COMPSCI 514: ALGORITHMS FOR DATA SCIENCE

Andrew McGregor

Lecture 1

MOTIVATION FOR THIS CLASS

MOTIVATION FOR THIS CLASS

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

MOTIVATION FOR THIS CLASS

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?

MOTIVATION FOR THIS CLASS

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?

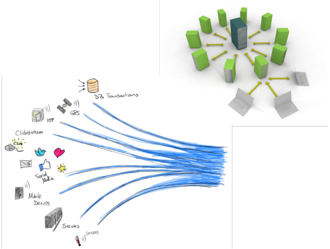
A NEW PARADIGM FOR ALGORITHM DESIGN

- 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).

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.

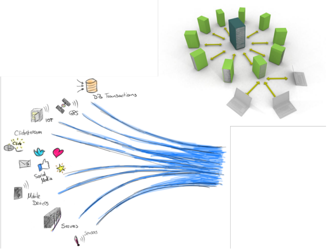
vs.



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.



- 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.

A NEW PARADIGM FOR ALGORITHM DESIGN

For example:

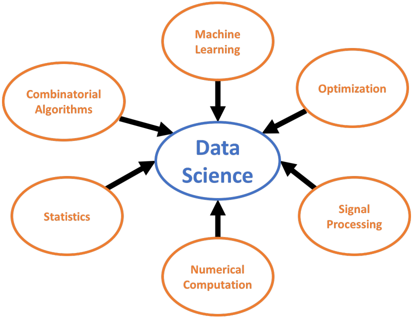
- How can Twitter rapidly detect if an incoming Tweet is an exact duplicate of another Tweet made in the last year? Given that no machine can store all Tweets made in a year.
- How can Google estimate the number of unique search queries that are made in a given week? Given that no machine can store the 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 ~ 8 million audio files in its database.

MOTIVATION FOR THIS CLASS

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

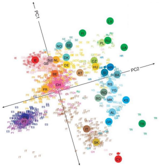
How can we efficiently compress large data sets in a way that lets us answer important algorithmic questions rapidly?

Section 1: Randomized Methods & Sketching

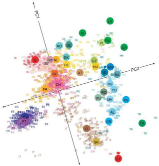
How can we efficiently compress large data sets in a way that lets us answer important algorithmic questions rapidly?

- 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.

Section 2: Spectral Methods

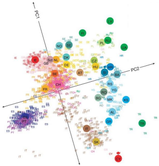


Section 2: Spectral Methods



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

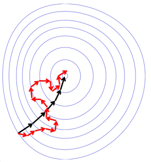
Section 2: Spectral Methods



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

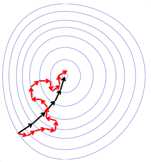
- Principal component analysis, low-rank approximation, dimensionality reduction.
- Singular value decomposition (SVD) and its applications to PCA, low-rank approximation, LSA, MDS, . . .
- Spectral graph theory. Spectral clustering, community detection, network visualization.
- Computing the SVD on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

Section 3: Optimization



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

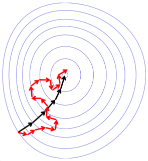
Section 3: Optimization



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 651.

IMPORTANT TOPICS WE WON'T COVER

IMPORTANT TOPICS WE WON'T COVER

- **Systems/Software Tools.**

IMPORTANT TOPICS WE WON'T COVER

- **Systems/Software Tools.**

  - COMPSCI 532: Systems for Data Science

IMPORTANT TOPICS WE WON'T COVER

- **Systems/Software Tools.**
  - COMPSCI 532: Systems for Data Science
- **Machine Learning/Data Analysis Methods and Models.**

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.

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, perhaps the only theory course you'll take at UMass if you're a Masters student.

STYLE OF THE COURSE

This is a theory course, perhaps the only theory course you'll take at UMass if you're a Masters student.

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

Style of the Course

This is a theory course, perhaps the only theory course you'll take at UMass if you're a Masters student.

- 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. There'll may be some coding but not much.

STYLE OF THE COURSE

This is a theory course, perhaps the only theory course you'll take at UMass if you're a Masters student.

- 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. There'll may be some coding but not much.
- 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, perhaps the only theory course you'll take at UMass if you're a Masters student.

- 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. There'll may be some coding but not much.
- 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 math background (particularly in linear algebra and probability) **are required**. See Moodle for revision notes.
- UMass prereqs: COMPSCI 240 and COMPSCI 311.

This is a theory course, perhaps the only theory course you'll take at UMass if you're a Masters student.

- 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. There'll may be some coding but not much.
- 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 math background (particularly in linear algebra and probability) **are required**. See Moodle for revision notes.
- UMass prereqs: COMPSCI 240 and COMPSCI 311.

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

COURSE LOGISTICS

See course webpage for lecture slides and related readings:

https://people.cs.umass.edu/~mcgregor/CS514S23/

See Moodle page for this link if you lose it.

