# COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Prof. Cameron Musco

University of Massachusetts Amherst. Fall 2020.

Lecture 1

# MOTIVATION FOR THIS CLASS

People are increasingly interested in analyzing and learning from massive datasets.

**People are increasingly interested in analyzing and learning from massive datasets.**

- Twitter receives 6,000 tweets per second, 500 million/day. Google receives 60,000 searches per second, 5.6 billion/day.
  - How do they process them to target advertisements? To predict trends? To improve their products?

People are increasingly interested in analyzing and learning from massive datasets.

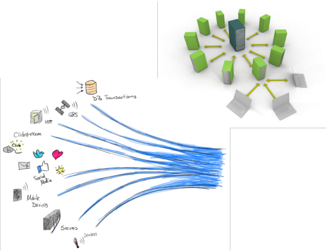
- Twitter receives 6,000 tweets per second, 500 million/day. Google receives 60,000 searches per second, 5.6 billion/day.
  - How do they process them to target advertisements? To predict trends? To improve their products?
- The Large Synoptic Survey Telescope will take high definition photographs of the sky, producing 15 terabytes of data/night.
  - How do they denoise and compress the images? How do they detect anomalies such as changing brightness or position of objects to alert researchers?

- Traditionally, algorithm design focuses on fast computation when data is stored in an efficiently accessible centralized manner (e.g., in RAM on a single machine).

# A NEW PARADIGM FOR ALGORITHM DESIGN

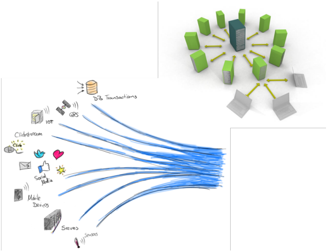
- Traditionally, algorithm design focuses on fast computation when data is stored in an efficiently accessible centralized manner (e.g., in RAM on a single machine).
- Massive data sets require storage in a distributed manner or processing in a continuous stream.

vs.



- Traditionally, algorithm design focuses on fast computation when data is stored in an efficiently accessible centralized manner (e.g., in RAM on a single machine).
- Massive data sets require storage in a distributed manner or processing in a continuous stream.

vs.




- Even ‘simple’ problems become very difficult in this setting.

For example:

For example:

- How can Twitter rapidly detect if an incoming Tweet is an exact duplicate of another Tweet made in the last year? Given that no machine can store all Tweets made in a year.

For example:

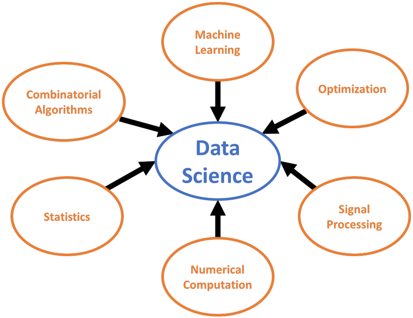
- How can Twitter rapidly detect if an incoming Tweet is an exact duplicate of another Tweet made in the last year? Given that no machine can store all Tweets made in a year.
- How can Google estimate the number of unique search queries that are made in a given week? Given that no machine can store the full list of queries.

For example:

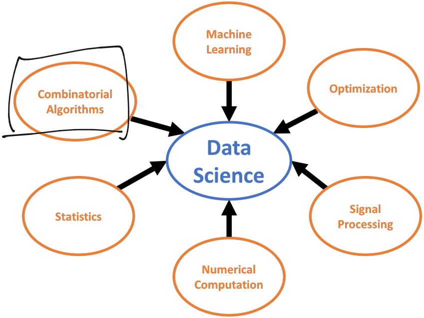
- How can Twitter rapidly detect if an incoming Tweet is an exact duplicate of another Tweet made in the last year? Given that no machine can store all Tweets made in a year.
- How can Google estimate the number of unique search queries that are made in a given week? Given that no machine can store the full list of queries.
- When you use Shazam to identify a song from a recording, how does it provide an answer in < 10 seconds, without scanning over all $\sim$ 8 million audio files in its database.

