Introduction to Probability

Lecture 8: Basic Inequalities and Law of Large Numbers

Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk

Easter 2024

UNIVERSITY OF CAMBRIDGE

Outline

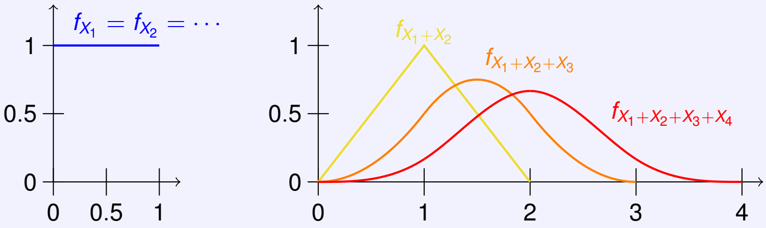
Intro: Sum of Independent (Uniform) Random Variables

Example 1

Let X_1 and X_2 be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X_1 + X_2$ look like? What happens for $X_1 + X_2 + X_3$ etc.?

Answer

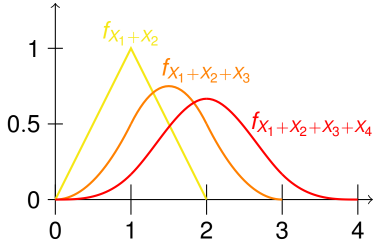
Let us try to sketch the densities without explicit computations[a]

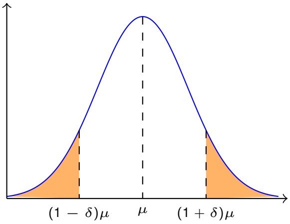


[a] This is also called "convolution". The detailed calculation for $f_{X_1+X_2}$ can be found at the end of these slides. The exact distribution is known for any number of random variables under the name Irwin-Hall distribution.

Motivation

We will study sums of independent and identically distributed variables. How does their distribution look like, and how well do they concentrate around the expectation?





1. Markov's inequality
2. Chebyshev's inequality
3. Law of Large Numbers
4. Central Limit Theorem

Re-use concepts from previous lectures:

1. Independence (Random Var.) (Lec. 1, 7)
2. Expectation and Variance (Lec. 2, 3)
3. Normal Distribution (Lec. 5)
4. Sums of Random Variables (Lec. 6)

Outline

Introduction

Markov's Inequality and Chebyshev's Inequality

Weak Law of Large Numbers

Markov's Inequality

Markov's Inequality

For any non-negative random variable X with finite $\mathbf{E}[X]$, it holds for any $a > 0$,

$$\mathbf{P}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

Markov's inequality is a so-called tail-bound: it upper bounds the probability that the random variable exceeds its mean

A. Markov (1856-1922)

Comments:

- Markov's inequality can be rewritten as: for any $\delta > 0$,

$$\mathbf{P}[X \geq \delta \cdot \mathbf{E}[X]] \leq 1/\delta.$$

- Advantage: Very basic inequality, we only need to know $\mathbf{E}[X]$
- Downside: For many distributions, the tail bound might be quite loose
- Proof is similar to the proof of Chebyshev's inequality (Exercise!)

Applying Markov's Inequality

Example 2

Consider throwing an unbiased, six-sided dice 120 times and let X denote the number of times we obtain a six.

1. Derive an upper bound on $\mathbf{P}[\,X \geq 30\,]$.
2. Can you also derive an upper bound on $\mathbf{P}[\,X \leq 10\,]$?

Answer

Chebyshev's Inequality

Chebyshev's Inequality

For any random variable X with finite $\mathbf{E}[X]$ and $\mathbf{V}[X]$, for any $a > 0$,

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{V}[X]/a^2.$$

P. Chebyshev (1821-1894)

Chebyshev's Inequality

Chebyshev's Inequality

For any random variable X with finite $\mathbf{E}[X]$ and $\mathbf{V}[X]$, for any $a > 0$,

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{V}[X]/a^2.$$

P. Chebyshev (1821-1894)

Comments:

- can be rewritten as:

The "$\mu \pm$ a few σ" rule. Most of the probability mass is within a few standard deviations from μ.

$$\mathbf{P}\left[|X - \mathbf{E}[X]| \geq \sqrt{\delta \cdot \mathbf{V}[X]}\right] \leq 1/\delta.$$

Chebyshev's Inequality

Chebyshev's Inequality

For any random variable X with finite $\mathbf{E}[X]$ and $\mathbf{V}[X]$, for any $a > 0$,

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{V}[X]/a^2.$$

P. Chebyshev (1821-1894)

Comments:

- can be rewritten as:

The "$\mu \pm$ a few σ" rule. Most of the probability mass is within a few standard deviations from μ.

$$\mathbf{P}\left[|X - \mathbf{E}[X]| \geq \sqrt{\delta \cdot \mathbf{V}[X]}\right] \leq 1/\delta.$$

- Unlike Markov, Chebyshev's inequality is two-sided and also holds for random variables with negative values
- In most cases, Chebyshev's inequality yields much stronger bounds than Markov (however, it requires knowledge not only of $\mathbf{E}[X]$ but also $\mathbf{V}[X]$!)

Chebyshev's Inequality

Chebyshev's Inequality

For any random variable X with finite $\mathbf{E}[X]$ and $\mathbf{V}[X]$, for any $a > 0$,

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq a] \leq \mathbf{V}[X]/a^2.$$

P. Chebyshev (1821-1894)

Comments:

- can be rewritten as:

The "$\mu \pm$ a few σ" rule. Most of the probability mass is within a few standard deviations from μ.

$$\mathbf{P}\left[|X - \mathbf{E}[X]| \geq \sqrt{\delta \cdot \mathbf{V}[X]}\right] \leq 1/\delta.$$

- Unlike Markov, Chebyshev's inequality is two-sided and also holds for random variables with negative values
- In most cases, Chebyshev's inequality yields much stronger bounds than Markov (however, it requires knowledge not only of $\mathbf{E}[X]$ but also $\mathbf{V}[X]$!)
- Chebyshev's inequality is also known as Second Moment Method

Derivation of Chebychev's inequality

Proof

Derivation of Chebychev's inequality

Proof

Exercise: Can you find a proof that uses Markov's inequality?

Example: Chebychev is (usually) much stronger than Markov

Example 3

Throw an unbiased coin n times and let X be the total number of heads. In an experiment, with n large, we would usually expect a number of heads that is close to the expectation. Can we justify that?

Answer

Outline

Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2.

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2.

