Measure and Concentration

Bhaumik and Jian

Introduction

In 1829, Dirichlet proved the following theorem:

Theorem (Dirichlet, 1829)

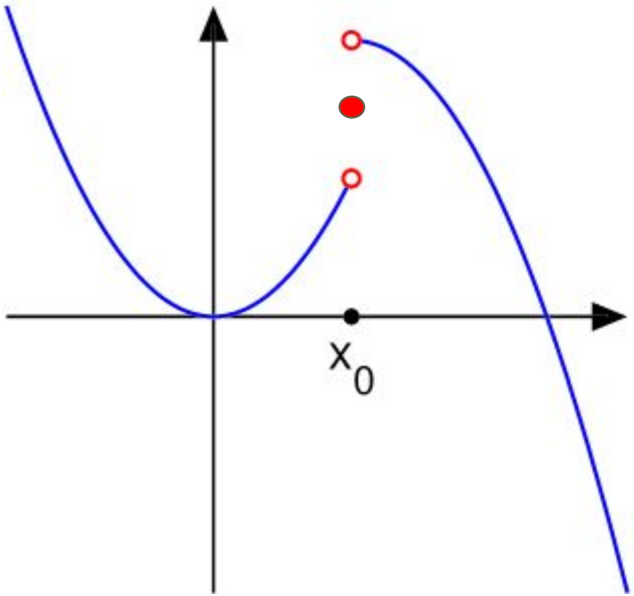
The Fourier series of a piecewise-smooth f on $[-\pi, \pi]$ converges to

$$\frac{f(x_+) + f(x_-)}{2}.$$

That is, the Fourier series of f converges at every point of continuity. At discontinuities, it takes the middle value.

Introduction

Theorem (Dirichlet, 1829): The Fourier series of a piecewise-smooth f converges at every point of continuity. At discontinuities, it takes the middle value.



Introduction

Theorem (Dirichlet, 1829): The Fourier series of a piecewise-smooth f converges at every point of continuity. At discontinuities, it takes the middle value.

Why is this interesting?

Introduction

Theorem (Dirichlet, 1829): The Fourier series of a piecewise-smooth f converges at every point of continuity. At discontinuities, it takes the middle value.

Why is this interesting?

- First ever theorem about Fourier series

Introduction

Theorem (Dirichlet, 1829): The Fourier series of a piecewise-smooth f converges at every point of continuity. At discontinuities, it takes the middle value.

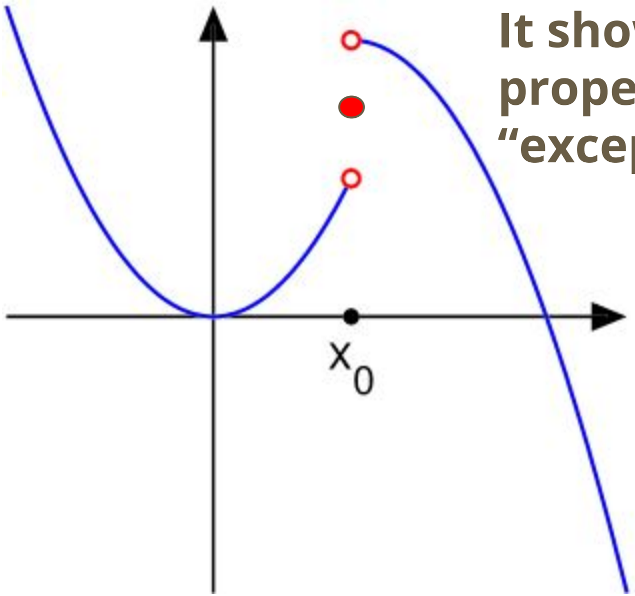
Why is this interesting?

- First ever theorem about Fourier series
- But more importantly…

Introduction

Theorem (Dirichlet, 1829): The Fourier series of a piecewise-smooth f converges at every point of continuity. At discontinuities, it takes the middle value.

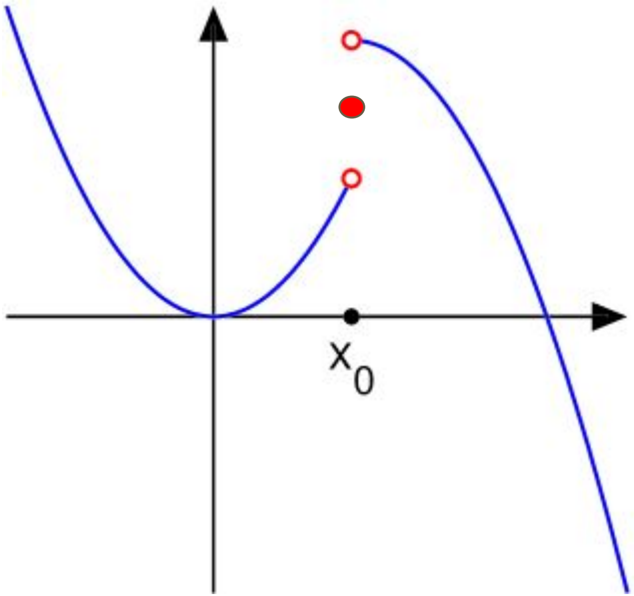
But more importantly…



It showed that some desired properties do not hold at "exceptional yet negligible sets."