**A Second Motivation:** Data Science is highly interdisciplinary.

**A Second Motivation:** Data Science is highly interdisciplinary.



**A Second Motivation:** Data Science is highly interdisciplinary.




- Many techniques that aren't covered in the traditional CS algorithms curriculum.
- Emphasis on building comfort with mathematical tools that underly data science and machine learning.

# WHAT WE'LL COVER

Section 1: Randomized Methods & Sketching

Section 1: Randomized Methods & Sketching

How can we efficiently compress large data sets in a way that lets us answer important algorithmic questions rapidly?

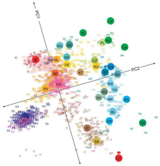
## Section 1: Randomized Methods & Sketching

How can we efficiently compress large data sets in a way that lets us answer important algorithmic questions rapidly?

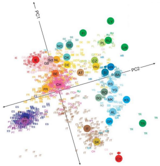
- Probability tools and concentration inequalities.
- Randomized hashing for efficient lookup, load balancing, and estimation. Bloom filters.
- Locality sensitive hashing and nearest neighbor search.
- Streaming algorithms: identifying frequent items in a data stream, counting distinct items, etc.
- Random compression of high-dimensional vectors: the Johnson-Lindenstrauss lemma, applications, and connections to the weirdness of high-dimensional geometry.

# WHAT WE'LL COVER

Section 2: Spectral Methods

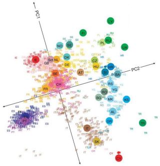


Section 2: Spectral Methods



How do we identify the most important features of a dataset using linear algebraic techniques?

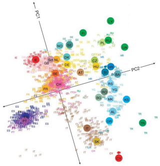
## Section 2: Spectral Methods



How do we identify the most important features of a dataset using linear algebraic techniques?

- Principal component analysis, low-rank approximation, dimensionality reduction.
- The singular value decomposition (SVD) and its applications to PCA, low-rank approximation, LSI, MDS, …
- Spectral graph theory. Spectral clustering, community detection, network visualization.
- Computing the SVD on large datasets via iterative methods.

## Section 2: Spectral Methods



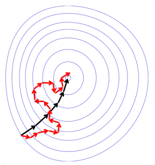
How do we identify the most important features of a dataset using linear algebraic techniques?

*If you open up the codes that are underneath [most data science applications] this is all linear algebra on arrays.*

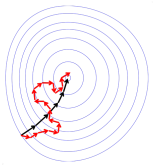
– Michael Stonebraker

# WHAT WE'LL COVER

Section 3: Optimization

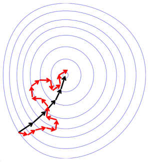


Section 3: Optimization



Fundamental continuous optimization approaches that drive methods in machine learning and statistics.

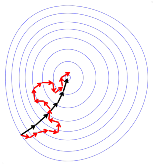
## Section 3: Optimization



Fundamental continuous optimization approaches that drive methods in machine learning and statistics.

- Gradient descent. Analysis for convex functions.
- Stochastic and online gradient descent.
- Focus on convergence analysis.

## Section 3: Optimization



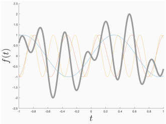
Fundamental continuous optimization approaches that drive methods in machine learning and statistics.

- Gradient descent. Analysis for convex functions.
- Stochastic and online gradient descent.
- Focus on convergence analysis.

A small taste of what you can find in COMPSCI 590OP or 690OP.

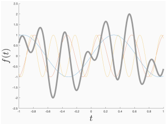
# WHAT WE'LL COVER

Section 4: Assorted Topics



Depending on pacing may have time for a few bonus classes.

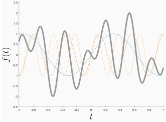
Section 4: Assorted Topics




Depending on pacing may have time for a few bonus classes.