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[\, |\overline{X}_n - \mu| > \epsilon \,\right] = 0$$

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] = 0$$

$\forall \epsilon > 0:$

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] = 0$$

$\forall \epsilon > 0 : \forall \delta > 0 :$

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^n X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

$\forall \epsilon > 0\colon \forall \delta > 0\colon \exists N > 0\colon$

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[\, |\overline{X}_n - \mu| > \epsilon \,\right] = 0$$

$\forall \epsilon > 0 \colon \forall \delta > 0 \colon \exists N > 0 \colon \forall n \geq N \colon$

Law of Large Numbers

$=$ independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

$$\forall\epsilon > 0\colon \forall\delta > 0\colon \exists N > 0\colon \forall n \geq N\colon\ \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] \leq \delta$$

Law of Largeb Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] = 0$$

$$\forall \epsilon > 0 \colon \forall \delta > 0 \colon \exists N > 0 \colon \forall n \geq N \colon \; \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] \leq \delta$$

- "Power of Averaging": repeated samples allow us to estimate μ

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

$$\forall \epsilon > 0 \colon \forall \delta > 0 \colon \exists N > 0 \colon \forall n \geq N \colon \; \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] \leq \delta$$

- "Power of Averaging": repeated samples allow us to estimate μ

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

J. Bernoulli (1655-1705)

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^n X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

$$\forall \epsilon > 0\colon \forall \delta > 0\colon \exists N > 0\colon \forall n \geq N\colon \; \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] \leq \delta$$

- "Power of Averaging": repeated samples allow us to estimate μ
- A similar statement holds even if the X_i's are not identically distributed

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

J. Bernoulli (1655-1705)

Law of Large Numbers

= independent and identically distributed

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d.with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] = 0$$

$$\forall \epsilon > 0\colon \forall \delta > 0\colon \exists N > 0\colon \forall n \geq N\colon\ \mathbf{P}\left[\,|\overline{X}_n - \mu| > \epsilon\,\right] \leq \delta$$

- "Power of Averaging": repeated samples allow us to estimate μ
- A similar statement holds even if the X_i's are not identically distributed
- There is also a strong law of large numbers:

$$\mathbf{P}\left[\lim_{n \to \infty} \overline{X}_n = \mu\right] = 1.$$

"For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one's goal."

J. Bernoulli (1655-1705)

Illustration of Weak Law of Large Numbers (1/4)

Illustration of Weak Law of Large Numbers (1/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each

Illustration of Weak Law of Large Numbers (1/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each
- Consider $\widetilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$

Illustration of Weak Law of Large Numbers (1/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each
- Consider $\widetilde{X}_n := \sum_{i=1}^{n} X_i$ for any $n = 0, 1, \ldots, 200$

How does a “typical” realisation look like?

Illustration of Weak Law of Large Numbers (2/4)

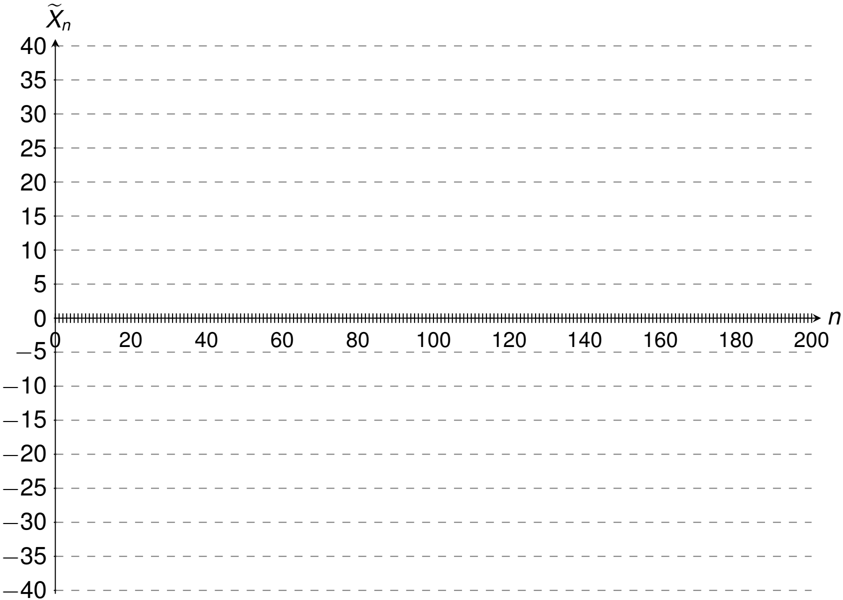


Illustration of Weak Law of Large Numbers (2/4)

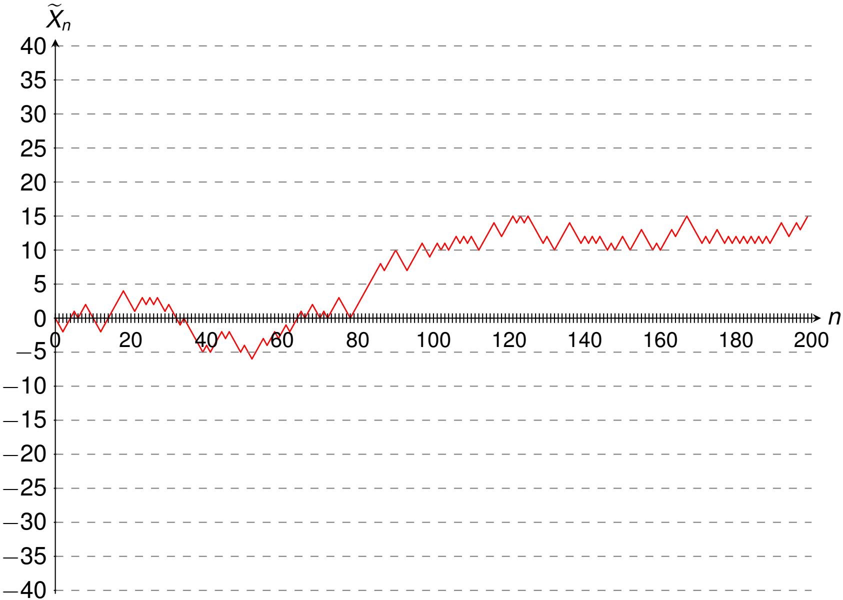


Illustration of Weak Law of Large Numbers (2/4)

