CS 121: Lecture 23
Probability Review

Madhu Sudan

https://madhu.seas.Harvard.edu/courses/Fall2020

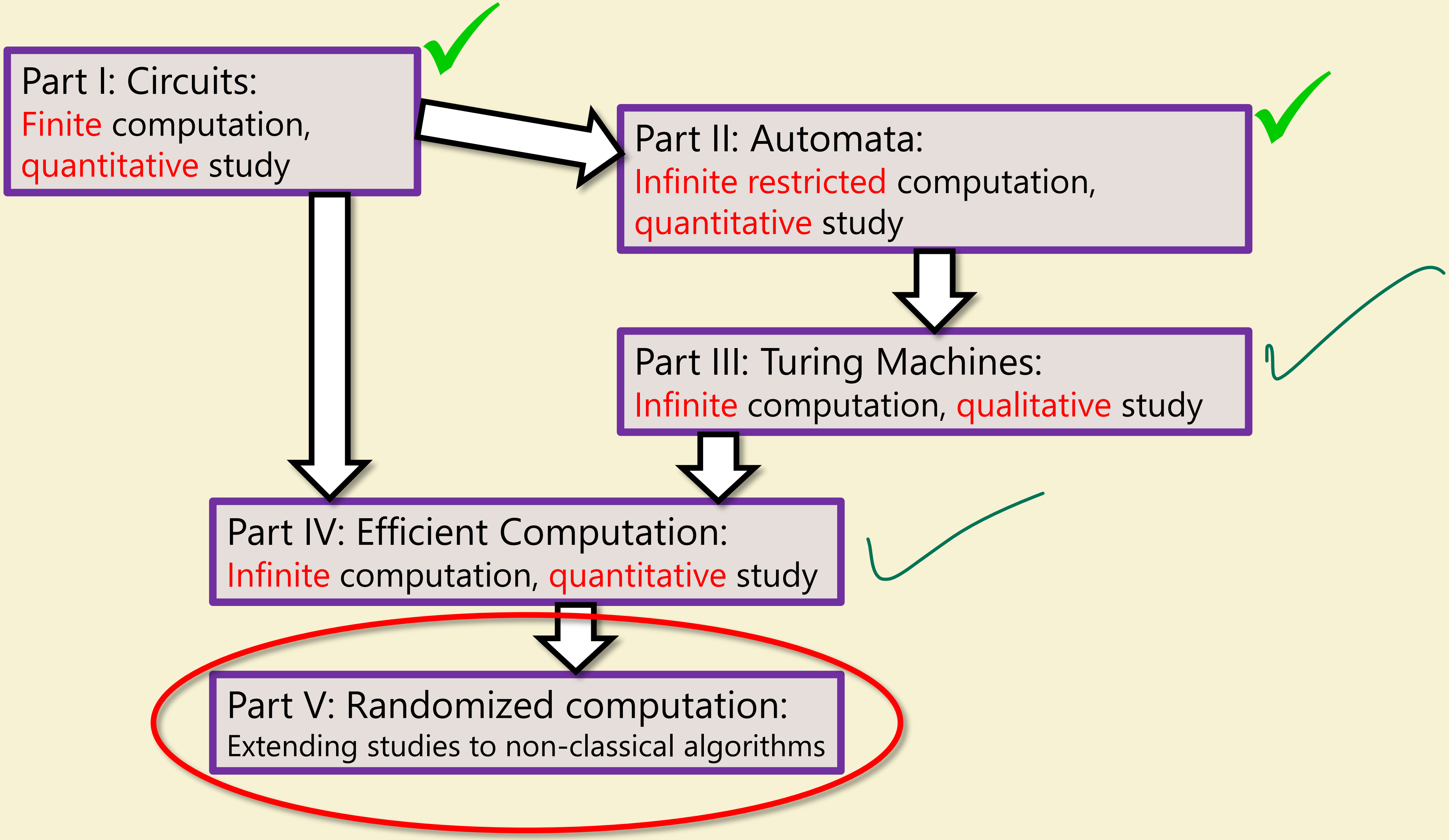
Book: https://introtcs.org

How to contact us
- The whole staff (faster response): CS 121 Piazza
- Only the course heads (slower): cs121.fall2020.course.heads@gmail.com

Announcements:

- Advanced section: Roei Tell (De)Randomization
- 2nd feedback survey
- Homework 6 out today. Due 12/3/2020
- Section 11 week starts

Where we are:


Part I: Circuits:
Finite computation, quantitative study
Part II: Automata:
Infinite restricted computation, quantitative study
Part III: Turing Machines:
Infinite computation, qualitative study
Part IV: Efficient Computation:
Infinite computation, quantitative study
Part V: Randomized computation:
Extending studies to non-classical algorithms

Review of course so far

- Circuits:
  - Compute every finite function … but no infinite function
  - Measure of complexity: SIZE. $\mathrm{ALL}_n \subseteq \mathrm{SIZE}\left(O\left(\frac{2^n}{n}\right)\right)$; $\exists f \notin \mathrm{SIZE}\left(o\left(\frac{2^n}{n}\right)\right)$ — Counting
- Finite Automata:
  - Compute some infinite functions (all Regular functions)
  - Do not compute many (easy) functions (e.g. $\mathrm{EQ}(x, y) = 1 \Leftrightarrow x = y$) — Pumping/Pigeonhole
- Turing Machines:
  - Compute everything computable (definition/thesis)
  - HALT is not computable: — Diagonalization, Reductions
- Efficient Computation:
  - P Polytime computable Boolean functions – our notion of efficiently computable
  - NP Efficiently verifiable Boolean functions. Our wishlist …
  - NP-complete: Hardest problems in NP: 3SAT, ISET, MaxCut, EU3SAT, Subset Sum — Polytime Reductions

Last Module: Challenges to STCT

- Strong Turing-Church Thesis: Everything physically computable in polynomial time can be computable in polynomial time on Turing Machine.
- Challenges:
  - Randomized algorithms: Already being computed by our computers:
    - Weather simulation, Market prediction, …
  - Quantum algorithms:
- Status:
  - Randomized: May not add power?
  - Quantum: May add power?
  - Still can be studied with same tools. New classes: BPP, BQP:
    - B – bounded error; P – Probabilistic, Q - Quantum

Today: Review of Basic Probability

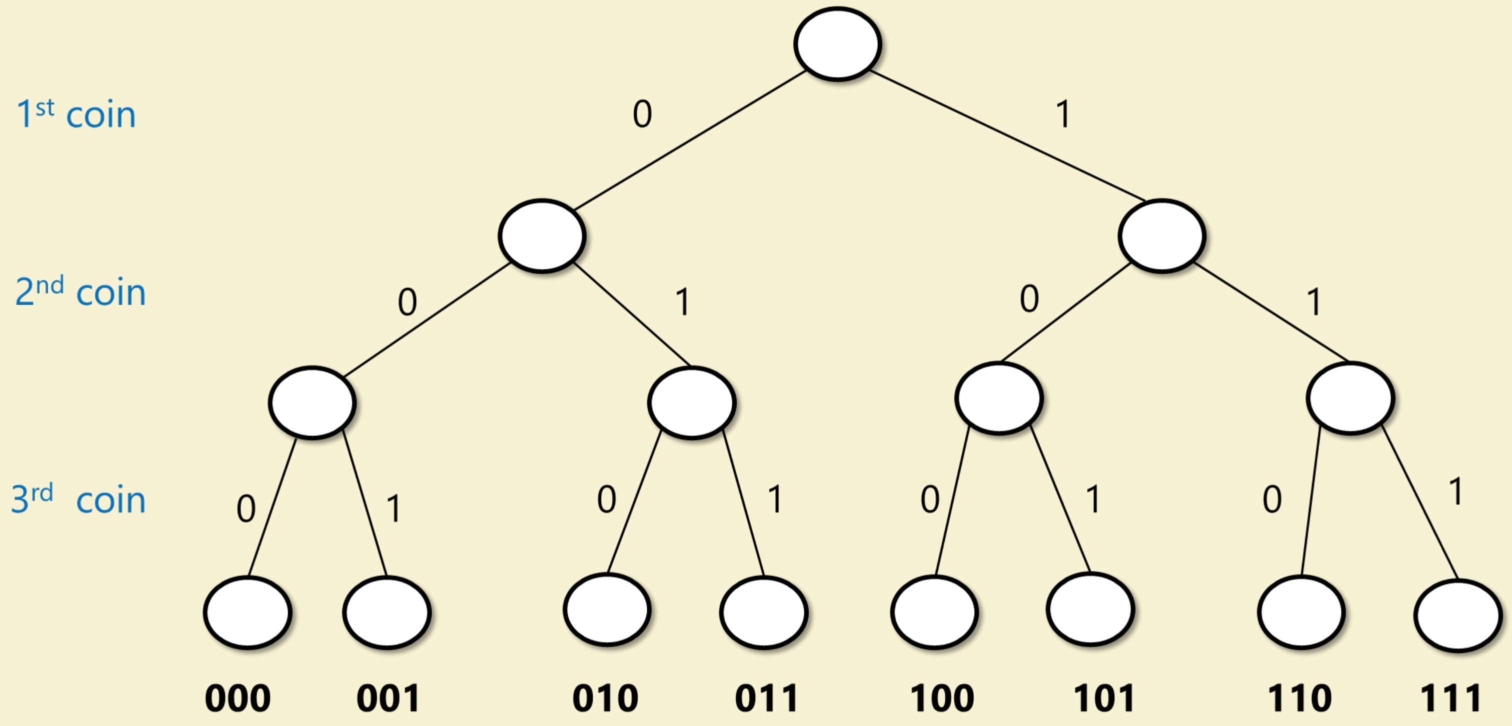
Read Chapter 18

- Sample space
- Events
- Union/intersection/negation – AND/OR/NOT of events
- Random variables
- Expectation
- Concentration / tail bounds

Throughout this lecture

Probabilistic experiment: tossing n independent unbiased coins.