- Compressed sensing, restricted isometry property, basis pursuit.
- Discrete Fourier transform, fast Fourier transform.
- Differential privacy, algorithmic fairness.

**Section 4: Assorted Topics**




Depending on pacing may have time for a few bonus classes.

- Compressed sensing, restricted isometry property, basis pursuit.
- Discrete Fourier transform, fast Fourier transform.
- Differential privacy, algorithmic fairness.

Some flexibility here. Let me know what you are interested in!

# IMPORTANT TOPICS WE WON'T COVER

# IMPORTANT TOPICS WE WON'T COVER

- Systems/Software Tools.

hadoop TensorFlow

APACHE Spark™ Flink APACHE GIRAPH

# IMPORTANT TOPICS WE WON'T COVER

- Systems/Software Tools.

  - COMPSCI 532: Systems for Data Science

- **Systems/Software Tools.**
  - COMPSCI 532: Systems for Data Science
- **Machine Learning/Data Analysis Methods and Models.**
  - E.g., regression methods, kernel methods, random forests, SVM, deep neural networks.

# IMPORTANT TOPICS WE WON'T COVER

- **Systems/Software Tools.**
  - COMPSCI 532: Systems for Data Science
- **Machine Learning/Data Analysis Methods and Models.**
  - E.g., regression methods, kernel methods, random forests, SVM, deep neural networks.
  - COMPSCI 589/689: Machine Learning

# STYLE OF THE COURSE

This is a theory course.

This is a theory course.

- Build general mathematical tools and algorithmic strategies that can be applied to a wide range of problems.

This is a theory course.

- Build general mathematical tools and algorithmic strategies that can be applied to a wide range of problems.
- Assignments will emphasize algorithm design, correctness proofs, and asymptotic analysis (minimal required coding).

**This is a theory course.**

- Build general mathematical tools and algorithmic strategies that can be applied to a wide range of problems.
- Assignments will emphasize algorithm design, correctness proofs, and asymptotic analysis (minimal required coding).
- The homework is designed to make you think beyond what is taught in class. You will get stuck, and not see the solutions right away. This is the best (only?) way to build mathematical and algorithm design skills.

# STYLE OF THE COURSE

This is a theory course.

- Build general mathematical tools and algorithmic strategies that can be applied to a wide range of problems.
- Assignments will emphasize algorithm design, correctness proofs, and asymptotic analysis (minimal required coding).
- The homework is designed to make you think beyond what is taught in class. You will get stuck, and not see the solutions right away. This is the best (only?) way to build mathematical and algorithm design skills.
- A strong algorithms and mathematical background (particularly in linear algebra and probability) **are required**.
- UMass prereqs: COMPSCI 240 and COMPSCI 311.

This is a theory course.

- Build general mathematical tools and algorithmic strategies that can be applied to a wide range of problems.
- Assignments will emphasize algorithm design, correctness proofs, and asymptotic analysis (minimal required coding).
- The homework is designed to make you think beyond what is taught in class. You will get stuck, and not see the solutions right away. This is the best (only?) way to build mathematical and algorithm design skills.
- A strong algorithms and mathematical background (particularly in linear algebra and probability) **are required**.
- UMass prereqs: COMPSCI 240 and COMPSCI 311.

**For example:** Baye's rule in conditional probability. What it means for a vector $x$ to be an eigenvector of a matrix $A$, orthogonal projection, greedy algorithms, divide-and-conquer algorithms.

See course webpage for logistics, policies, lecture notes, assignments, etc.:

http://people.cs.umass.edu/~cmusco/CS514F20/

See Moodle page for this link if you lose it and the lecture password.

**Professor:** Cameron Musco

- Email: cmusco@cs.umass.edu
- Office Hours: Tuesdays, 8am-9am and 2:15pm-3:15pm. See website for Zoom link (different than lecture)
- I encourage you to come as regularly as possible to ask questions/work together on practice problems.

**TAs:**

- Pratheba Selvaraju
- Shiv Shankar

See website for office hours, contact info, and Zoom links.