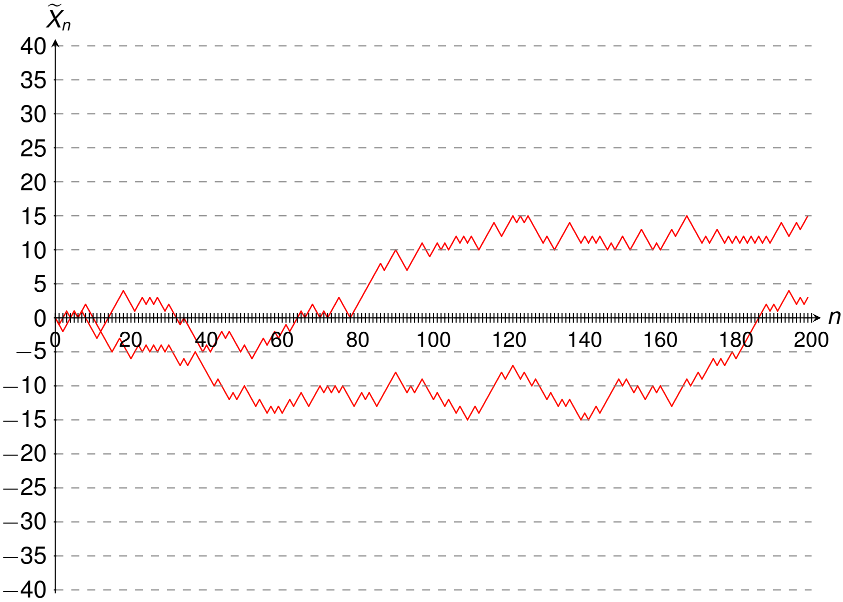


Illustration of Weak Law of Large Numbers (2/4)

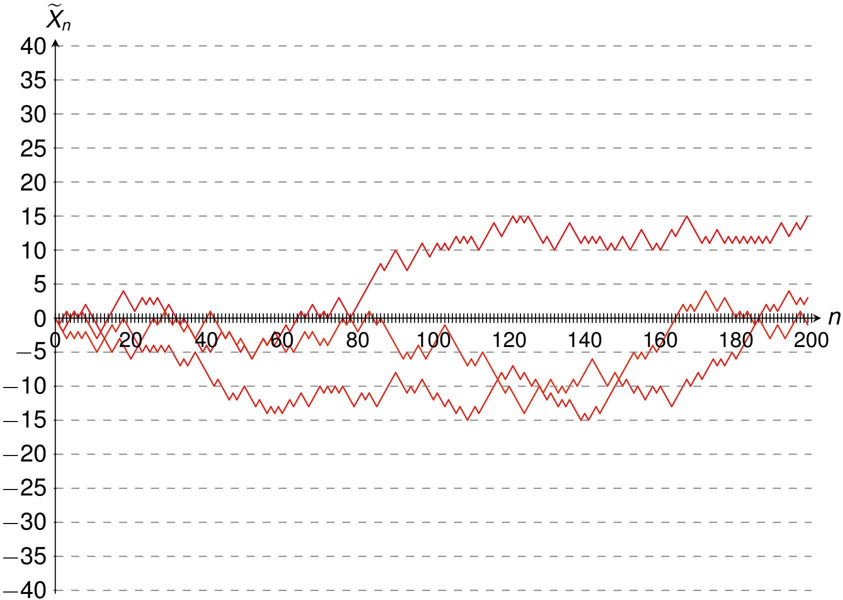


Illustration of Weak Law of Large Numbers (2/4)

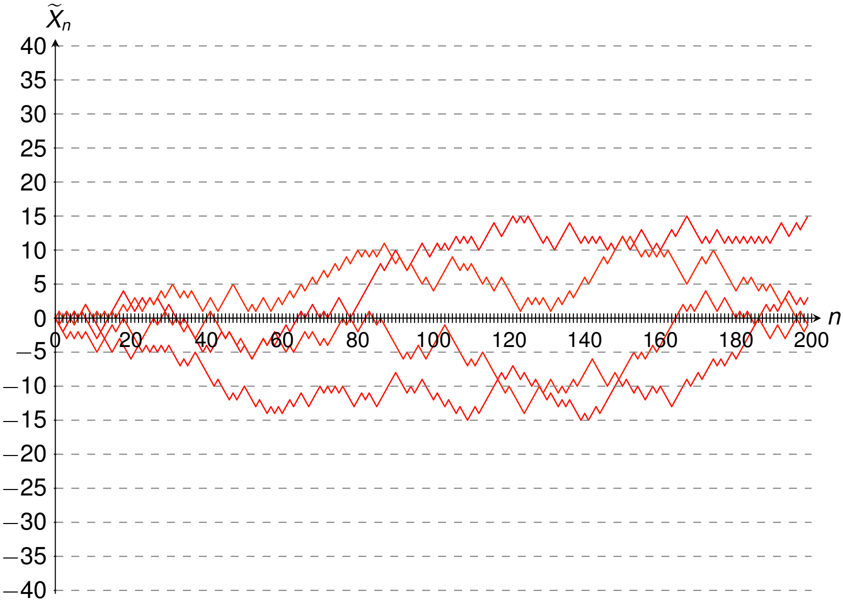


Illustration of Weak Law of Large Numbers (2/4)

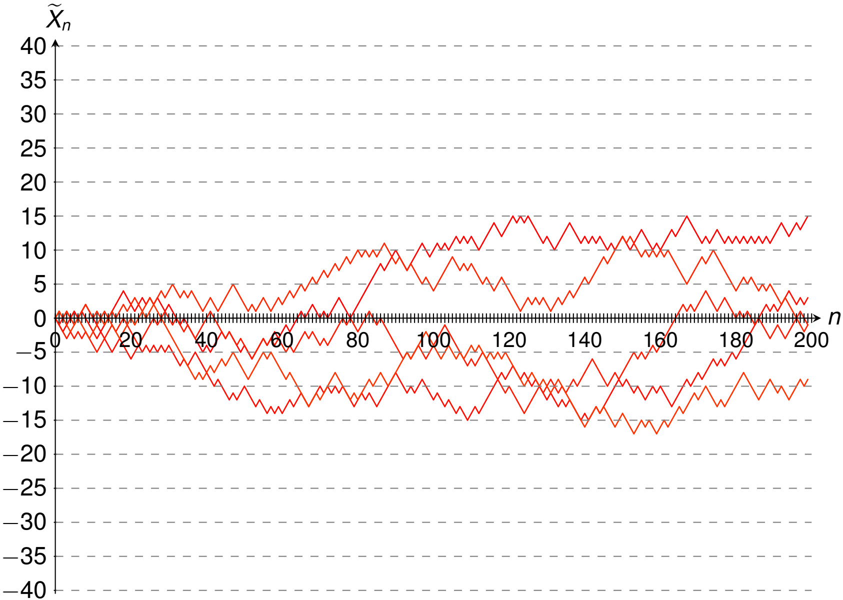


Illustration of Weak Law of Large Numbers (2/4)

