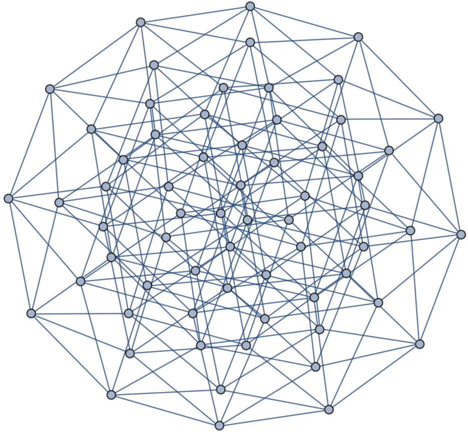


The effect is greater for larger n

Why: because $\mathbb{F}_2^n$ is really, really highly connected. In this sense $\mathbb{F}_2^{256}$ is ‘larger’ than $\mathbb{R}^4$.

In[∘]:= `HypercubeGraph[6]`

Out[∘]=

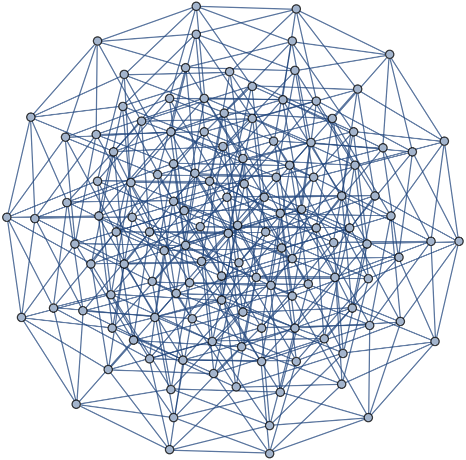


The effect is greater for larger n

Why: because $\mathbb{F}_2^n$ is really, really highly connected. In this sense $\mathbb{F}_2^{256}$ is 'larger' than $\mathbb{R}^4$.

In[∘]:= `HypercubeGraph[7]`

Out[∘]=

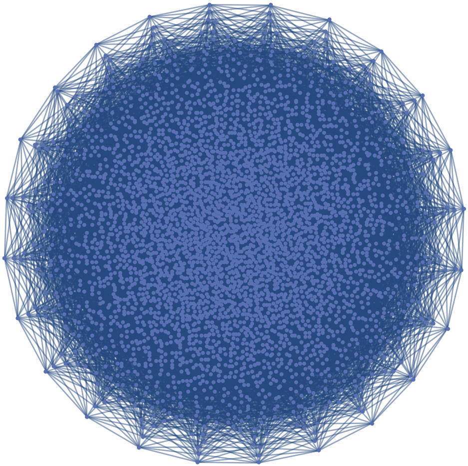


The effect is greater for larger n

Why: because $\mathbb{F}_2^n$ is really, really highly connected. In this sense $\mathbb{F}_2^{256}$ is 'larger' than $\mathbb{R}^4$.

In[∘]:= `HypercubeGraph[12]`

Out[∘]=



Any questions?

Any questions?

My blog post, see `wildonblog.wordpress.com`, has outline proofs of the special cases of Shannon's Noisy Coding Theorem and the Plotkin bound. Also the connection with cryptography:

- why $\mathbb{F}_2^{56}$ is tiny and $\mathbb{F}_2^{256}$ might as well be $\mathbb{F}_2^{\infty}$,

and computation:

- the amazing sense in which $2^{2^{\mathbb{N}}}$ (meaning definable subsets of the Cantor set) is a smaller computational space than $2^{\mathbb{N}}$.