# PIAZZA AND PARTICIPATION

We will use Piazza for class discussion and questions.

- See website for link to sign up.

We will use Piazza for class discussion and questions.

- See website for link to sign up.

You may earn up to 5% extra credit for participation.

- Asking good clarifying questions and answering questions during the live lecture or on Piazza.
- Actively participating in office hours.
- Answering other students' or instructor questions on Piazza.
- Posting helpful/interesting links on Piazza, e.g., resources that cover class material, research articles related to the topics covered in class, etc.

- If you can, please turn on your video, but keep your mic muted when not asking a question.
- If you want to ask a question, either raise your hand, just unmute and interrupt me, and ask in chat.
- I encourage you to answer each others questions/discuss in chat during lecture.
- I will sometimes create random breakout rooms for you to work together/discuss problems. So if you are signed in to lecture please be ready to participate in these.

## TEXTBOOKS AND MATERIALS

We will use material from two textbooks (links to free online versions on the course webpage): *Foundations of Data Science* and *Mining of Massive Datasets*, but will follow neither closely.

- I will post optional readings a few days prior to each class.
- Lecture notes will be posted before each class, and annotated notes posted after class.

We will have 6 problem sets, which you may complete in **groups of up to 3 students**.

We will have 6 problem sets, which you may complete in **groups of up to 3 students**.

- We strongly encourage working in groups, as it will make completing the problem sets much easier/more educational.
- Collaboration with students outside your group is limited to discussion at a high level. You may not work through problems in detail or write up solutions together.
- See Piazza for a thread to help you organize groups.

We will have 6 problem sets, which you may complete in **groups of up to 3 students**.

- We strongly encourage working in groups, as it will make completing the problem sets much easier/more educational.
- Collaboration with students outside your group is limited to discussion at a high level. You may not work through problems in detail or write up solutions together.
- See Piazza for a thread to help you organize groups.

Problem set submissions will be via Gradescope.

- See website for a link to join. **Entry Code: 9DV6G5**
- Since your emails, names, and grades will be stored in Gradescope we need your consent to use. See Piazza for a poll to give consent. Please complete by **next Thursday 9/3**.

I will release a multiple choice quiz in Moodle each Thursday after lecture, due the next Monday at 8pm.

I will release a multiple choice quiz in Moodle each Thursday after lecture, due the next Monday at 8pm.

- Designed as a check-in that you are following the material, and to help me make adjustments as needed.
- Will take around 15-30 minutes per week, open notes.
- Will also include free response check-in questions to get your feedback on how the course is going, what material from the past week you find most confusing, interesting, etc.

Grade Breakdown:

- Problem Sets (6 total): 40%, weighted equally.
- Weekly Quizzes: 10%, weighted equally.
- Midterm (early October, take home): 25%.
- Final (early December, take home): 25%.
- Extra Credit: Up to 5% for participation, and lots more available on problem sets, for questions asked in class, etc.

Grade Breakdown:

- Problem Sets (6 total): 40%, weighted equally.
- Weekly Quizzes: 10%, weighted equally.
- Midterm (early October, take home): 25%.
- Final (early December, take home): 25%.
- Extra Credit: Up to 5% for participation, and lots more available on problem sets, for questions asked in class, etc.

Academic Honesty:

- A first violation cheating on a homework, quiz, or other assignment will result in a 0 on that assignment.
- A second violation, or cheating on an exam will result in failing the class.

# DISABILITY SERVICES AND ACCOMODATIONS

UMass Amherst is committed to making reasonable, effective, and appropriate accommodations to meet the needs to students with disabilities.

- If you have a documented disability **on file with Disability Services**, you may be eligible for reasonable accommodations in this course.
- If your disability requires an accommodation, please notify me by **next Thursday 9/3** so that we can make arrangements.

## DISABILITY SERVICES AND ACCOMODATIONS

UMass Amherst is committed to making reasonable, effective, and appropriate accommodations to meet the needs to students with disabilities.