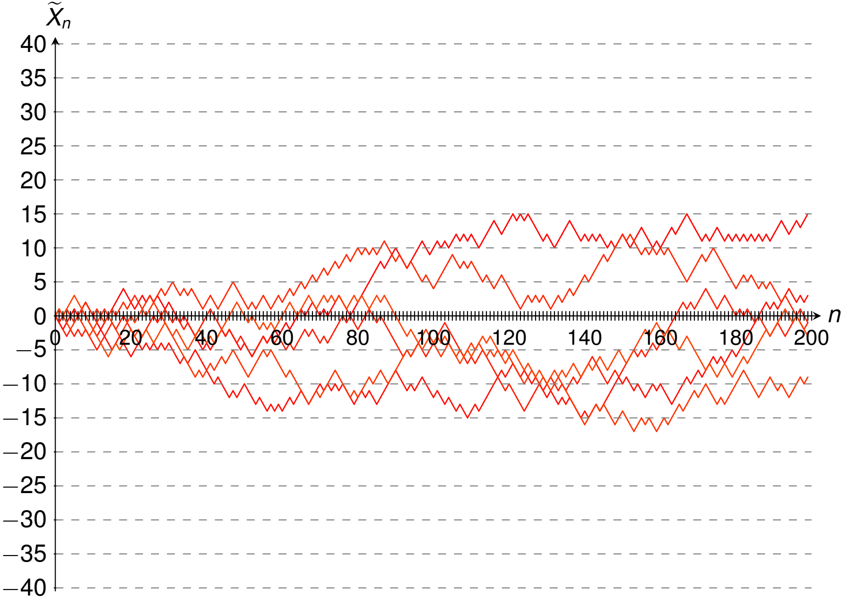


Illustration of Weak Law of Large Numbers (2/4)

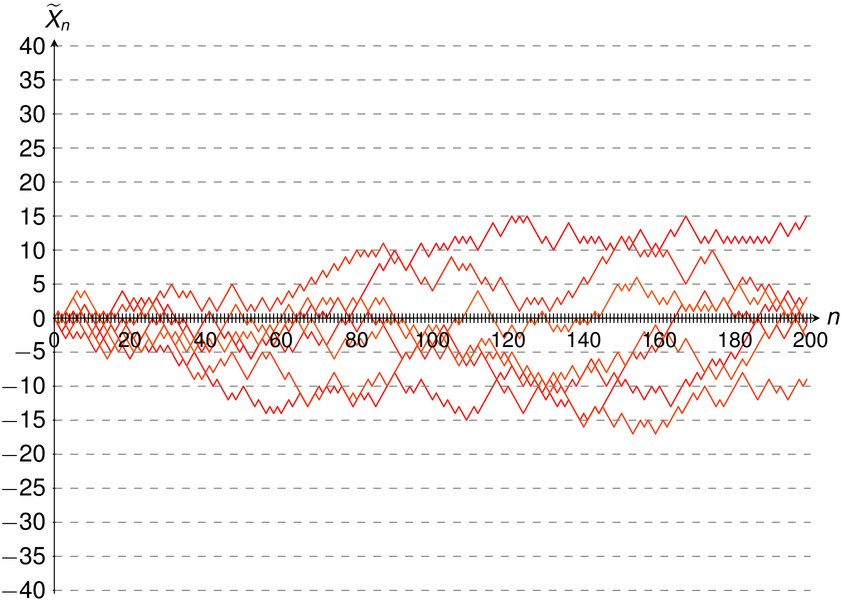


Illustration of Weak Law of Large Numbers (2/4)

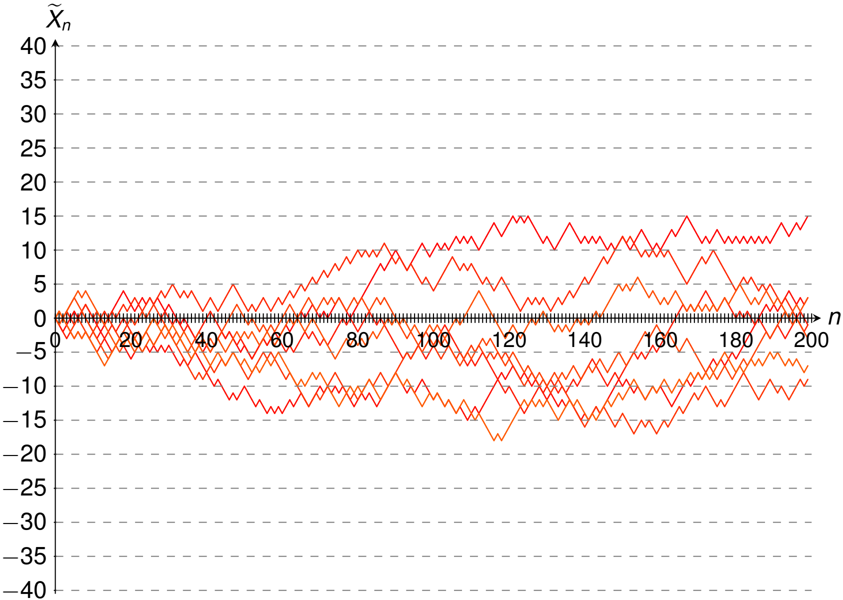


Illustration of Weak Law of Large Numbers (2/4)