Equivalently: Choose $x \sim \{0,1\}^n$

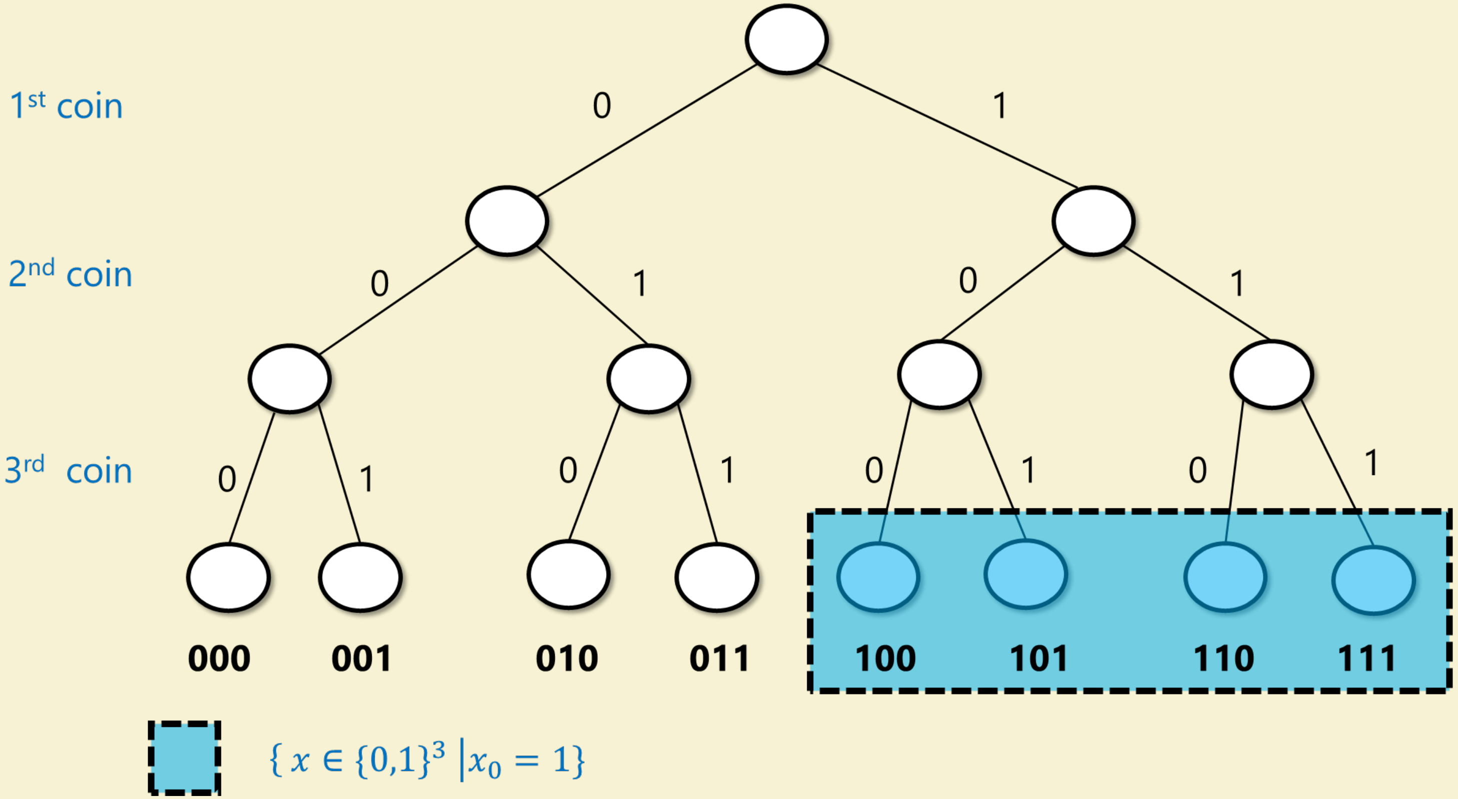


Events

Fix sample space to be $x \sim \{0,1\}^n$

An event is a set $A \subseteq \{0,1\}^n$. Probability that A happens is $\Pr[A] = \frac{|A|}{2^n}$

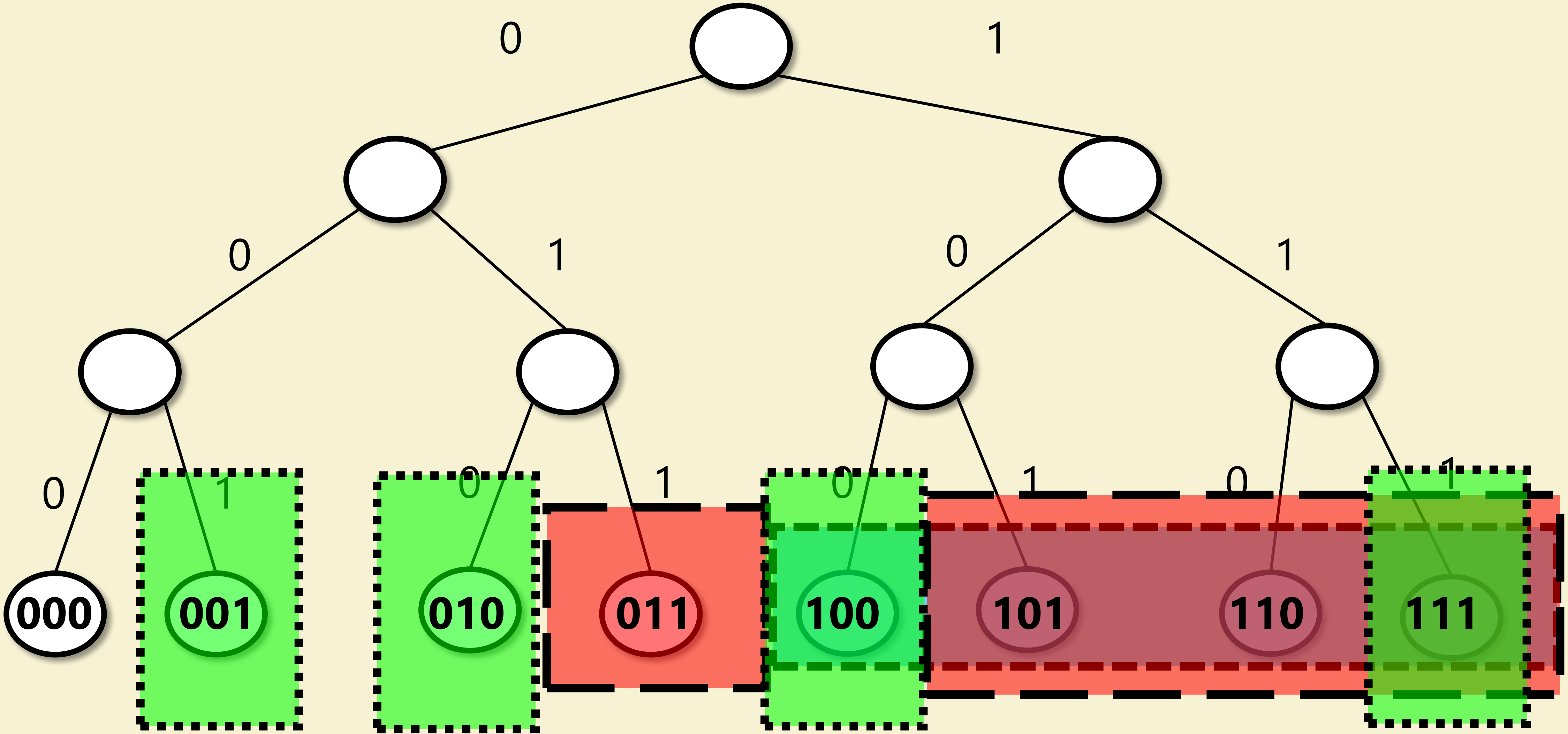
Example: If $x \sim \{0,1\}^3$, $\Pr[x_0 = 1] = \frac{4}{8} = \frac{1}{2}$



Q: Let $n = 3$,

$A = \{x_0 = 1\}$, $B = \{x_0 + x_1 + x_2 \geq 2\}$, $C = \{x_0 + x_1 + x_2 = 1\ mod\ 2\}$.

What are (i) $\Pr[B]$, (ii) $\Pr[C]$, (iii) $\Pr[A \cap B]$, (iv) $\Pr[A \cap C]$ (v) $\Pr[B \cap C]$?