Introduction

This made it clear to analysts that to rigorously establish convergence results in general, we need to be able to rule out “exceptional yet negligible sets.”



Why measure theory in probability?

This made it clear to analysts that to rigorously establish convergence results, we need to be able to rule out "exceptional yet negligible sets."

Now consider the Law of Large Numbers:

Why measure theory in probability?

This made it clear to analysts that to rigorously establish convergence results, we need to be able to rule out “exceptional yet negligible sets.”

Now consider the Law of Large Numbers:

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}X_1$$

Why measure theory in probability?

This made it clear to analysts that to rigorously establish convergence results, we need to be able to rule out "exceptional yet negligible sets."

We simplify to the case of coin flips:

$$\frac{number\ of\ heads}{number\ of\ flips} \longrightarrow \frac{1}{2}$$

Why measure theory in probability?

$$\frac{number\ of\ heads}{number\ of\ flips} \longrightarrow \frac{1}{2}$$

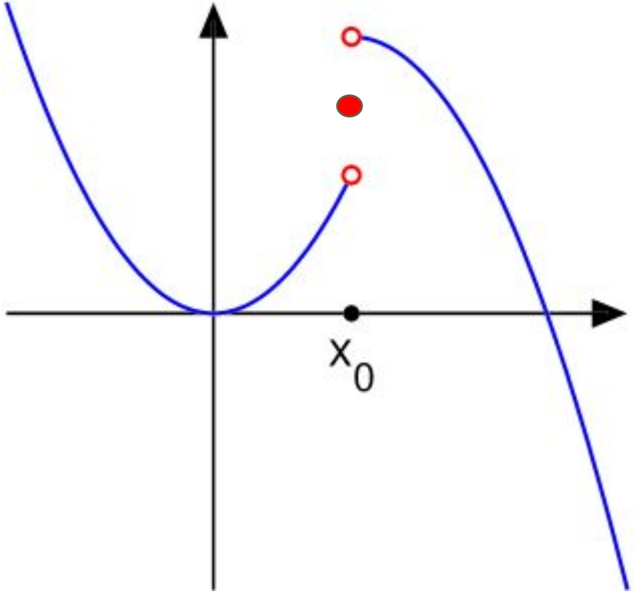
Why measure theory in probability?

$$\cancel{\frac{number\ of\ heads}{number\ of\ flips} \longrightarrow \frac{1}{2}}$$

What happens when we only flip heads????

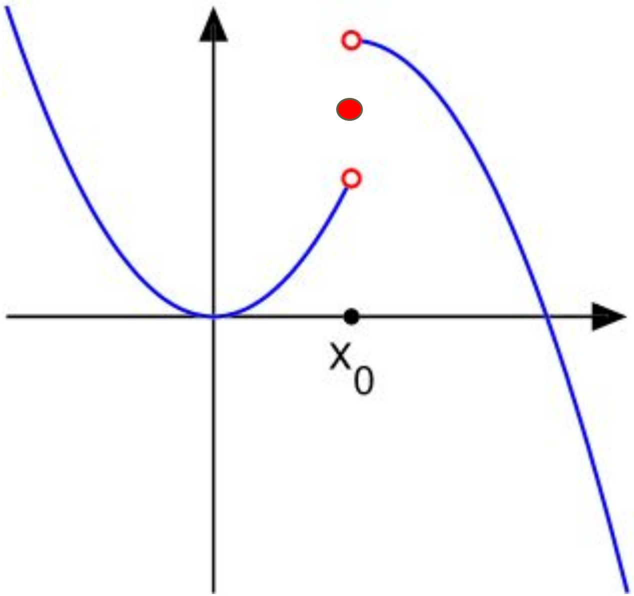
Why measure theory in probability?

Recall the idea of “exceptional yet negligible sets.”