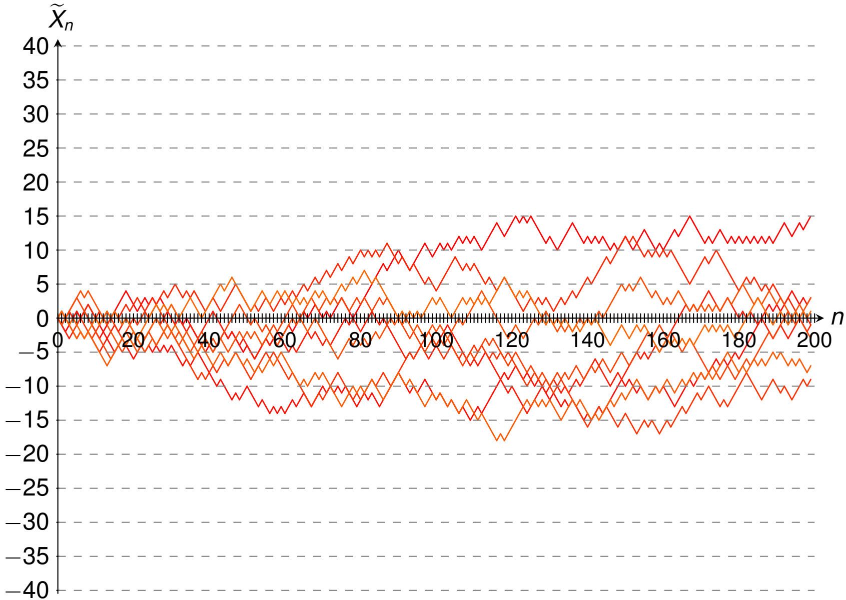


Illustration of Weak Law of Large Numbers (2/4)

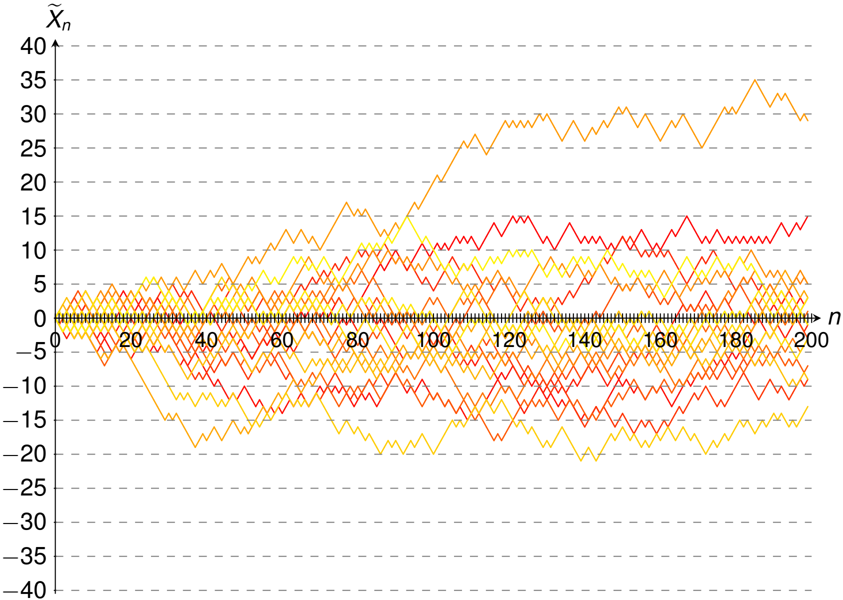
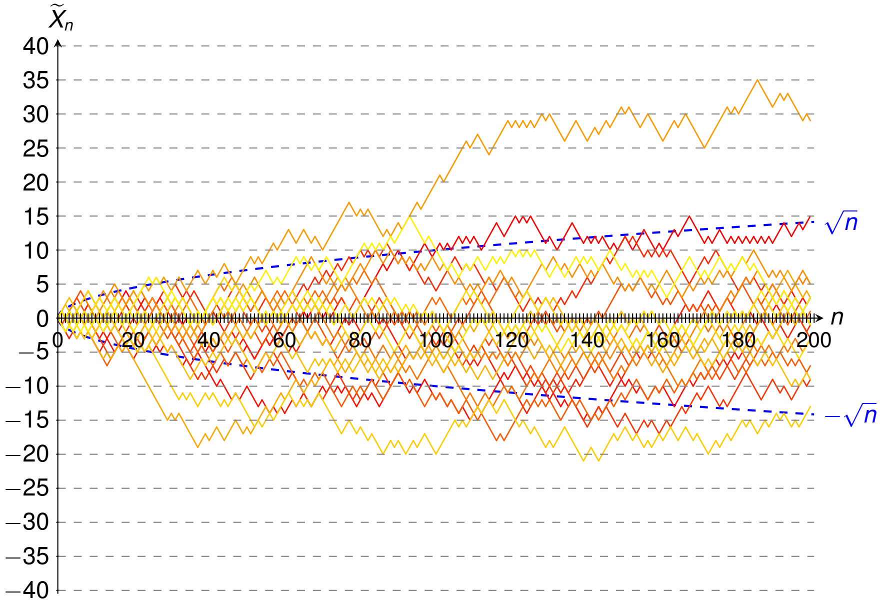
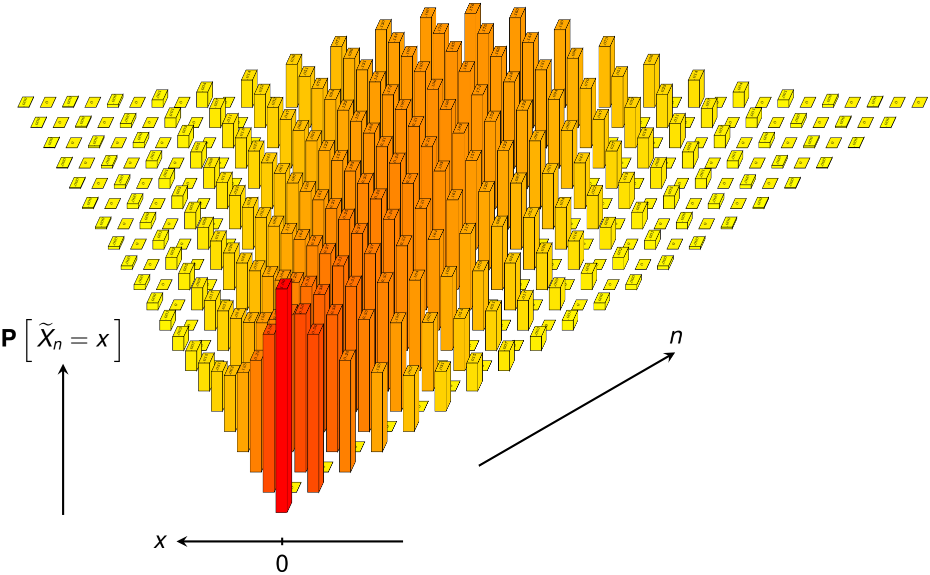