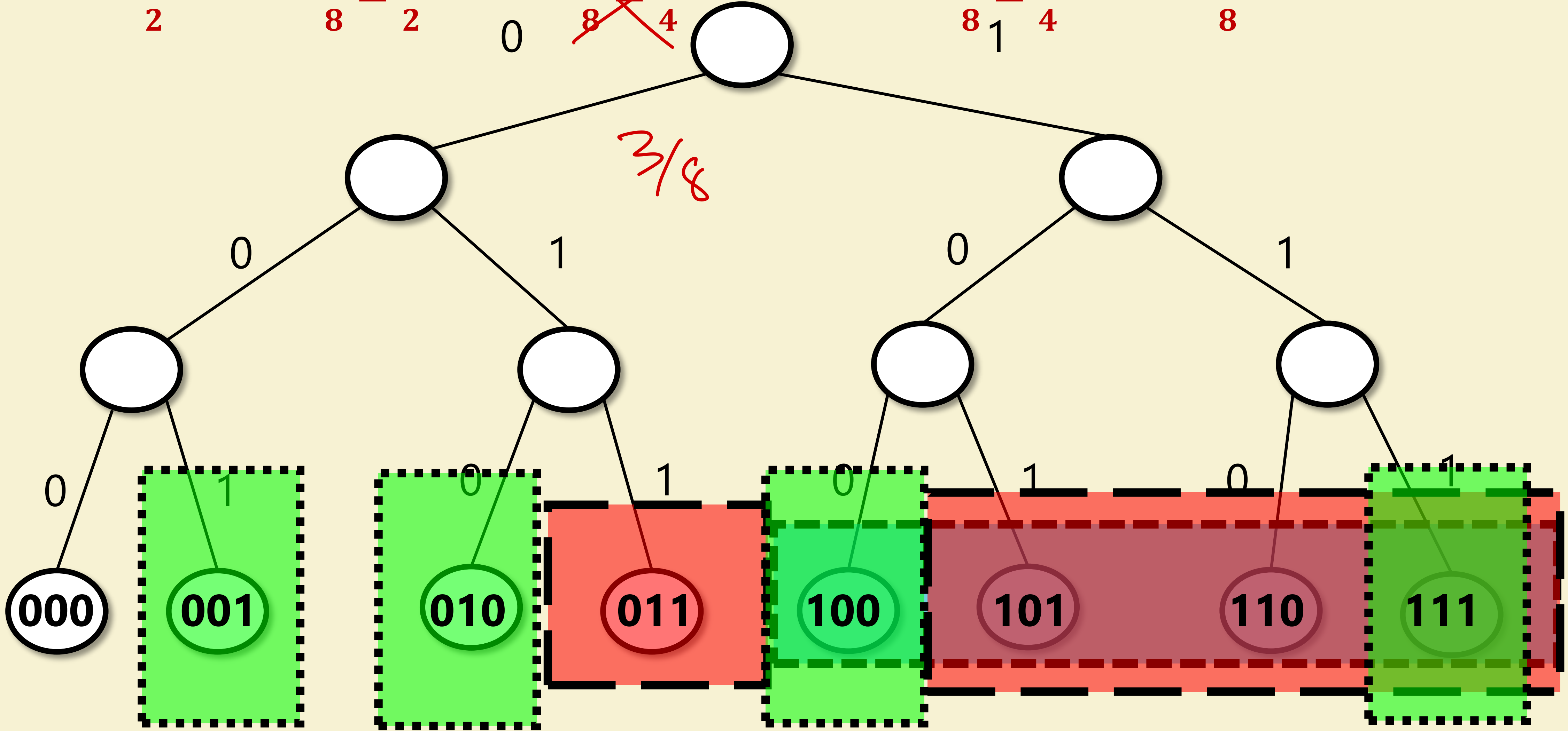


Q: Let $n = 3$,

$\boxed{A = \{x_0 = 1\}}$, $\boxed{B = \{x_0 + x_1 + x_2 \geq 2\}}$, $\boxed{C = \{x_0 + x_1 + x_2 = 1 \bmod 2\}}$.

What are (i) $\Pr[B]$, (ii) $\Pr[C]$, (iii) $\Pr[A \cap B]$, (iv) $\Pr[A \cap C]$ (v) $\Pr[B \cap C]$?

$\frac{1}{2}$ $\quad$ $\frac{4}{8} = \frac{1}{2}$ $\quad$ $\cancel{\frac{2}{8} = \frac{1}{4}}$ $\quad$ $\frac{2}{8} = \frac{1}{4}$ $\quad$ $\frac{1}{8}$

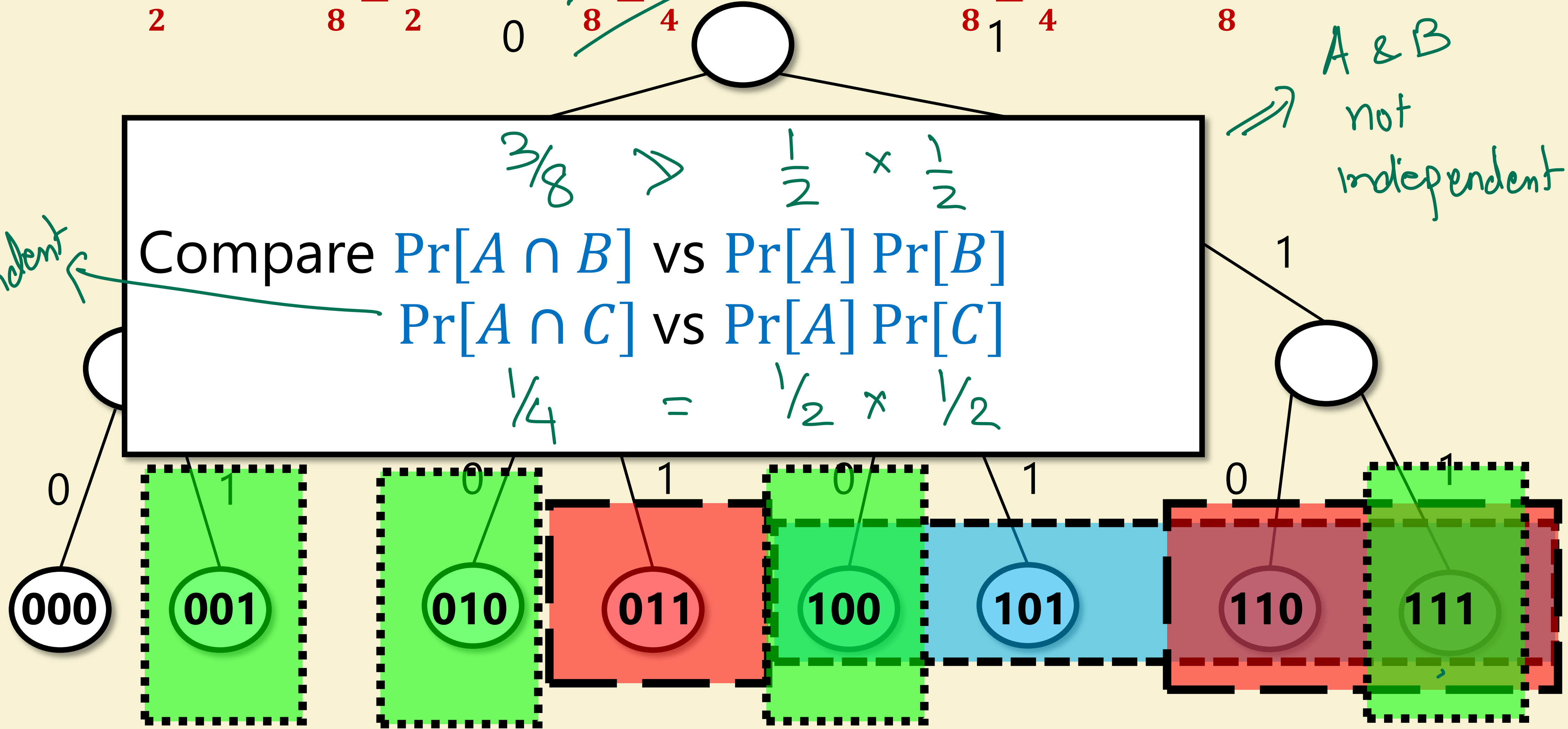


Q: Let $n = 3$,

$A = \{x_0 = 1\}$, $B = \{x_0 + x_1 + x_2 \geq 2\}$, $C = \{x_1 + x_2 + x_3 = 1 \bmod 2\}$.

What are (i) $\Pr[B]$, (ii) $\Pr[C]$, (iii) $\Pr[A \cap B]$, (iv) $\Pr[A \cap C]$ (v) $\Pr[B \cap C]$?

$\frac{1}{2}$ $\quad \frac{4}{8} = \frac{1}{2}$ $\quad \cancel{\frac{2}{8} = \frac{1}{4}}$ 3/8 $\quad \frac{2}{8} = \frac{1}{4}$ $\quad \frac{1}{8}$



3/8 $\gg$ $\frac{1}{2} \times \frac{1}{2}$

Compare $\Pr[A \cap B]$ vs $\Pr[A]\Pr[B]$

$\Pr[A \cap C]$ vs $\Pr[A]\Pr[C]$

1/4 = 1/2 × 1/2

A & C are independent

$\Rightarrow$ A & B not independent

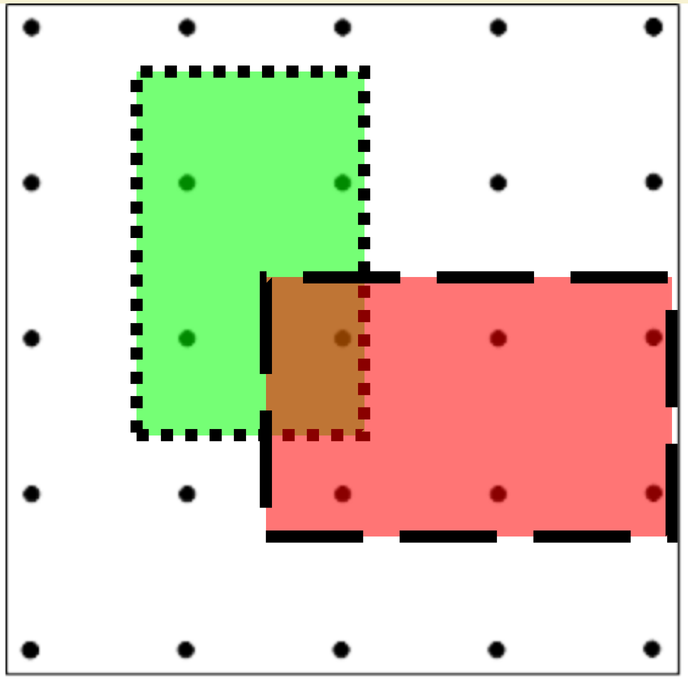
Operations on events

$\Pr[\,A\ or\ B\ happens] = \Pr[\,A \cup B]$

$\Pr[\,A\ and\ B\ happen] = \Pr[\,A \cap B]$

Exercise

$\Pr[\,A\ doesn't\ happen] = \boxed{\Pr[\,\overline{A}]} = \Pr[\{0,1\}^n \setminus A] = \underline{1 - \Pr[A]}$



Q: Prove the union bound: $\Pr[A \vee B] \leq \Pr[A] + \Pr[B]$

Example: $\Pr[A] = 4/25$, $\Pr[B] = 6/25$, $\Pr[A \cup B] = 9/25$

$|A \cup B| = |A| + |B| - |A \cap B| \leq |A| + |B|$

Independence

$\Pr[B|A] = \Pr[B]$

"Conditioning is the soul of statistics"

Informal: Two events A, B are independent if knowing A happened doesn't give any information on whether B happened.