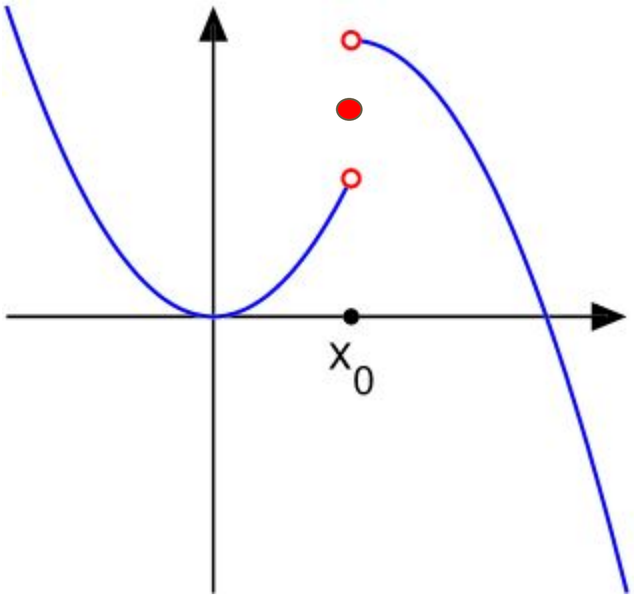
Why measure theory in probability?

The set of points where the Fourier series does not converge, or the set of events that violates the LLN, is an "exceptional yet negligible set."



Why measure theory in probability?

The event we only flip heads, or any event that violates the LLN,
has "measure zero," or "probability zero."



Why measure theory in probability?

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}X_1$$

If a property holds except on a set of measure zero, we say the property holds **"almost everywhere,"** or **"almost surely."**

Why measure theory in probability?

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}X_1$$

If a property holds except on a set of measure zero, we say the property holds **"almost everywhere,"** or **"almost surely."**

So the LLN holds with probability 1, or almost surely.

And we are allowed to say this thanks to measure theory.

Why measure theory in probability?

In 1933, Kolmogorov saw that measure and integration theory, originally developed to solve technical issues in Fourier analysis, could be used to make probability rigorous.

Why measure theory in probability?

In 1933, Kolmogorov saw that measure and integration theory, originally developed to solve technical issues in Fourier analysis, could be used to make probability rigorous.

Besides the ruling out of problematic sets that we have explored, measure and integration are fundamental to the description of basic notions of probability such as expectation, random variables, convergence results, etc.

A Weird Problem

Problem 1

Suppose

$$X_n = \begin{cases} n^2 - 1, & \text{with probability } n^{-2}, \\ -1, & \text{with probability } 1 - n^{-2}. \end{cases}$$

Show that

$$\mathbb{E}[X_n] = 0 \quad \text{but} \quad n^{-1} S_n \to -1 \ \text{ a.s.}$$

Result from Measure Theory

Lemma 1

If

$$Z_k \geq 0 \quad \text{and} \quad \sum \mathbb{E}[Z_k] < \infty$$

Then

$$\sum Z_k < \infty \quad \text{(a.s)} \quad \text{so} \quad Z_k \to 0 \quad \text{(a.s)}$$

Borel Cantelli Lemma 1

Borel Cantelli Lemma 1

If Events E_n have

$$\sum \mathbb{P}[E_n] < \infty = 0$$

Then

$$\mathbb{P}[E_n \ \text{i.o}]$$

In other words,

$$\mathbb{P}[\{\omega : \omega \in E_n \ \text{for infinitely many n}\}] = 0$$

Back to Problem!

Problem 1

Suppose

$$X_n = \begin{cases} n^2 - 1, & \text{with probability } n^{-2}, \\ -1, & \text{with probability } 1 - n^{-2}. \end{cases}$$

Show that

$$\mathbb{E}[X_n] = 0 \quad \text{but} \quad n^{-1} S_n \to -1 \ \text{ a.s.}$$

Why does Strong Law of Large Numbers fail?

- What are the conditions for SLLN to hold?

Talk to the person next to you!

Why do we care about the Law of Large Numbers?

What is the Law of Large Numbers saying?

What is the Law of Large Numbers saying?