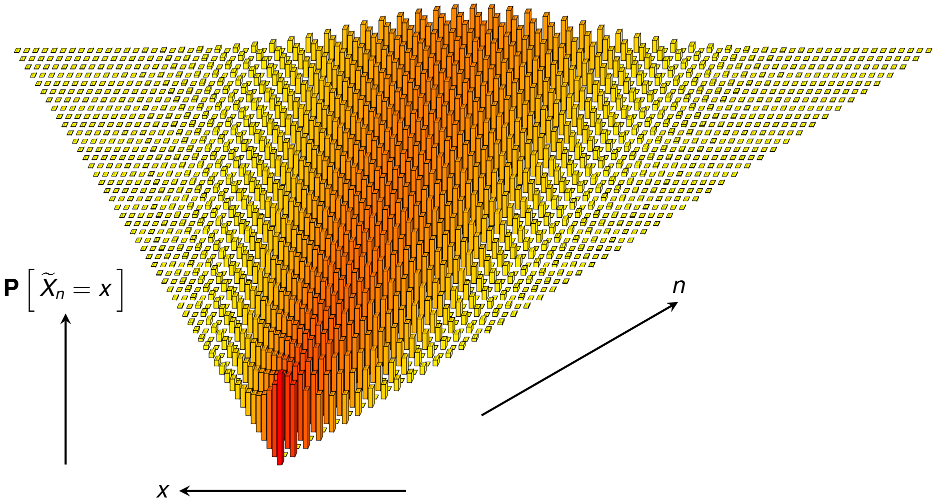
Illustration of Weak Law of Large Numbers (2/4)



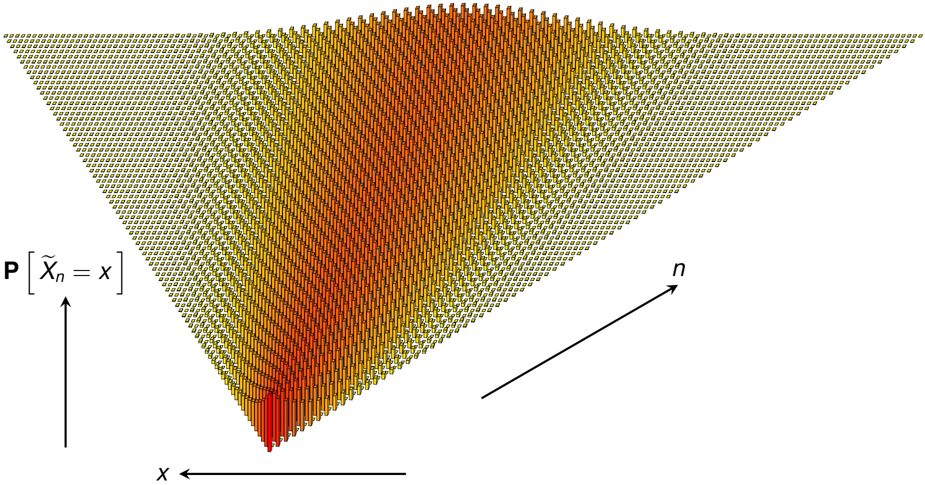
Plot of the Distributions for $n = 0, 1, \ldots, 20$



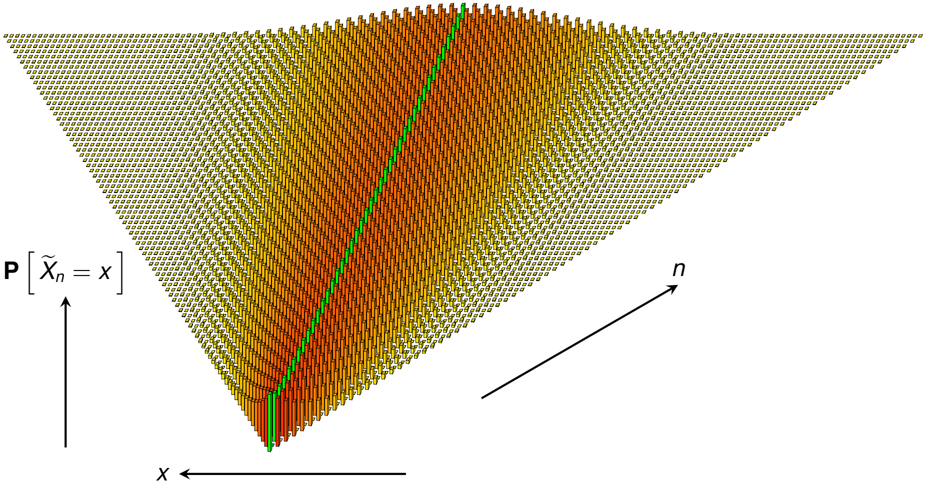
Plot of the Distributions for $n = 0, 1, \ldots, 50$



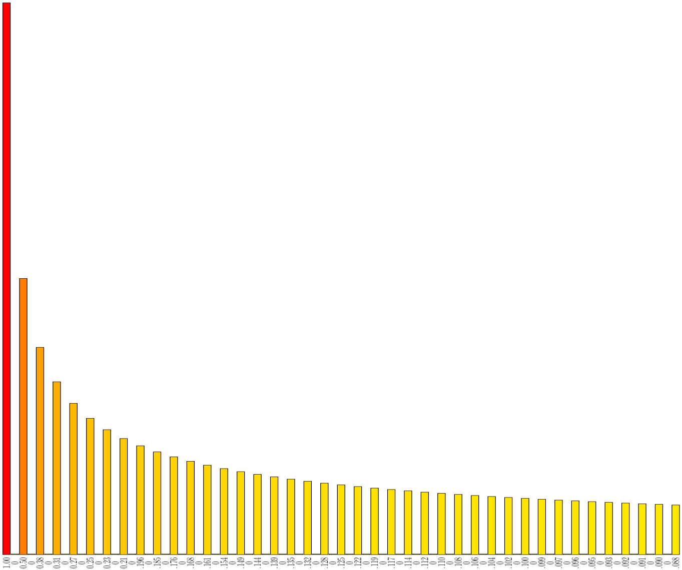
Plot of the Distributions for $n = 0, 1, \ldots, 80$



Plot of the Distributions for $n = 0, 1, \ldots, 80$



Interlude: Approximation of $\mathbf{P}[\,\widetilde{X}_n = 0\,]$



Interlude: Approximation of $\mathbf{P}[\,\widetilde{X}_n = 0\,]$

Exercise

Try to find an expression for $\mathbf{P}\left[\widetilde{X}_n = 0\right]$. Using Stirling's approximation for $n!$, conclude that $\mathbf{P}\left[\widetilde{X}_n = 0\right] = \Theta(1/\sqrt{n})$ for even integers n.