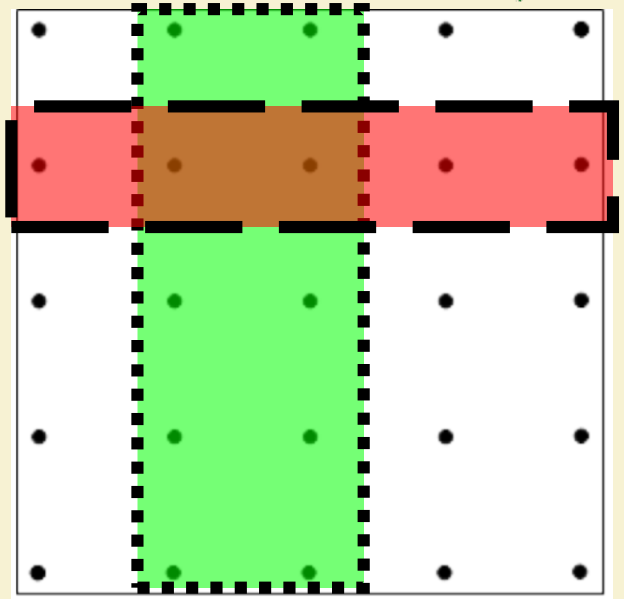
Formal: Two events A, B are independent if $\Pr[A \cap B] = \Pr[A]\Pr[B]$

Q: Which pairs are independent?

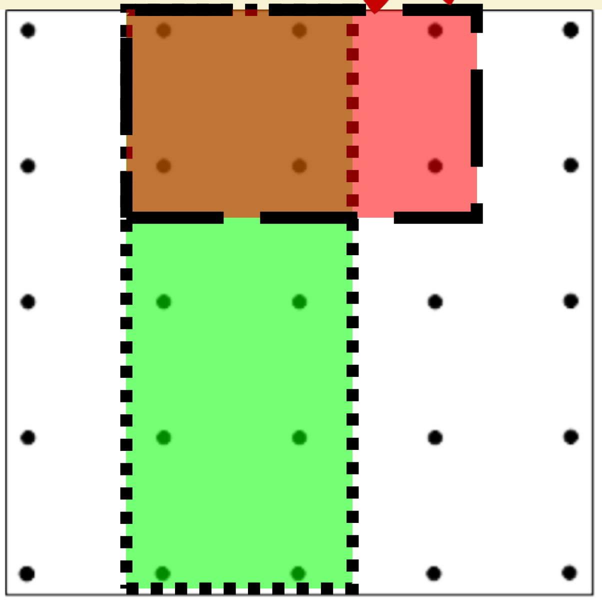
A = Green

B = Red

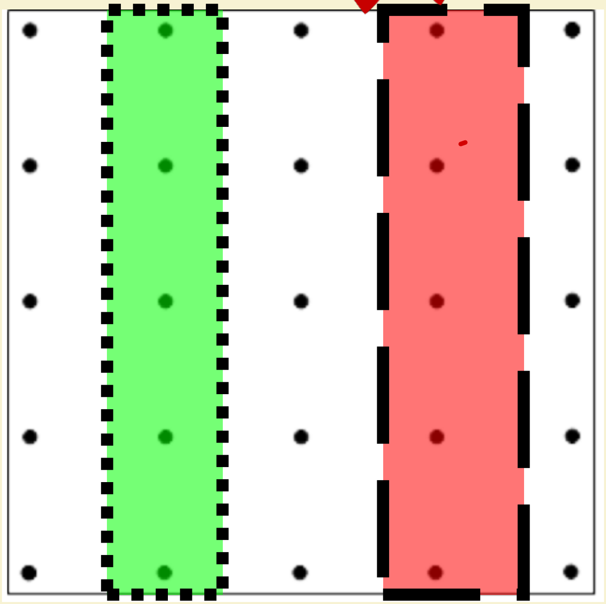
(i)



(ii)



(iii)

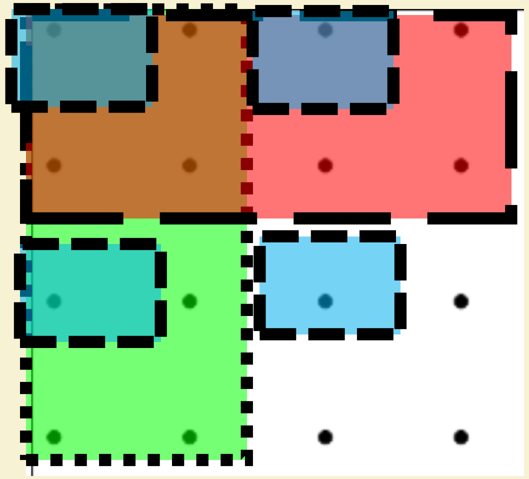


More than 2 events

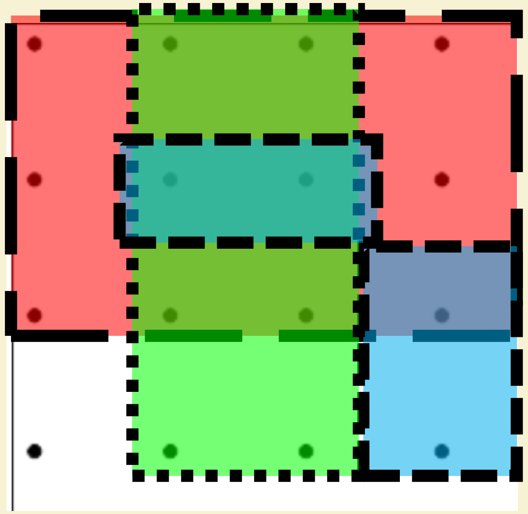
Informal: Two events A, B are independent if knowing whether A happened doesn't give any information on whether B happened.

Formal: Three events A, B, C are independent if every pair A, B , A, C and B, C is independent and $\Pr[A \cap B \cap C] = \Pr[A]\Pr[B]\Pr[C]$

(i)



(ii)



Random variables

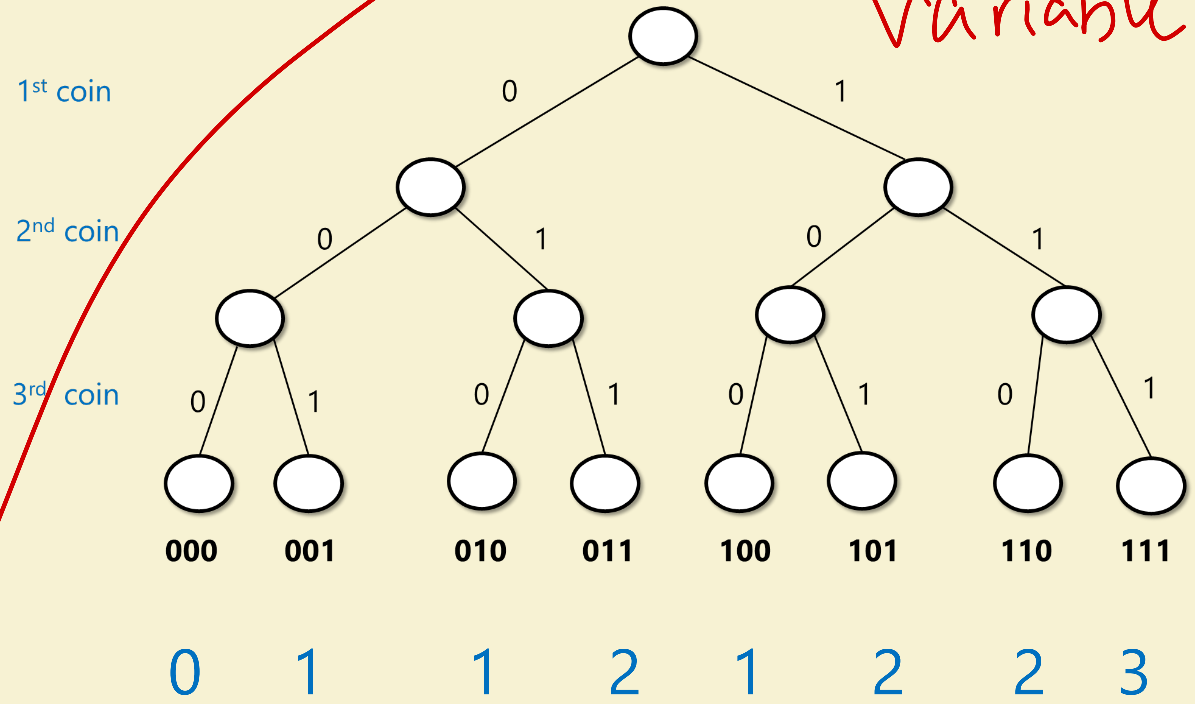
$f: \{0,1\}^n \to \mathbb{R}$ is a random variable

Assign a number to every outcome of the coins.