- If you have a documented disability **on file with Disability Services**, you may be eligible for reasonable accommodations in this course.
- If your disability requires an accommodation, please notify me by **next Thursday 9/3** so that we can make arrangements.

I understand that people have different learning needs, home situations, etc. If something isn't working for you in the class, please reach out and let's try to work it out.

Questions?

Get to know your classmates:

Get to know your classmates:

- I’m going to split you into random breakout rooms.
- Come up with a list of as many things as you can that are common to all members of your group.
- They can’t be things that are common to say more than 50% of this class.
- Designate a scribe to write them down.
- The group with the largest list wins.

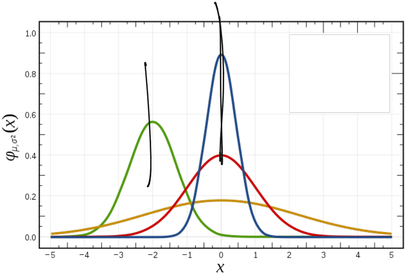
# Section 1: Randomized Methods & Sketching

# SOME PROBABILITY REVIEW

Consider a random X variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

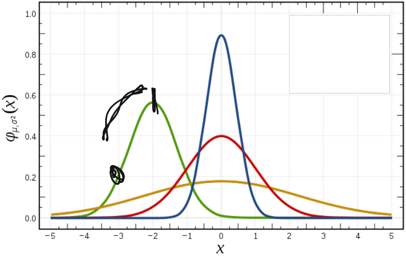
Consider a random X variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

- **Expectation:** $\mathbb{E}[X] = \sum_{s \in S} \Pr(X = s) \cdot s.$



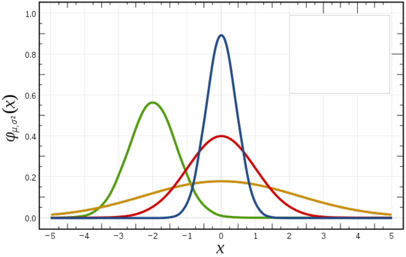
Consider a random $X$ variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

- **Expectation:** $\mathbb{E}[X] = \sum_{s \in S} \Pr(X = s) \cdot s.$
- **Variance:** $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$



Consider a random $X$ variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

- **Expectation:** $\mathbb{E}[X] = \sum_{s \in S} \Pr(X = s) \cdot s.$
- **Variance:** $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2].$




**Exercise:** Show that for any scalar $\alpha$, $\mathbb{E}[\alpha \cdot X] = \alpha \cdot \mathbb{E}[X]$ and $\text{Var}[\alpha \cdot X] = \alpha^2 \cdot \text{Var}[X]$.

Consider two random events $A$ and $B$.

$A \cap B$: event that both events $A$ *and* $B$ happen.

Consider two random events $A$ and $B$.

- **Conditional Probability:**

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

A ∪ B

$A \cap B$: event that both events $A$ *and* $B$ happen.

Consider two random events $A$ and $B$.

- **Conditional Probability:**

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

- **Independence:** $A$ and $B$ are independent if:

$$\Pr(A|B) = \Pr(A).$$

$A \cap B$: event that both events *A and B* happen.

Consider two random events $A$ and $B$.

- **Conditional Probability:**

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

- **Independence:** $A$ and $B$ are independent if:

$$\Pr(A|B) = \Pr(A).$$

Using the definition of conditional probability, independence means:

$$\frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A) \implies \Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

$A \cap B$: event that both events $A$ *and* $B$ happen.

For Example: What is the probability that for two independent dice rolls the first is a 6 and the second is odd?