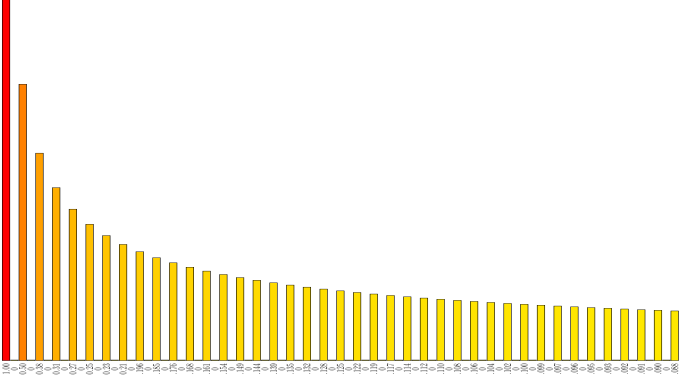


Illustration of Weak Law of Large Numbers (3/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each
- Consider $\widetilde{X}_n := \sum_{i=1}^{n} X_i$ for any for any $n = 0, 1, \ldots, 200$

Illustration of Weak Law of Large Numbers (3/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each
- Consider $\widetilde{X}_n := \sum_{i=1}^{n} X_i$ for any for any $n = 0, 1, \ldots, 200$

This does **not** converge!

Illustration of Weak Law of Large Numbers (3/4)

- Let X_i be independent random variables taking values $\in \{-1, +1\}$ with probability $1/2$ each
- Consider $\widetilde{X}_n := \sum_{i=1}^{n} X_i$ for any for any $n = 0, 1, \ldots, 200$

This does **not** converge!

Consider now the average (sample mean): $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$.

Illustration of Weak Law of Large Numbers (4/4)

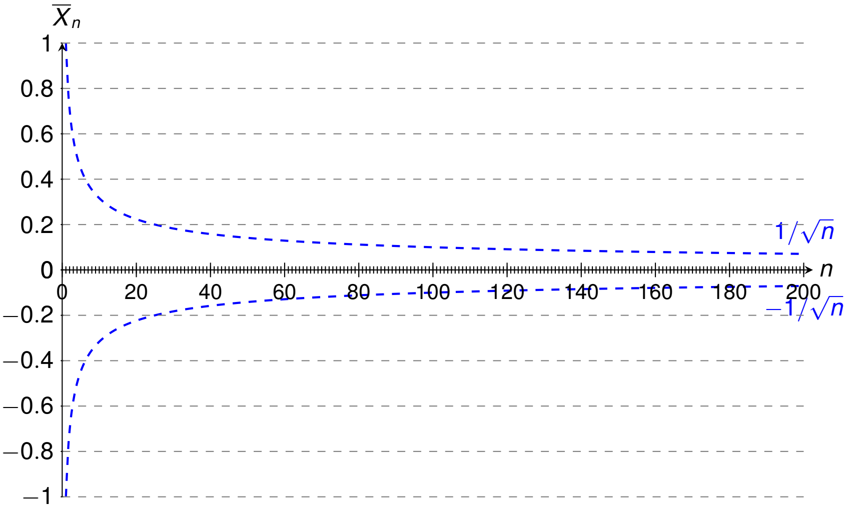


Illustration of Weak Law of Large Numbers (4/4)

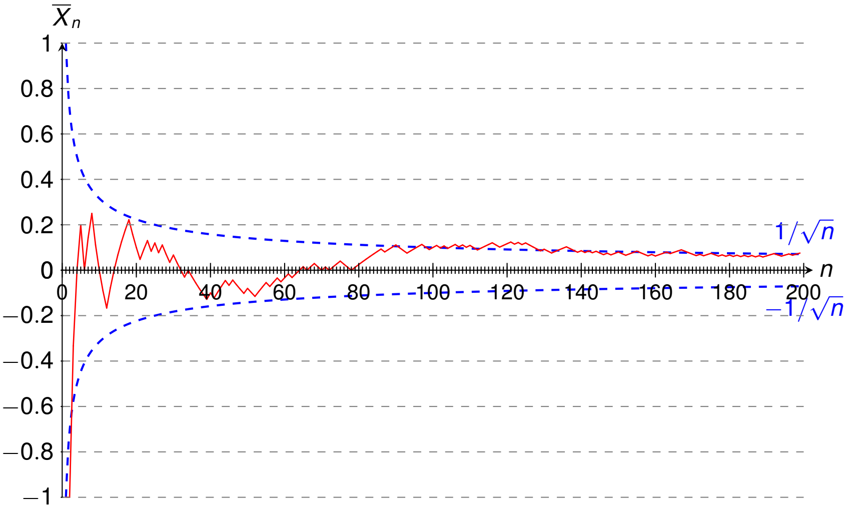


Illustration of Weak Law of Large Numbers (4/4)

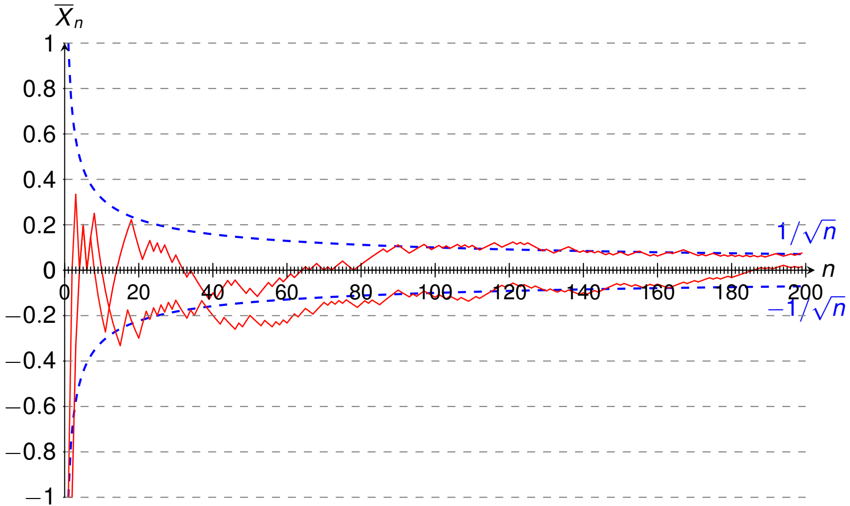


Illustration of Weak Law of Large Numbers (4/4)