Formally r.v. is $X: \{0,1\}^n \to \mathbb{R}$

Example: $X(x) = x_0 + x_1 + x_2$

v	$Pr[X = v]$
0	1/8
1	3/8
2	3/8
3	1/8



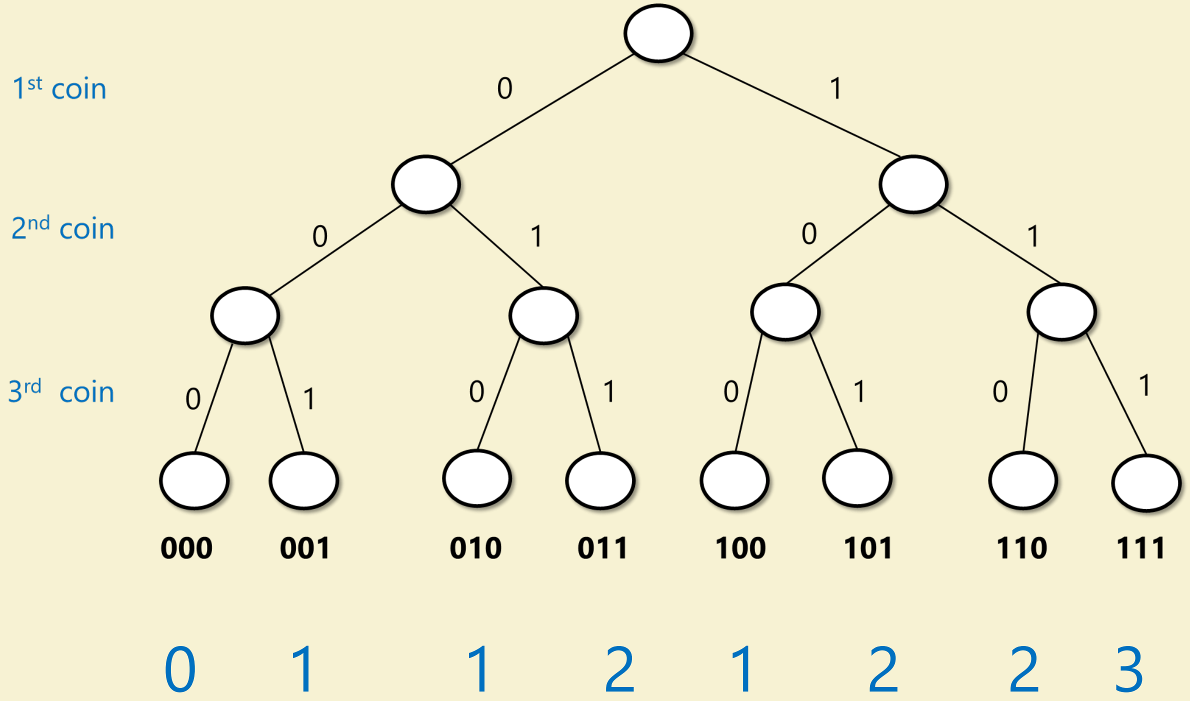
if $\{0,1\}^n$ is our prob. space.

Expectation

Average value of X : $\mathbb{E}[X] = \sum_{x\in\{0,1\}^n} 2^{-n} X(x) = \sum_{v\in\mathbb{R}} v \cdot \Pr[X = v]$

Example: $X(x) = x_0 + x_1 + x_2$

$\boldsymbol{v}$	$\boldsymbol{Pr[X = v]}$
0	1/8
1	3/8
2	3/8
3	1/8



Q: What is $\mathbb{E}[X]$? $0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$

Linearity of expectation

Lemma: $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Corollary: $\mathbb{E}[x_0 + x_1 + x_2] = \mathbb{E}[x_0] + \mathbb{E}[x_1] + \mathbb{E}[x_2] = 1.5$

Proof:

$$\mathbb{E}[X+Y] = 2^{-n} \sum_x (X(x) + Y(x)) = 2^{-n} \sum_x X(x) + 2^{-n} \sum_x Y(x) = \mathbb{E}[X] + \mathbb{E}[Y]$$

Independent random variables

Def: X, Y are independent if $\{X = u\}$ and $\{Y = v\}$ are independent $\forall u, v$

Def: $X_0, \dots, X_{k-1}$, independent if $\{X_0 = v_0\}, \dots, \{X_{k-1} = v_{k-1}\}$ ind. $\forall u_0 \dots u_{k-1}$

i.e. $\forall v, \dots, v_{k-1}\ \forall S \subseteq [k]$

$$\Pr\left[\bigwedge_{i \in S} X_i = v_i\right] = \prod_{i \in S} \Pr[X_i = v_i]$$

Q: Let $x \sim \{0,1\}^n$. Let $X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}$

Let $Y_0 = Y_1 = \cdots = Y_{k-1} = x_0$

Are $X_0, \dots, X_{n-1}$ independent?

Are $Y_0, \dots, Y_{n-1}$ independent?

Independence and concentration

Q: Let $x \sim \{0,1\}^n$. Let $X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$

Let $Y_0 = Y_1 = \cdots = Y_{k-1} = x_0$

Let $X = X_0 + \cdots + X_{n-1}$, $Y = Y_0 + \cdots + Y_{n-1}$

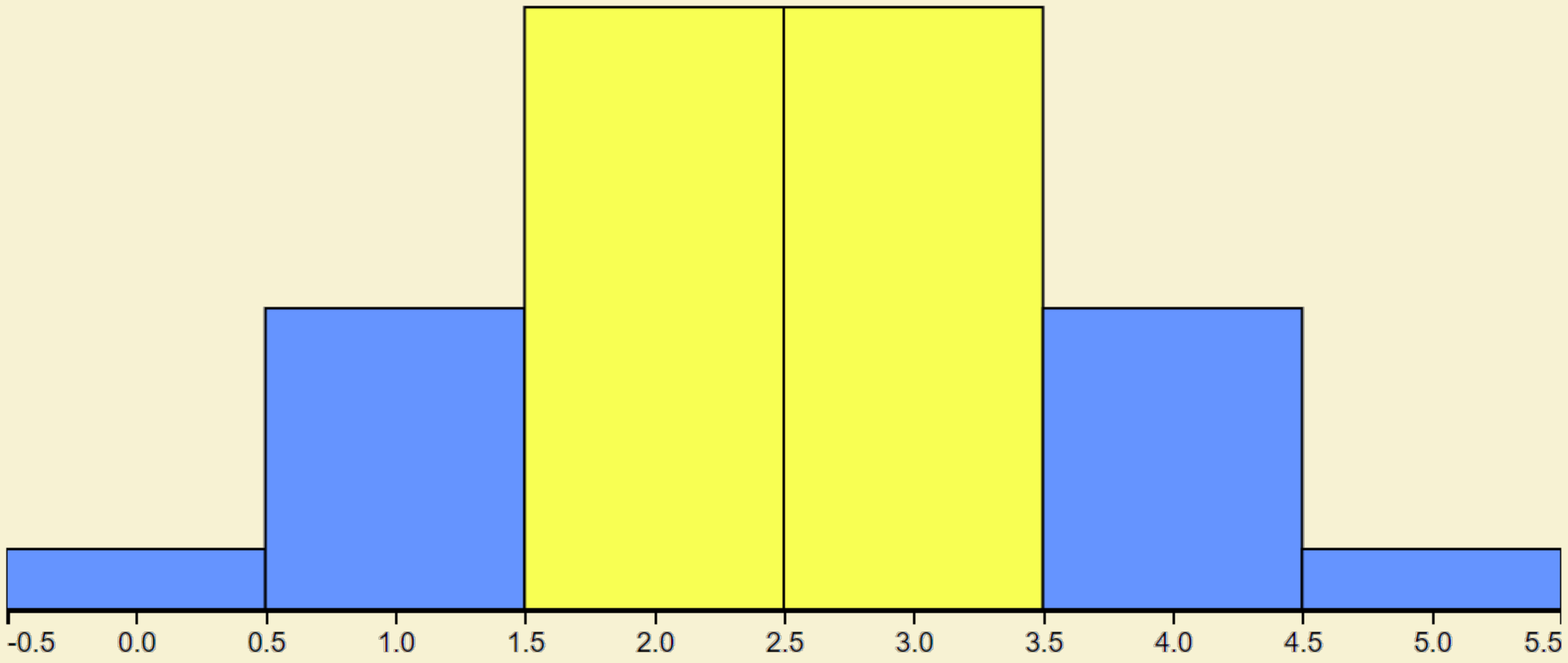
Compute $\mathbb{E}[X]$, compute $\mathbb{E}[Y]$

0

For $n = 100$, estimate $\Pr[Y \notin (0.4n, 0.6n)]$, $\Pr[X \notin (0.4n, 0.6n)]$