PERSONNEL

Professor: Andrew McGregor

TAs: Shib Dasgupta, Weronkia Nguyen, and Chenghao Lyu.

Together we've offer seven office hours, four in-person and three over Zoom. See Moodle page for locations, links, and times.

PIAZZA AND PARTICIPATION

We will use Piazza for class discussion and questions.

- See Moodle for link to sign up.

PIAZZA AND PARTICIPATION

We will use Piazza for class discussion and questions.

- See Moodle for link to sign up.

You're helping yourself and others if you:

- Ask good clarifying questions and answering questions during lectures.
- Answer other students' or instructor questions on Piazza.
- Post helpful/interesting links on Piazza, e.g., resources covering class material, research articles related to class topics.

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 sometimes post optional readings a few days prior to each class.
- Draft lecture notes will be posted before each class and potentially updated afterwards if necessary.

HOMEWORK

We will have 4-5 problem sets, which you may complete in **groups of up to 3 students**.

We will have 4-5 problem sets, which you may complete in **groups of up to 3 students**.

- You may not discuss homework with people outside your group (except the instructor and TAs) until the solutions are released.
- Use are not allowed to consult previous solutions from the class or search for solutions online.
- Work together on each question rather than dividing the questions between group members.
- Use, e.g., Piazza, to help organize groups.

HOMEWORK

We will have 4-5 problem sets, which you may complete in **groups of up to 3 students**.

- You may not discuss homework with people outside your group (except the instructor and TAs) until the solutions are released.
- Use are not allowed to consult previous solutions from the class or search for solutions online.
- Work together on each question rather than dividing the questions between group members.
- Use, e.g., Piazza, to help organize groups.

Problem set submissions will be via Gradescope.

- See Moodle for a link to join.

WEEKLY QUIZZES

Will release a Moodle quiz most Fridays. It's due Monday at 8pm.

WEEKLY QUIZZES

Will release a Moodle quiz most Fridays. It's due Monday at 8pm.

- Designed as a check-in that you are following the material, and to help me make adjustments as needed.
- Should take under an hour per week. Open notes unless specified otherwise. Most questions easy but some more challenging ones.

GRADING

Grade Breakdown:

- Problem Sets: 25%. (One-time “lifeline extension” of 48 hours.)
- Weekly Quizzes: 20%. (No extensions but we’ll drop lowest quiz.)
- Midterm: 25%. Thursday October 20th
- Cumulative Final: 25%. (During Final’s Week.)
- Piazza Participation: 5%.

GRADING

Grade Breakdown:

- Problem Sets: 25%. (One-time "lifeline extension" of 48 hours.)
- Weekly Quizzes: 20%. (No extensions but we'll drop lowest quiz.)
- Midterm: 25%. Thursday October 20th
- Cumulative Final: 25%. (During Final's Week.)
- Piazza Participation: 5%.

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 ACCOMMODATIONS

UMass 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 **Friday 2/17** so that we can make arrangements.

DISABILITY SERVICES AND ACCOMMODATIONS

UMass 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 **Friday 2/17** so that we can make arrangements.

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

Questions?

Section 1: Randomized Methods & Sketching

SOME PROBABILITY REVIEW

SOME PROBABILITY REVIEW

Consider a random $\mathbf{X}$ 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 $\mathbf{X}$ 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:** $\mathbb{E}[\mathbf{X}] = \sum_{s \in S} \Pr(\mathbf{X} = s) \cdot s$.

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

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

E.g., if **X** takes the values $1, 2, 4$ with probabilities $1/3, 1/2, 1/6$ then

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

E.g., if $\mathbf{X}$ takes the values $1, 2, 4$ with probabilities $1/3, 1/2, 1/6$ then

$$\mathbb{E}[\mathbf{X}] = 1/3 \cdot 1 + 1/2 \cdot 2 + 1/6 \cdot 4 = 2$$

$$\text{Var}[\mathbf{X}] = 1/3 \cdot (1 - 2)^2 + 1/2 \cdot (2 - 2)^2 + 1/6 \cdot (4 - 2)^2 = 1$$

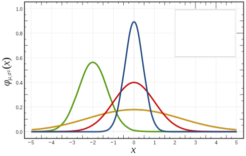
Consider a random $\mathbf{X}$ variable taking values in some finite set $S \subset \mathbb{R}$.
E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

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

E.g., if $\mathbf{X}$ takes the values $1, 2, 4$ with probabilities $1/3, 1/2, 1/6$ then

$$\mathbb{E}[\mathbf{X}] = 1/3 \cdot 1 + 1/2 \cdot 2 + 1/6 \cdot 4 = 2$$

$$\text{Var}[\mathbf{X}] = 1/3 \cdot (1 - 2)^2 + 1/2 \cdot (2 - 2)^2 + 1/6 \cdot (4 - 2)^2 = 1$$



SOME PROBABILITY REVIEW

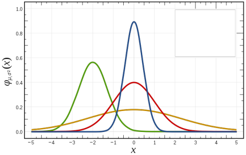
Consider a random $\mathbf{X}$ variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$.