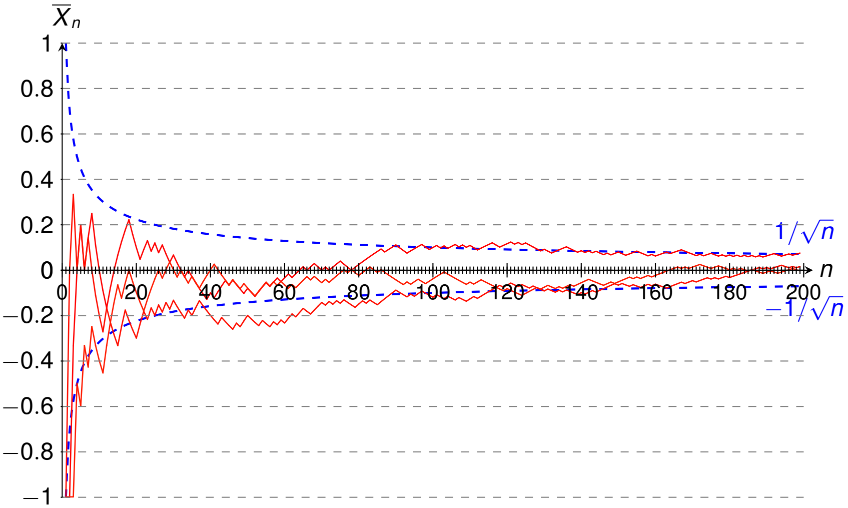


Illustration of Weak Law of Large Numbers (4/4)

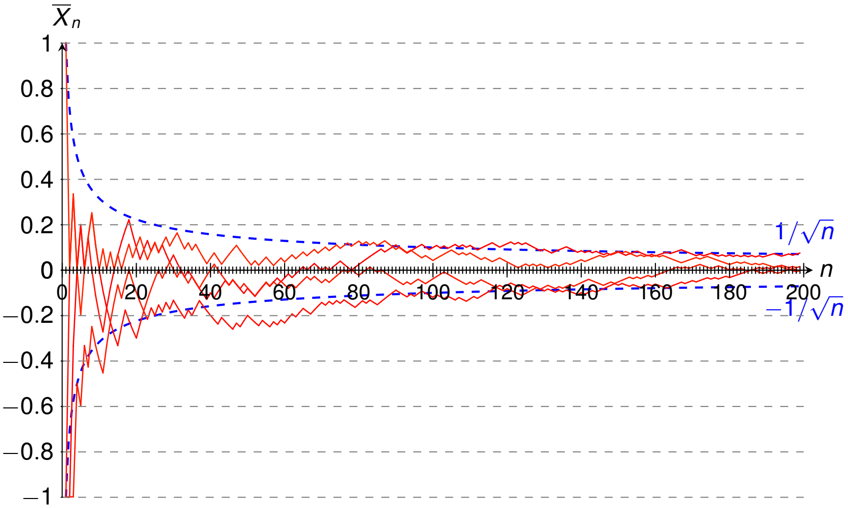
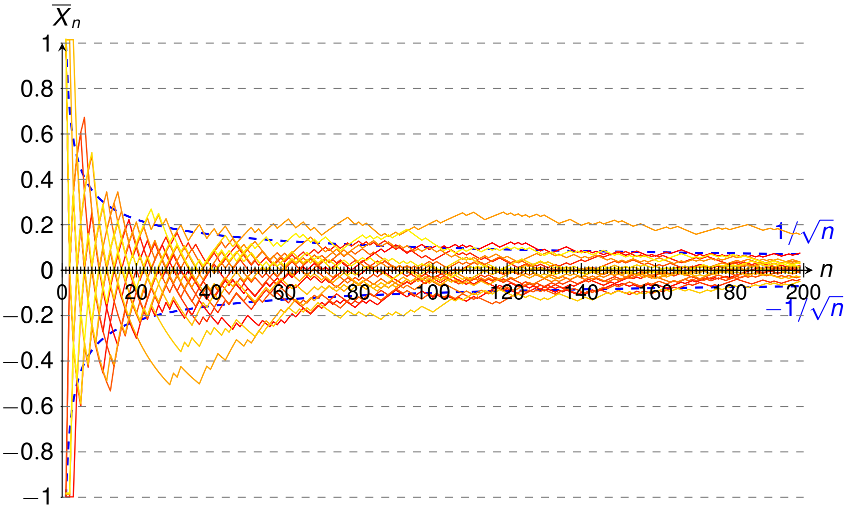


Illustration of Weak Law of Large Numbers (4/4)



Proof of the Weak Law of Large Numbers

The Weak Law of Large Numbers

Let $\overline{X}_n := 1/n \cdot \sum_{i=1}^{n} X_i$, where the X_i's are i.i.d. with finite expectation μ and finite variance σ^2. Then, for any $\epsilon > 0$,

$$\lim_{n \to \infty} \mathbf{P}\left[|\overline{X}_n - \mu| > \epsilon \right] = 0$$

Proof

Inferring Probabilities of an Event

Example 4

Suppose that, instead of the expectation μ, we want to estimate the probability of an event, e.g.,

$$p := \mathbf{P}\left[X \in (a, b]\right], \text{ where } a < b.$$

How can we use the Law of Large Numbers?

Answer

Appendix: Sum of Two Uniform R.V. (non-examinable)

Example

Let X and Y be two independent random variables, both uniformly distributed on $[0, 1]$. How does the probability density of $X + Y$ look like?

Answer

Appendix: Sum of Two Uniform R.V. (non-examinable)

Example

Let X and Y be two independent random variables, both uniformly distributed on $[0,1]$. How does the probability density of $X+Y$ look like?

Answer

We have

$$f_{X+Y}(a) \overset{(\star)}{=} \int_{-\infty}^{+\infty} f_X(a-y) f_Y(y) dy,$$

where for $(\star)$, see Chapter 6.3 in Ross (Chapter 11.2 in Dekking et al.). Since $f_Y(y) = 1$ if $0 \leq y \leq 1$ and $f_Y(y) = 0$ otherwise, we have

$$f_{X+Y}(a) = \int_0^1 f_X(a-y) dy.$$

Further, for $0 \leq a \leq 1$ we have $f_X(a-y) = 1$ and $f_X(a-y) = 0$ otherwise, and thus

$$f_{X+Y}(a) = \int_0^a dy = a.$$

Similarly, for $1 < a < 2$, $f_{X+Y}(a) = \int_a^2 dy = 2 - a$. Therefore,

$$f_{X+Y}(a) = \begin{cases} a & \text{if } 0 \leq a \leq 1, \\ 2-a & \text{if } 1 \leq a \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$