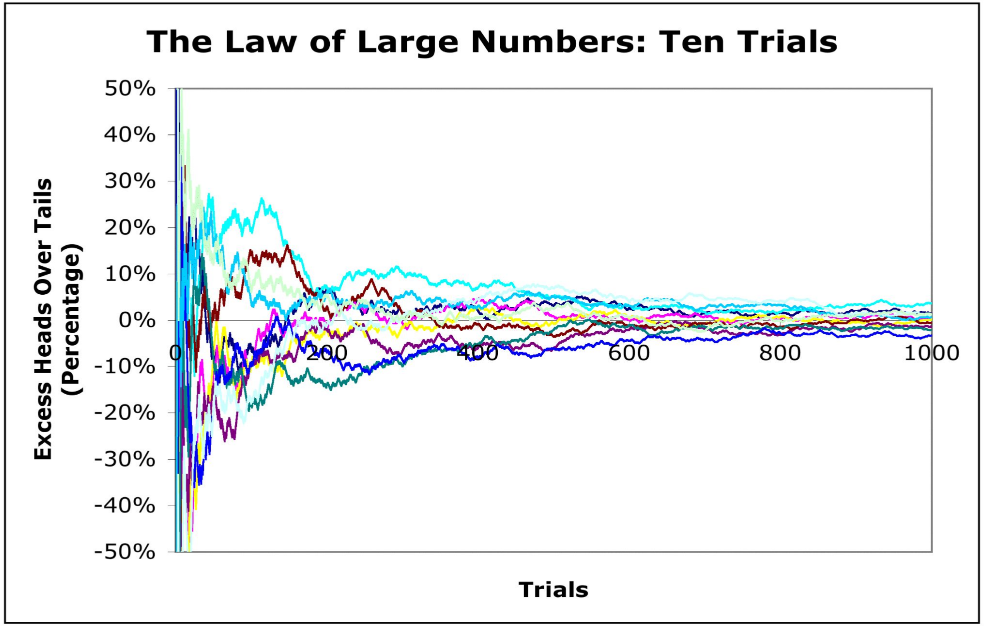
$$\frac{X_1 + ... + X_n}{n} \longrightarrow \mathbb{E}X_1$$

What is the Law of Large Numbers saying?

$$\frac{X_1 + \ldots + X_n}{n} \longrightarrow \mathbb{E}X_1$$

The aggregate of independent random variables behave predictably.

What is the Law of Large Numbers saying?



The aggregate of independent random variables behave predictably.

Why do we care about the Law of Large Numbers?

Intuitively, adding randomness to an already random system should increase randomness—

Why do we care about the Law of Large Numbers?

Intuitively, adding randomness to an already random system should increase randomness—

But the opposite happens: randomness concentrates around typical behavior, making outcomes almost completely predictable.

Probabilists call this “Concentration”

Probabilists call this “Concentration”

“A random variable … that depends on the influence of many independent variables is essentially constant.”

-Michel Talagrand

Concentration Principle

If $X_1, \ldots, X_n$ are independent (or weakly dependent) random variables, then the random variable $f(X_1, \ldots, X_n)$ is "close" to its mean $\mathbb{E}[f(X_1, \ldots, X_n)]$ provided that the function $f(x_1, \ldots, x_n)$ is not too "sensitive" to any of the coordinates x_i.

Source: Direct quote from the lecture notes (Van Handel, 2016)

The Standard Concentration Statement

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

$$X \approx \mathbb{E}X$$

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

$$X \approx \mathbb{E}X$$

$$X - \mathbb{E}X \approx 0$$

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

$$X \approx \mathbb{E}X$$

$$X - \mathbb{E}X \approx 0$$

$$\mathbb{P}\Big(X - \mathbb{E}X \approx 0\Big) = big$$

The Standard Concentration Statement

$$\mathbb{P}\Big(X - \mathbb{E}X \approx 0\Big) = big$$

The Standard Concentration Statement

$$\mathbb{P}\Big(X - \mathbb{E}X \approx 0\Big) = big$$

$$\mathbb{P}\Big(|X - \mathbb{E}X| \geq t\Big) = small$$

The Standard Concentration Statement

$$\mathbb{P}\Big(|X - \mathbb{E}X| \geq t\Big) = small$$

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

$$\Updownarrow$$

$$\mathbb{P}\Big(|X - \mathbb{E}X| \geq t\Big) = small$$

The Standard Concentration Statement

$$Observed\ Value \approx Expected\ Value$$

$$\Updownarrow$$

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) = small$$

The Standard Concentration Statement

Concentration Principle

If $X_1, \ldots, X_n$ are independent (or weakly dependent) random variables, then the random variable $f(X_1, \ldots, X_n)$ is "close" to its mean $\mathbb{E}[f(X_1, \ldots, X_n)]$ provided that the function $f(x_1, \ldots, x_n)$ is not too "sensitive" to any of the coordinates x_i.

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) = small$$

Concentration Phenomenon, intuitively.

This begs the question: What do we mean by “small”? How “small”?

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) = small$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) = small$$

For simplicity, consider the case when $f(x_1, \ldots, x_n) = x_1 + \ldots + x_n$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) = small$$

Concentration Phenomenon, intuitively.

Chebyshev's Inequality:

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq \frac{Var(S_n)}{t^2}$$

Concentration Phenomenon, intuitively.

Chebyshev's Inequality:

$$\mathbb{P}\Big(\big|S_n - \mathbb{E}S_n\big| \geq t\Big) \leq \frac{Var(S_n)}{t^2}$$

So we have that the deviation decreases at least quadratically by Chebyshev's.