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

E.g., if $\mathbf{X}$ takes the values $1, 2, 4$ with probabilities $1/3, 1/2, 1/6$ then

$$\mathbb{E}[\mathbf{X}] = 1/3 \cdot 1 + 1/2 \cdot 2 + 1/6 \cdot 4 = 2$$

$$\text{Var}[\mathbf{X}] = 1/3 \cdot (1 - 2)^2 + 1/2 \cdot (2 - 2)^2 + 1/6 \cdot (4 - 2)^2 = 1$$



For any scalar α, $\mathbb{E}[\alpha \cdot \mathbf{X}] = \alpha \cdot \mathbb{E}[\mathbf{X}]$ and $\text{Var}[\alpha \cdot \mathbf{X}] = \alpha^2 \cdot \text{Var}[\mathbf{X}]$.

INDEPENDENCE

Consider two random events A and B.

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

INDEPENDENCE

Consider two random events A and B.

- **Conditional Probability:**

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

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

INDEPENDENCE

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.

INDEPENDENCE

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.

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

$$\begin{aligned}\Pr(D_1 \in \{1,3,5\} \cap D_2 \in \{1,3,5\}) &= \Pr(D_1 \in \{1,3,5\}) \cdot \Pr(D_2 \in \{1,3,5\}) \\ &= 1/4\end{aligned}$$

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

$$\begin{aligned}\Pr(D_1 \in \{1,3,5\} \cap D_2 \in \{1,3,5\}) &= \Pr(D_1 \in \{1,3,5\}) \cdot \Pr(D_2 \in \{1,3,5\}) \\ &= 1/4\end{aligned}$$

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

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

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

$$\begin{aligned}\Pr(D_1 \in \{1,3,5\} \cap D_2 \in \{1,3,5\}) &= \Pr(D_1 \in \{1,3,5\}) \cdot \Pr(D_2 \in \{1,3,5\}) \\ &= 1/4\end{aligned}$$

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

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

Example: What is the probability that for two independent dice rolls, the first is 1 and the sum of the two dice is odd?

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

$$\begin{aligned}\Pr(D_1 \in \{1,3,5\} \cap D_2 \in \{1,3,5\}) &= \Pr(D_1 \in \{1,3,5\}) \cdot \Pr(D_2 \in \{1,3,5\}) \\ &= 1/4\end{aligned}$$

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

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

Example: What is the probability that for two independent dice rolls, the first is 1 and the sum of the two dice is odd?

$$\begin{aligned}&\Pr(D_1 = 1 \cap D_1 + D_2 \text{ is odd}) \\ &= \Pr(D_1 = 1) \cdot \Pr(D_1 + D_2 \in \{1,3,5,7,9,11\} | D_1 = 1) \\ &= 1/6 \cdot \Pr(D_2 \in \{2,4,6\}) = 1/6 \cdot 1/2 = \Pr(D_1 = 1) \cdot \Pr(D_1 + D_2 \text{ is odd})\end{aligned}$$

LINEARITY OF EXPECTATION AND VARIANCE

When are the expectation and variance linear?

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

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

and

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

$\mathbf{X}, \mathbf{Y}$: any two random variables.

LINEARITY OF EXPECTATION

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

LINEARITY OF EXPECTATION

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

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

Proof:

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

Proof:

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

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

Proof:

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

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

Proof:

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

LINEARITY OF EXPECTATION

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

Proof:

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

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

Proof:

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

(law of total probability)

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

Proof:

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

LINEARITY OF VARIANCE

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

LINEARITY OF VARIANCE

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

LINEARITY OF VARIANCE

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

Exercise 1: $\text{Var}[\mathbf{X}] = \mathbb{E}[\mathbf{X}^2] - \mathbb{E}[\mathbf{X}]^2$

LINEARITY OF VARIANCE

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

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

LINEARITY OF VARIANCE

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

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

Exercise 2: $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

LINEARITY OF VARIANCE

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

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

Exercise 2: $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

LINEARITY OF VARIANCE

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

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

Exercise 2: $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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

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

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

Exercise 2: $\mathbb{E}[\mathbf{XY}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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

(linearity of expectation)

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

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

Exercise 2: $\mathbb{E}[\mathbf{XY}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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

LINEARITY OF VARIANCE

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

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

Exercise 2: $\mathbb{E}[\mathbf{XY}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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

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

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

Exercise 2: $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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

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

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

Exercise 2: $\mathbb{E}[\mathbf{X}\mathbf{Y}] = \mathbb{E}[\mathbf{X}] \cdot \mathbb{E}[\mathbf{Y}]$ when $\mathbf{X}, \mathbf{Y}$ are independent.

Together give:

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