**For Example:** What is the probability that for two independent dice rolls the first is a 6 and the second is odd?

$$\Pr(D_1 = 6 \cap D_2 \in \{1,3,5\}) = \Pr(D_1 = 6) \cdot \Pr(D_2 \in \{1,3,5\})$$

$$= \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

**For Example:** What is the probability that for two independent dice rolls the first is a 6 and the second is odd?

$$\Pr(D_1 = 6 \cap D_2 \in \{1, 3, 5\}) = \Pr(D_1 = 6) \cdot \Pr(D_2 \in \{1, 3, 5\})$$

**Independent Random Variables:** Two random variables X, Y are independent if for all $s, t$, $\mathrm{X} = s$ and $\mathrm{Y} = t$ are independent events. In other words:

$$\Pr(\mathrm{X} = s \cap \mathrm{Y} = t) = \Pr(\mathrm{X} = s) \cdot \Pr(\mathrm{Y} = t).$$

**Think-Pair-Share:** When are the expectation and variance linear?

I.e., under what conditions on $X$ and $Y$ do we have:

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

and

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y].$$

$X, Y$: any two random variables.

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

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

Proof:

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

**Proof:**

$$\mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t)$$

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

Proof:

$$\mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t)$$

$$= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot t$$

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables X and Y.

Proof:

$$\mathbb{E}[X+Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t)$$

$$= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot t$$

$$= \sum_{s \in S} s \cdot \sum_{t \in T} \Pr(X = s \cap Y = t) + \sum_{t \in T} t \cdot \sum_{s \in S} \Pr(X = s \cap Y = t)$$

P(X=S)

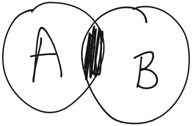
$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

Proof:

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t) \\
&= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot t \\
&= \sum_{s \in S} s \cdot \sum_{t \in T} \Pr(X = s \cap Y = t) + \sum_{t \in T} t \cdot \sum_{s \in S} \Pr(X = s \cap Y = t)
\end{aligned}$$

# LINEARITY OF EXPECTATION

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.




**Proof:**

$$\mathbb{E}[X+Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s+t)$$

$$= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot t$$

$$= \sum_{s \in S} s \cdot \sum_{t \in T} \Pr(X = s \cap Y = t) + \sum_{t \in T} t \cdot \sum_{s \in S} \Pr(X = s \cap Y = t)$$

$$= \sum_{s \in S} s \cdot \Pr(X = s) + \sum_{t \in T} t \cdot \Pr(Y = t)$$

(law of total probability)

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ for any random variables $X$ and $Y$.

**Proof:**

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{s\in S}\sum_{t\in T} \Pr(X = s \cap Y = t)\cdot(s+t) \\
&= \sum_{s\in S}\sum_{t\in T} \Pr(X = s \cap Y = t)\cdot s + \sum_{s\in S}\sum_{t\in T} \Pr(X = s \cap Y = t)\cdot t \\
&= \sum_{s\in S} s \cdot \sum_{t\in T} \Pr(X = s \cap Y = t) + \sum_{t\in T} t \cdot \sum_{s\in S} \Pr(X = s \cap Y = t) \\
&= \sum_{s\in S} s \cdot \Pr(X = s) + \sum_{t\in T} t \cdot \Pr(Y = t) \\
&\qquad\qquad \text{(law of total probability)} \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ when X and Y are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\text{Var}[X+Y] = \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2$$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\begin{aligned}\text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2\end{aligned}$$

(linearity of expectation)

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\begin{aligned} \text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &\qquad \text{(linearity of expectation)} \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \end{aligned}$$

$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\begin{aligned}
\text{Var}[X + Y] &= \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&\qquad\qquad\qquad \text{(linearity of expectation)} \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2
\end{aligned}$$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\begin{aligned} \text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &\qquad \text{(linearity of expectation)} \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \\ &= \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 \end{aligned}$$

$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$ when $X$ and $Y$ are independent.

**Claim 1: (exercise)** $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (via linearity of expectation)

**Claim 2: (exercise)** $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when $X, Y$ are independent.

**Together give:**

$$\begin{aligned}
\text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(linearity of expectation)} \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 \\
&= \text{Var}[X] + \text{Var}[Y].
\end{aligned}$$