Concentration Phenomenon, intuitively.

Chebyshev's Inequality:

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq \frac{Var(S_n)}{t^2}$$

So we have that the deviation decreases at least quadratically by Chebyshev's.

But can we do smaller?

Concentration Phenomenon, intuitively.

Recall the Central Limit Theorem:

Concentration Phenomenon, intuitively.

Recall the Central Limit Theorem:

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mathbb{E}S_n) \sim N(0, 1)$$

Concentration Phenomenon, intuitively.

Recall the Central Limit Theorem:

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mathbb{E}S_n) \sim N(0,1)$$

So with the change of variables $t \mapsto (\sigma\sqrt{n})t$,

Concentration Phenomenon, intuitively.

Recall the Central Limit Theorem:

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mathbb{E}S_n) \sim N(0,1)$$

So with the change of variables $t \mapsto (\sigma\sqrt{n})t$,

$$\mathbb{P}\Big(\big|S_n - \mathbb{E}S_n\big| \geq t\Big) \approx 2 \cdot \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}}\, dx$$

Concentration Phenomenon, intuitively.

Recall the Central Limit Theorem:

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mathbb{E}S_n) \sim N(0,1)$$

So with the change of variables $t \mapsto (\sigma\sqrt{n})t$,

$$\mathbb{P}\Big(\big|S_n - \mathbb{E}S_n\big| \geq t\Big) \approx 2 \cdot \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}}\, dx \leq e^{-\frac{t^2}{2}}$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq e^{-\frac{t^2}{2}}$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\left(|S_n - \mathbb{E}S_n| \geq t\right) \leq e^{-\frac{t^2}{2}}$$

We have exponential decay! Super small, fast, and good!!

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq e^{-\frac{t^2}{2}}$$

We have exponential decay! Super small, fast, and good!!

$$Observed\ Value \approx Expected\ Value$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(|S_n - \mathbb{E}S_n| \geq t\Big) \leq e^{-\frac{t^2}{2}}$$

Side note:

$$|F_n(x) - \Phi(x)| \leq \frac{C\rho}{\sigma^3\sqrt{n}}$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) = small$$

$$Observed\ Value \approx Expected\ Value$$

Concentration Phenomenon, intuitively.

$$\mathbb{P}\left(\left|f(X) - \mathbb{E}f(X)\right| \geq t\right) \lesssim e^{-\frac{t^2}{2}}$$

$$Observed\ Value \approx Expected\ Value$$

"A random variable … that depends on the influence of many independent variables is essentially constant."

-Michel Talagrand

"A random variable ... that depends on the influence of many independent variables satisfies Chernoff-type bounds."

-Michel Talagrand

"A random variable … that depends on the influence of many independent variables satisfies Chernoff-type bounds."

-Michel Talagrand

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) \lesssim e^{-\frac{t^2}{2}}$$

Important notes

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) \lesssim e^{-\frac{t^2}{2}}$$

Of course we can't expect concentration to hold for completely arbitrary f and X.

Important notes

$$\mathbb{P}\left(\left|f(X) - \mathbb{E}f(X)\right| \geq t\right) \lesssim e^{-\frac{t^2}{2}}$$

Of course we can't expect concentration to hold for completely arbitrary f and X.

- Function must be Lipschitz (Not too "sensitive" to any of our coordinates)

Important notes

$$\mathbb{P}\Big(\big|f(X) - \mathbb{E}f(X)\big| \geq t\Big) \lesssim e^{-\frac{t^2}{2}}$$

Of course we can’t expect concentration to hold for completely arbitrary f and X.

- Function must be Lipschitz (Not too “sensitive” to any of our coordinates)
- The coordinates must be sub-Gaussian, sub-exponential

Important notes

$$\mathbb{P}\left(\left|f(X) - \mathbb{E}f(X)\right| \geq t\right) \lesssim e^{-\frac{t^2}{2}}$$

Of course we can't expect concentration to hold for completely arbitrary f and X.

- Function must be Lipschitz (Not too "sensitive" to any of our coordinates)
- The coordinates must be sub-Gaussian, sub-exponential
- etc.

A Concentration Result

Again, why should we care?

A Concentration Result

We illustrate with an example:

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E} f(X)$?

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E} f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E} f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

To simplify our analysis, assume the coordinates are i.i.d. $X_i \sim N(0, 1)$.

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E}f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

To simplify our analysis, assume the coordinates are i.i.d. $X_i \sim N(0, 1)$

Then,

$$\mathbb{E}\|X\|_2^2$$

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E} f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

To simplify our analysis, assume the coordinates are i.i.d. $X_i \sim N(0, 1)$

Then,

$$\mathbb{E}\|X\|_2^2 = \mathbb{E}\sum_{i=1}^{n} X_i^2$$

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E} f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

To simplify our analysis, assume the coordinates are i.i.d. $X_i \sim N(0, 1)$

Then,

$$\mathbb{E}\|X\|_2^2 = \mathbb{E}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathbb{E}X_i^2$$

A Concentration Result

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ that maps $x \mapsto \|x\|_2$.