$\Pr[\text{not}\,(.4n \le Y \le .6n)]$

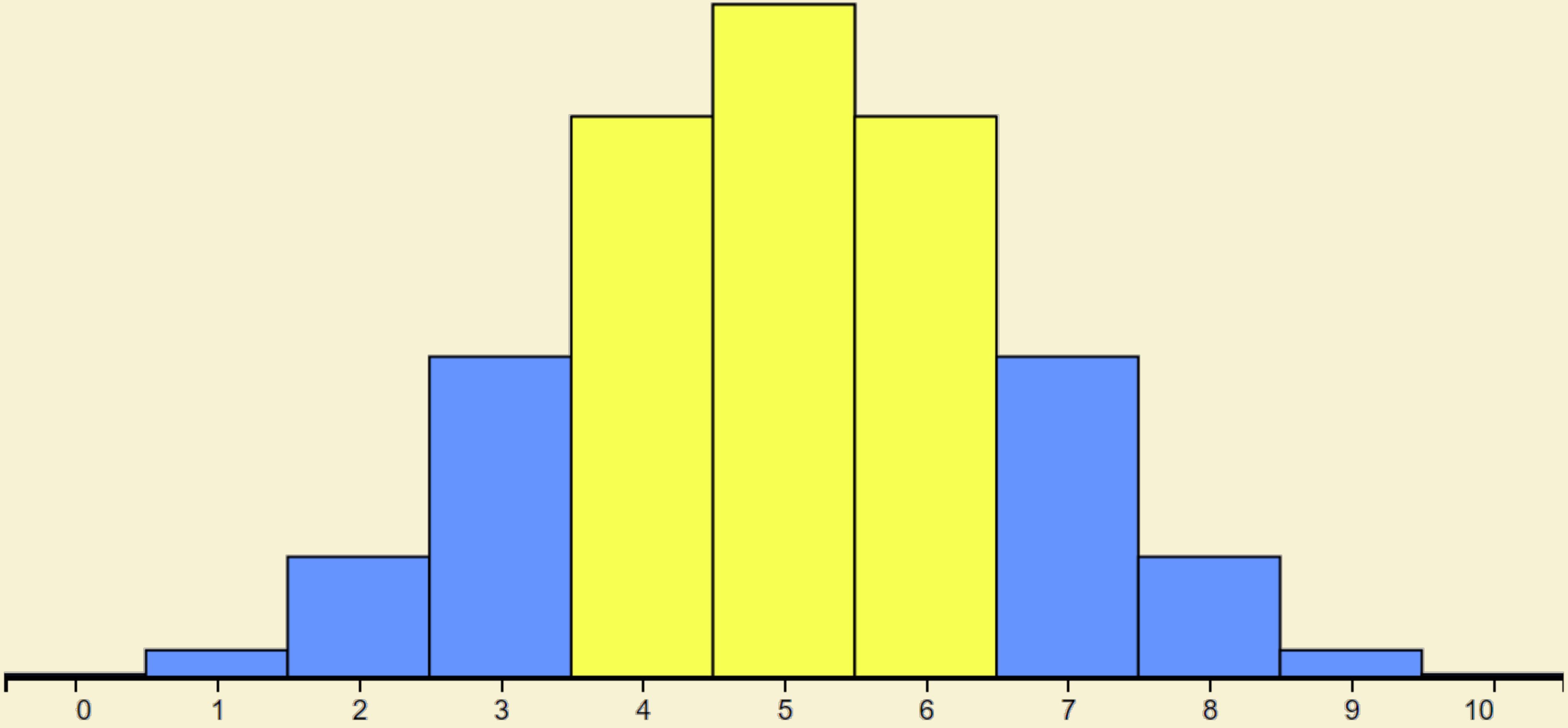
Sums of independent random variables



$n = 5$ $\Pr[X \notin (0.4, 0.6)n] = 37.5\%$

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

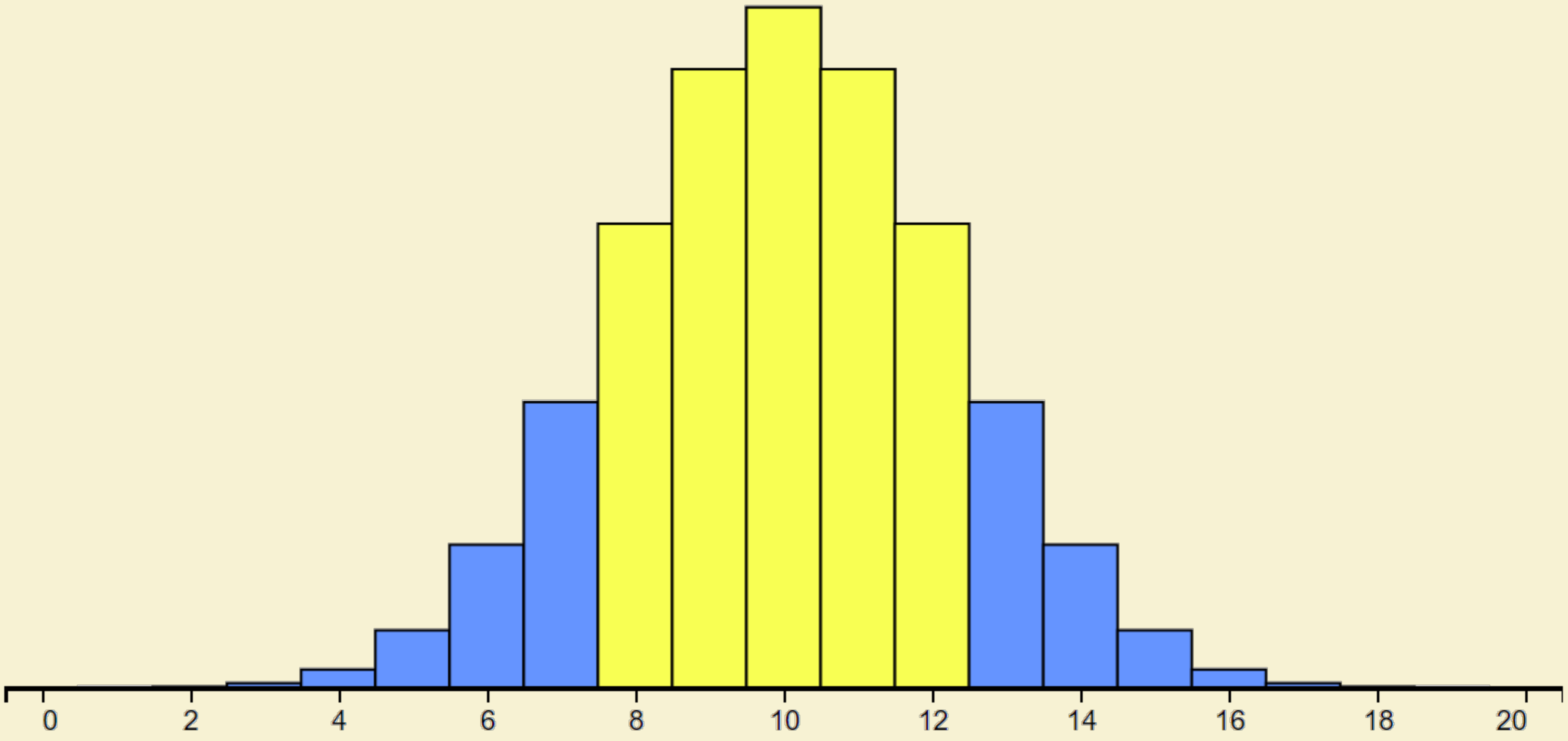
Sums of independent random variables



$n = 10$ $\Pr[X \notin (0.4, 0.6)n] = 34.4\%$

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

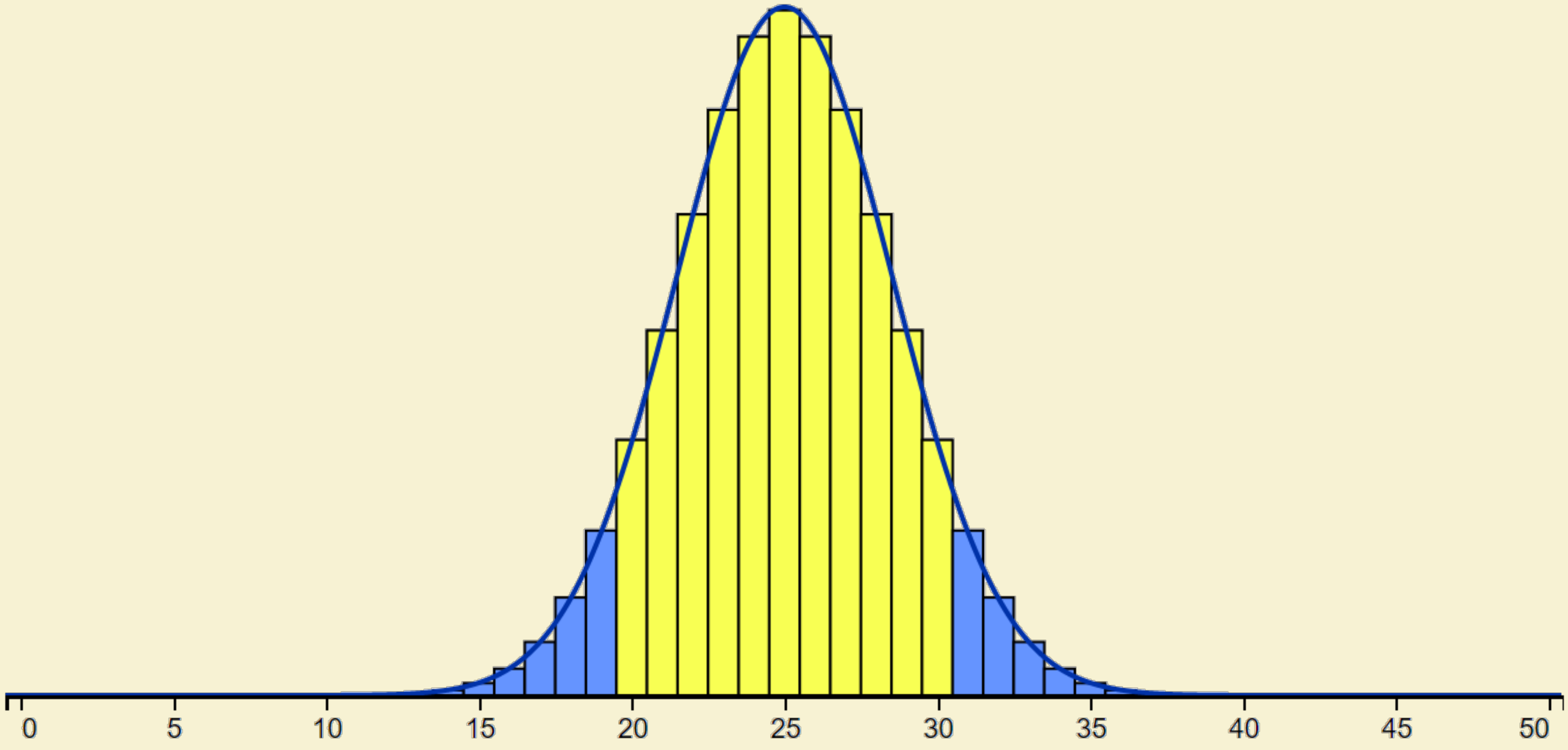
Sums of independent random variables



$n = 20$ $\Pr[X \notin (0.4,0.6)n] = 26.3\%$

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

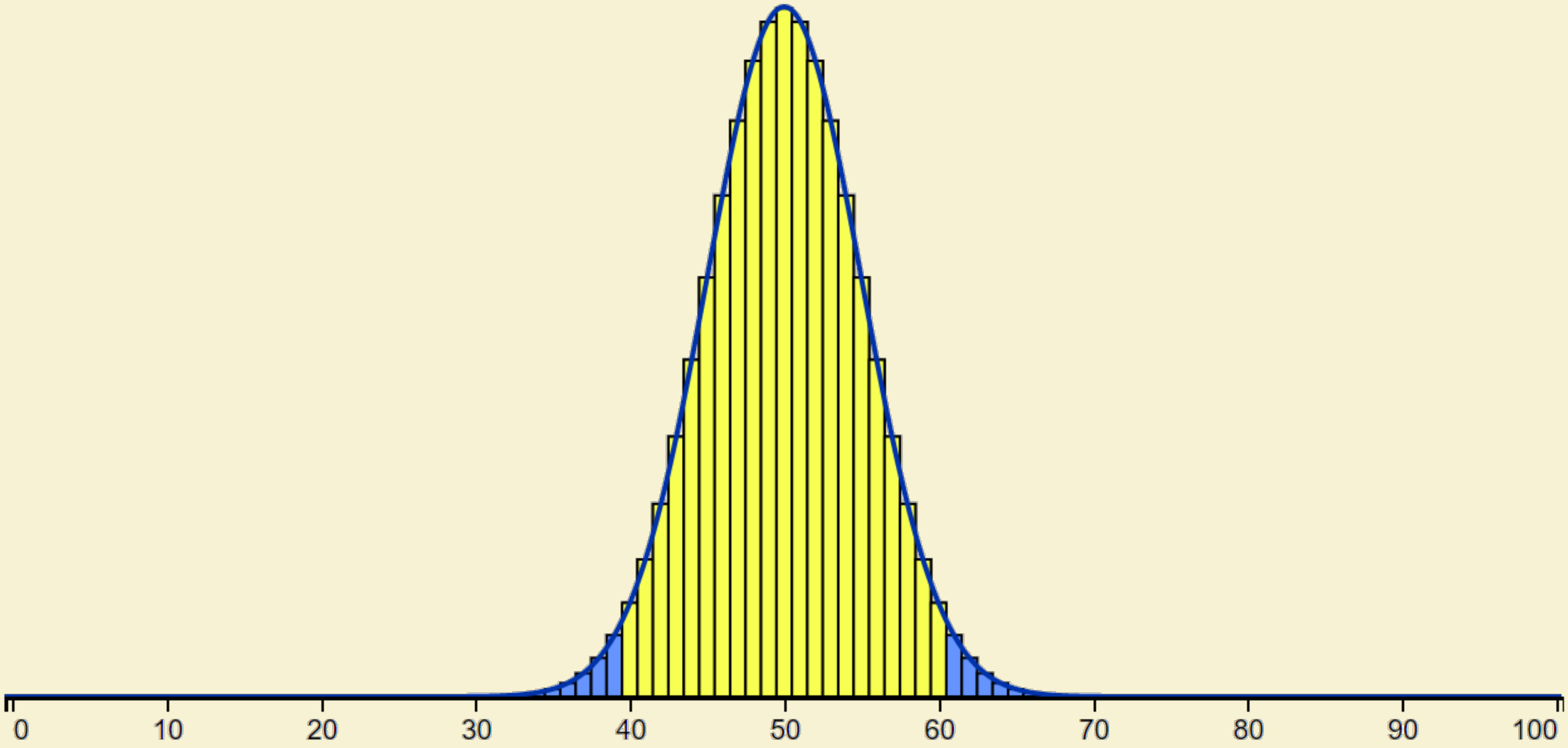
Sums of independent random variables



$n = 50$ $\Pr[X \notin (0.4, 0.6)n] =$ 11.9%

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

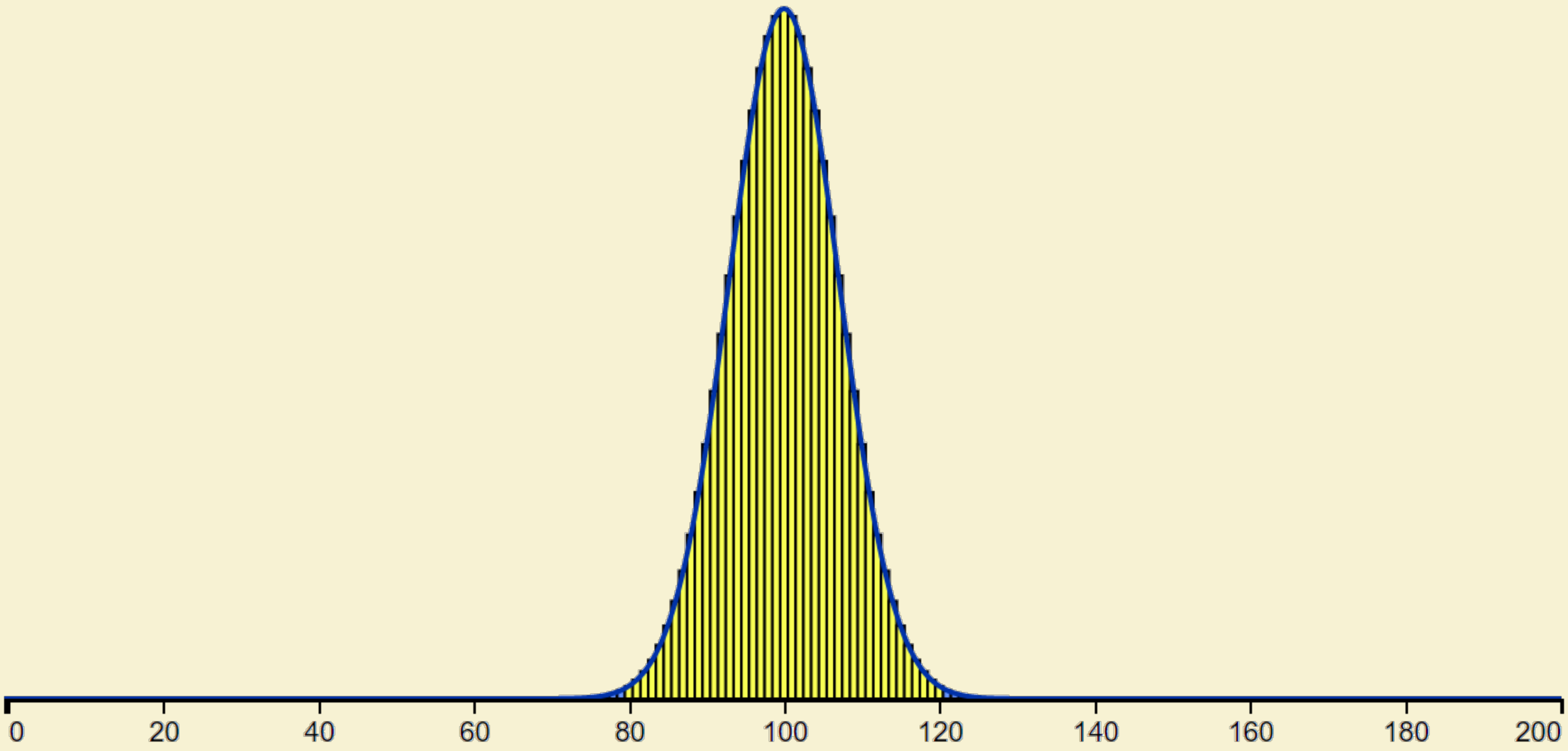
Sums of independent random variables



$n = 100$ $\Pr[X \notin (0.4, 0.6)n] =$ 3.6%

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

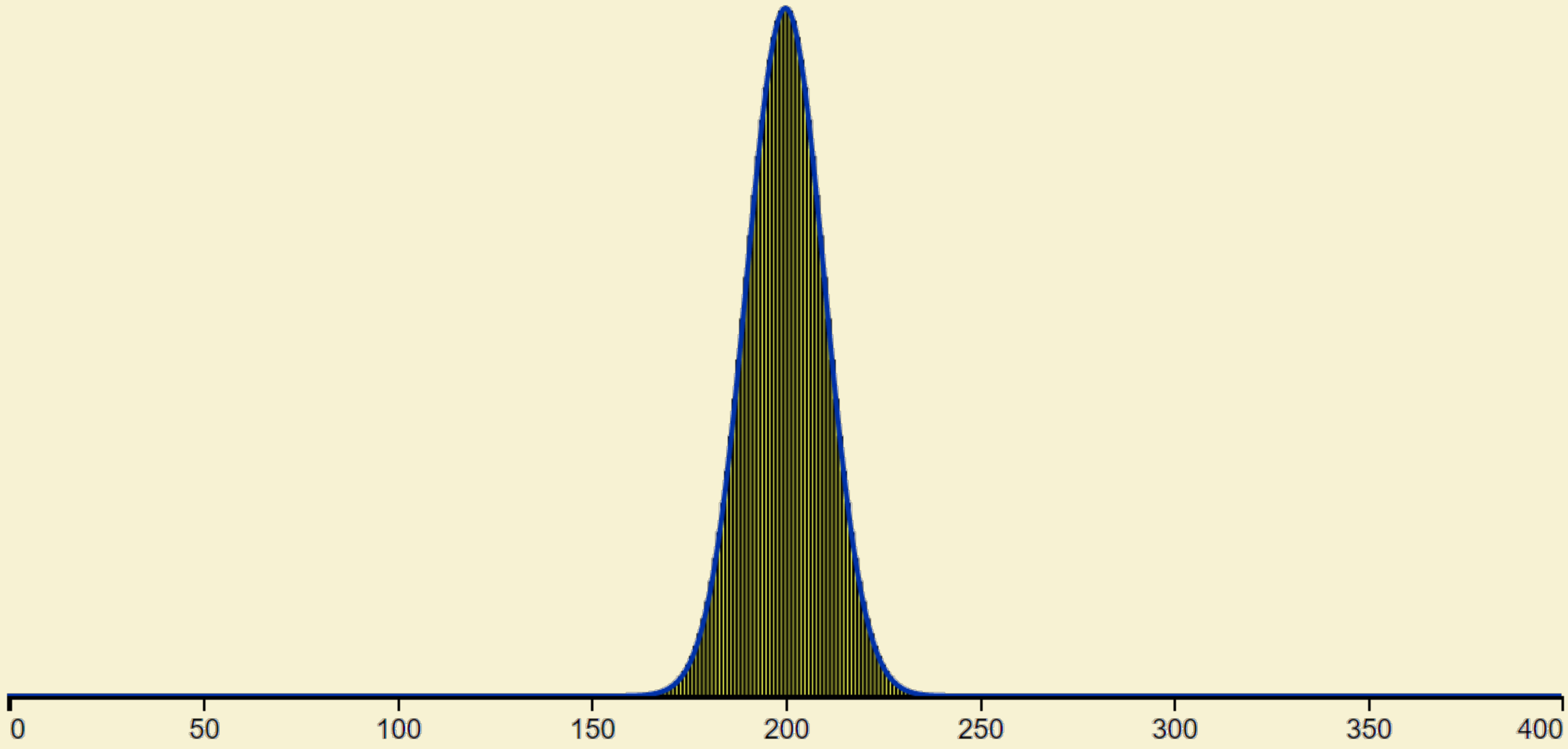
Sums of independent random variables



$n = 200$ $\Pr[X \notin (0.4,0.6)n] =$ 0.4%

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

Sums of independent random variables

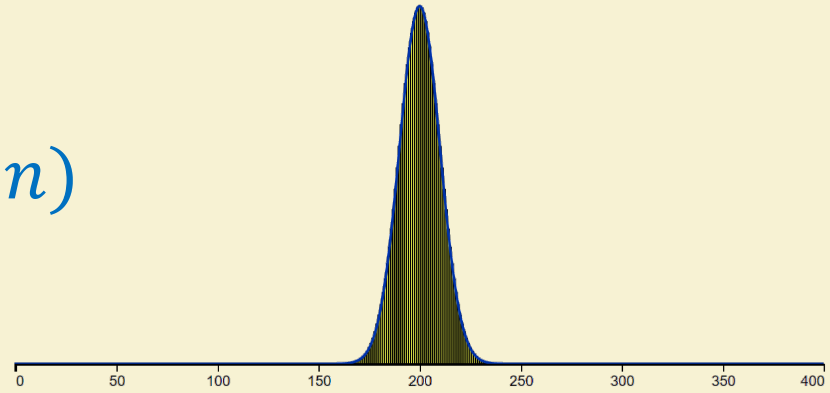