What is $\mathbb{E}f(X)$? That is, what is the expected length, or expected Euclidean norm, of random vector $X = (X_1, X_2, \ldots, X_n)$?

To simplify our analysis, assume the coordinates are i.i.d. $X_i \sim N(0,1)$

Then,

$$\mathbb{E}\|X\|_2^2 = \mathbb{E}\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} \mathbb{E}X_i^2 = n$$

A Concentration Result

$$\mathbb{E}\,\|X\|_2^2 = n$$

A Concentration Result

$$\mathbb{E}\,\|X\|_2^2 = n$$

$$\mathbb{E}\,\|X\|_2 = \sqrt{n}$$

A Concentration Result

$$\mathbb{E}\left\|X\right\|_2 = \sqrt{n}$$

A Concentration Result

$$\mathbb{E}\,\|X\|_2 = \sqrt{n}$$

So the length of random vector X should be

$$\|X\|_2 \approx \sqrt{n}$$

(Recall that $Observed\ Value \approx Expected\ Value$)

A Concentration Result

Formally, we can state $\|X\|_2 \approx \sqrt{n}$ in terms of our standard concentration statement:

A Concentration Result

Formally, we can state $\|X\|_2 \approx \sqrt{n}$ in terms of our standard concentration statement:

$$\mathbb{P}\Big(\big| \, \|X\|_2 - \sqrt{n} \big| \geq t \Big) = small$$

A Concentration Result

With a simple application of Bernstein's inequality we can get the following bound:

$$\mathbb{P}\Big(\big|\, \|X\|_2 - \sqrt{n}\big| \geq t\Big) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

A Concentration Result

$$\mathbb{P}\left(\big| \, \|X\|_2 - \sqrt{n} \big| \geq t \right) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

A Concentration Result

$$\mathbb{P}\Big(\big| \|X\|_2 - \sqrt{n} \big| \geq t \Big) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

We want to note a few things about the following inequality:

A Concentration Result

$$\mathbb{P}\left(\big| \, \|X\|_2 - \sqrt{n} \big| \geq t \right) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

We want to note a few things about the following inequality:

- The length of random vector X is “essentially constant,”

A Concentration Result

$$\mathbb{P}\Big(\big| \|X\|_2 - \sqrt{n} \big| \geq t \Big) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

We want to note a few things about the following inequality:

- The length of random vector X is “essentially constant,” that is to say, it is essentially equidistant from the origin, that is, it is close to a sphere of radius $\sqrt{n}$

A Concentration Result

$$\mathbb{P}\left(\left| \|X\|_2 - \sqrt{n} \right| \geq t \right) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

We want to note a few things about the following inequality:

- The length of random vector X is “essentially constant,” that is to say, it is essentially equidistant from the origin, that is, it is close to a sphere of radius $\sqrt{n}$
- The bound on our deviation does NOT depend on n.

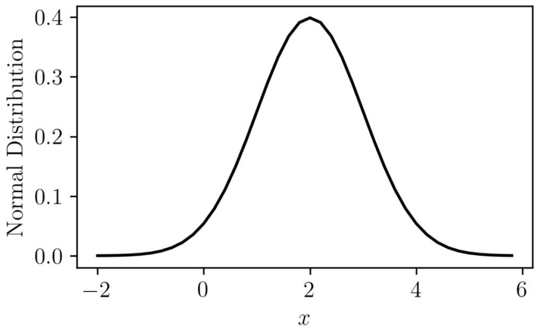
A Concentration Result

$$\mathbb{P}\left(\left| \, \|X\|_2 - \sqrt{n} \right| \geq t \right) \leq 2 \cdot exp(\frac{-ct^2}{2})$$

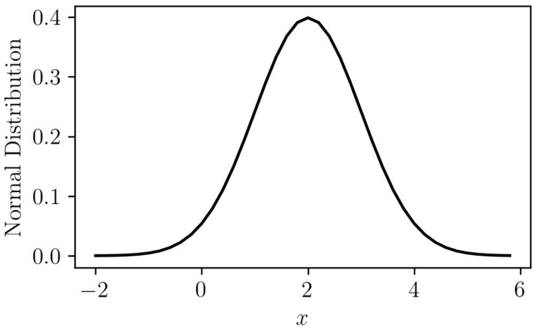
We want to note a few things about the following inequality:

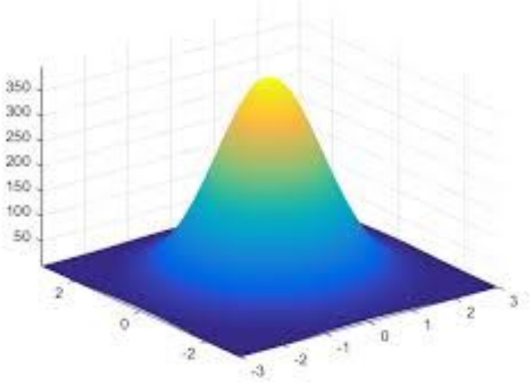
- The length of random vector X is “essentially constant,” that is to say, it is essentially equidistant from the origin, that is, it is close to a sphere of radius $\sqrt{n}$
- The bound on our deviation does NOT depend on n. As n grows, almost all of our observations stay within a constant distance from $\sqrt{n}S^{n-1}$

A Concentration Result

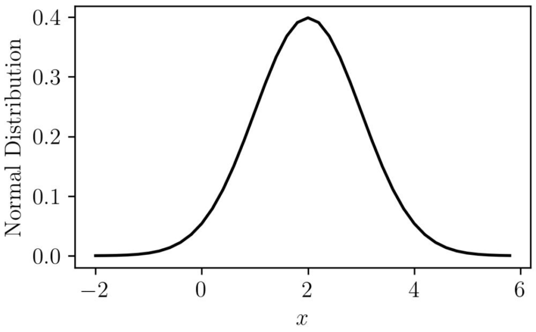


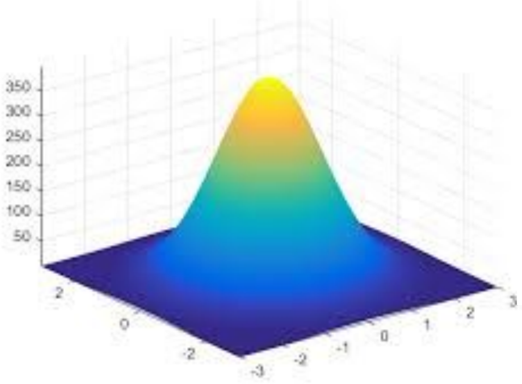
A Concentration Result



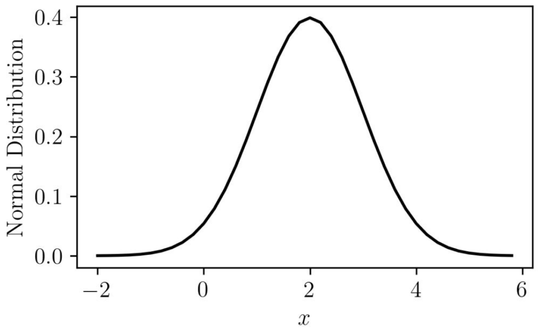


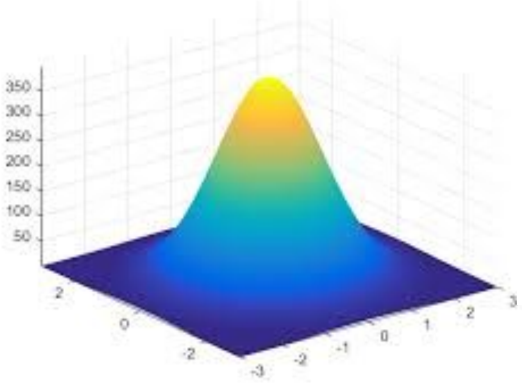
A Concentration Result

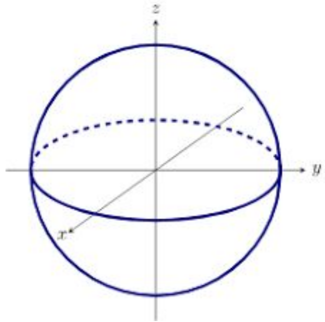




A Concentration Result

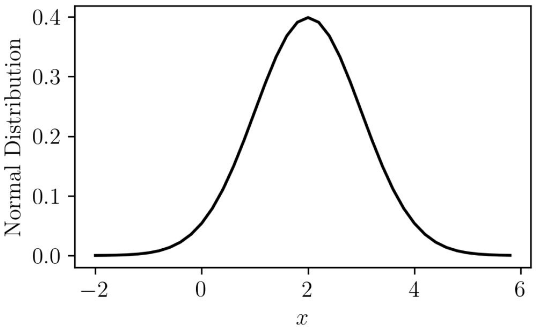


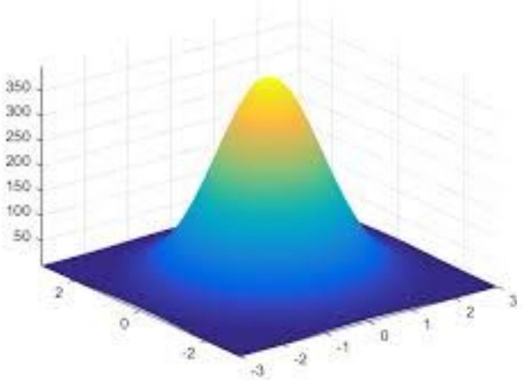


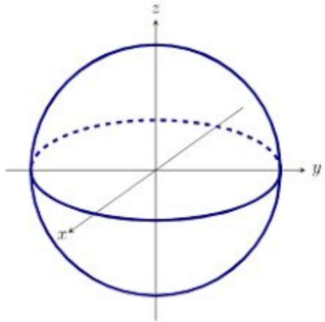


A Concentration Result

???







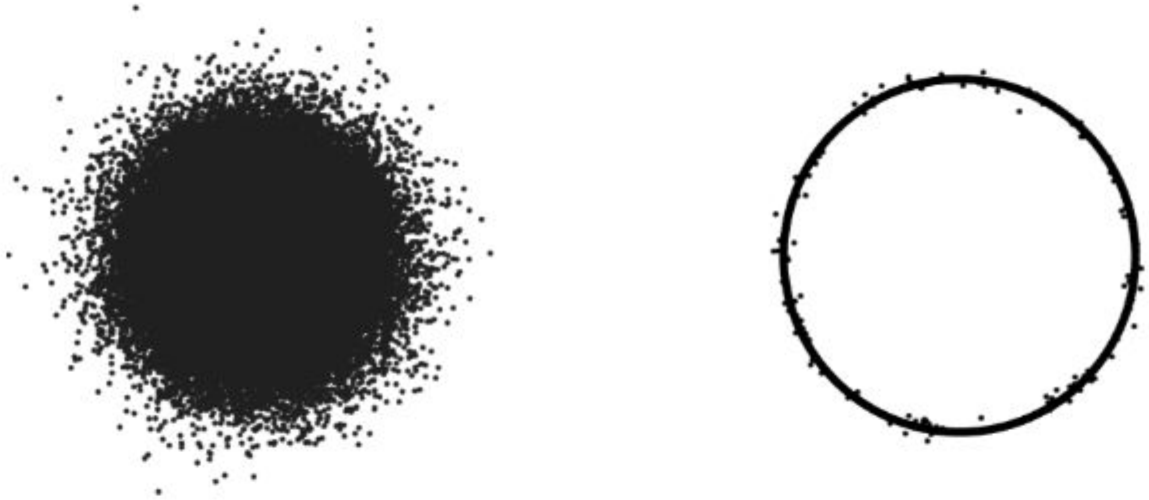


Figure 3.6 A Gaussian point cloud in two dimensions (left) and its intuitive visualization in high dimensions (right). In high dimensions, the standard normal distribution is very close to the uniform distribution on the sphere of radius $\sqrt{n}$.

Source: Vershynin, 2018

Applications to Statistics

Applications to Statistics

$$Observed\ Value \approx Expected\ Value$$

Applications to Statistics

$$Observed\ Value \approx Expected\ Value$$

$$Sample \approx Population$$ (With very high probability)

Applications to Statistics

- Covariance matrix estimation