$n = 400$ $\Pr[X \notin (0.4, 0.6)n] = 0.01\%$

https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm

Concentration bounds

X sum of n independent random variables – ≈"Normal/Gaussian/bell curve":



$$\Pr\left[\, X \notin (0.99, 1.01)\mathbb{E}[X]\right] < \exp(-\delta \cdot n)$$

Otherwise: weaker bounds

In concentrated r.v.'s, expectation, median, m

Independent & identically distributed

Chernoff Bound: Let $X_0, \dots, X_{n-1}$ i.i.d. r.v.'s with $X_i \in [0,1]$.
Then if $X = X_0 + \cdots + X_{n-1}$ and $p = \mathbb{E}[X_i]$ for every $\epsilon > 0$,

$0 \le p \le 1$

$$\Pr[\, |X - np| > \epsilon n] < 2 \cdot \exp(-2\epsilon^2 \cdot n)$$

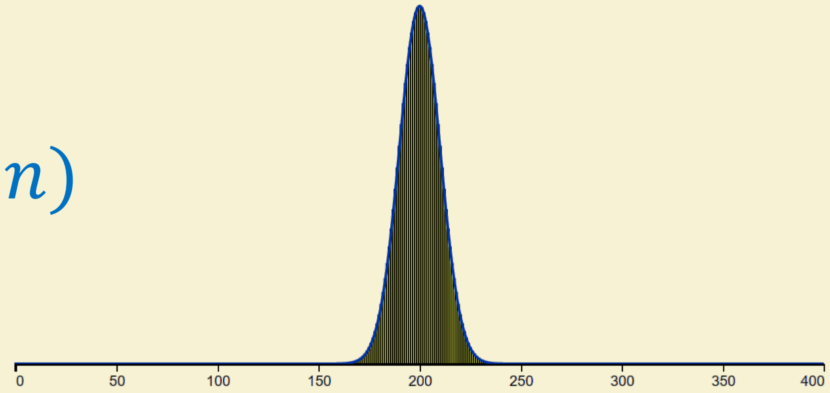
NOT($np - \epsilon n < X < np + \epsilon n$)

$\epsilon = .1$

Concentration bounds

X sum of n independent random variables – ≈“Normal/Gaussian/bell curve”:

$$\Pr\big[\, X \notin (0.99,1.01)\mathbb{E}[X]\big] < \exp(-\delta \cdot n)$$



Otherwise: weaker bounds

In concentrated r.v.'s, expectation, median, m

Independent & identically distributed

Chernoff Bound: Let $X_0, \ldots, X_{n-1}$ i.i.d. r.v.'s with $X_i \in [0,1]$.
Then if $X = X_0 + \cdots + X_{n-1}$ and $p = \mathbb{E}[X]$ for every $\epsilon > 0$,