- High-dimensional regression (e.g. LASSO)
- Empirical risk minimization
- etc.

Fun Fact


KARINE D.
COIFFURE • HOMME • FEMME • ENFANT
Allianz
Assurances, Santé & Services financiers


CENTRE HOSPITALIER EMILE BOREL
URGENCES
Accueil / Admissions
E.H.P.A.D. La Sorgues

Acknowledgements

First and foremost, we want to thank Leon and Nila for being excellent mentors and for their time and effort in teaching us probability. We also want to thank Ethan for organizing the Directed Reading Program.

References

Dirichlet, P. G. L. (1829). Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données. *Journal für die reine und angewandte Mathematik*, 4, 157–169.

Hawkins, T. (1979). *Lebesgue's theory of integration: Its origins and development.* University of Wisconsin Press.

Kolmogorov, A. N. (1933). Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer.

Talagrand, M. (1996). A new look at independence. *The Annals of Probability*, 24(1), 1–34. https://doi.org/10.1214/aop/1039639355

Van Handel, R. (2016). *High-dimensional probability: Lecture notes.* Princeton University. Retrieved from https://web.math.princeton.edu/~rvan/APC550.pdf

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science.* Cambridge University Press.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint.* Cambridge University Press.

Williams, D. (1991). *Probability with martingales.* Cambridge University Press.

Questions...?