$$\Pr[\, |X - np| > \epsilon n] < 2 \cdot \exp(-2\epsilon^2 \cdot n)$$

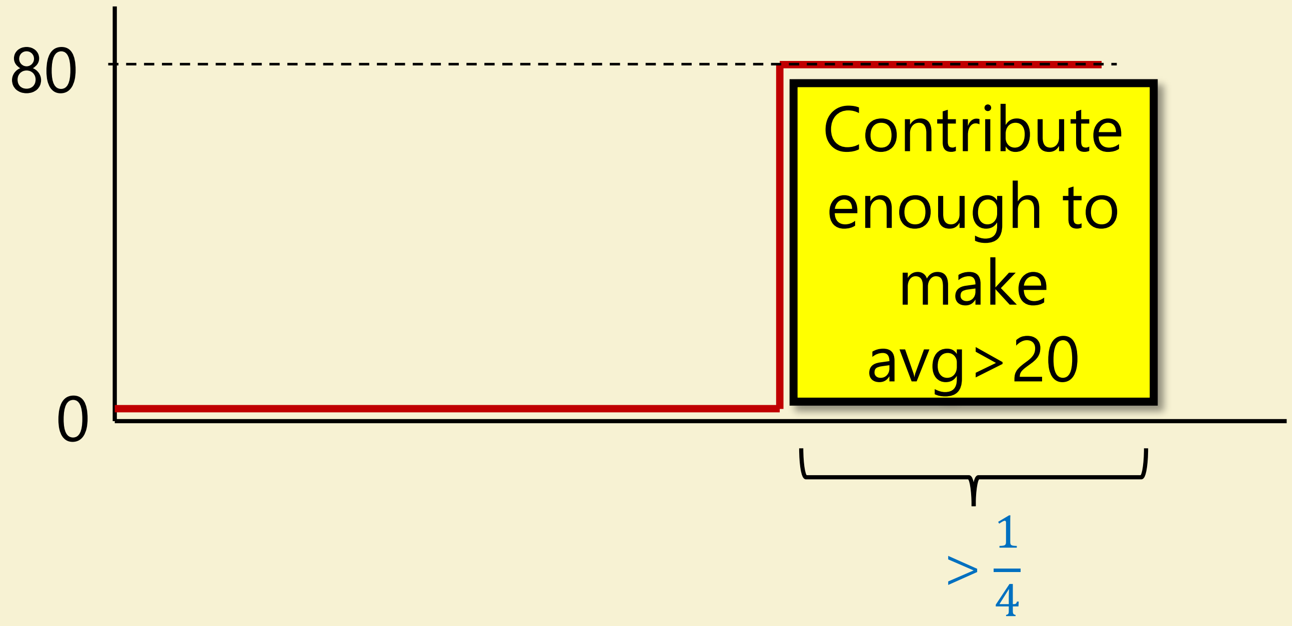
$$< \exp(-\epsilon^2 \cdot n) \left(\text{if } n > \frac{1}{\epsilon^2}\right)$$

Simplest Bound: Markov's Inequality

Q: Suppose that the average age in a neighborhood is 20. Prove that at most 1/4 of the residents are 80 or older.

A: Suppose otherwise:

$$\text{Avg} > \frac{1}{4} \cdot 80 = 20$$



Thm (Markov's Inequality): Let X be non-negative r.v. and $\mu = \mathbb{E}[X]$. Then for every $k > 1$, $\Pr[X \geq k\mu] \leq 1/k$

Proof: Same as question above

Variance & Chebychev

If X is r.v. with $\mu = \mathbb{E}[X]$ then $Var(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$$\sigma(X) = \sqrt{Var(X)}$$

Chebychev's Inequality: For every r.v. X

$$\Pr[\ X\ \geq k\ deviations\ from\ \mu] \leq \frac{1}{k^2}$$

(Proof: Markov on $Y = (X-\mu)^2$)

Compare with $X = \sum X_i$ i.i.d or other r.v.'s well approx. by Normal where

$$\Pr[\ X\ \geq k\ deviations\ from\ \mu] \approx \exp(-k^2)$$

Next Lectures

- Randomized Algorithms
  - Some examples
- Randomized Complexity Classes
  - $\mathrm{BPTIME}\big(T(n)\big)$, BPP
  - Properties of randomized computation (Reducing error …)