Algorithms M2 IF

More on Randomized Algorithms

Michael Lampis

Fall 2019

Review of Probability Theory

Definition of Probability

To define probabilities we need:

- A “universe” of events Ω
- A collection of **events** $\mathcal{E}$ such that for $E \in \mathcal{E}$ we have $E \subseteq \Omega$
- A **probability** function $Pr : \mathcal{E} \to [0, 1]$

Definition of Probability

To define probabilities we need:

- A "universe" of events Ω
- A collection of **events** $\mathcal{E}$ such that for $E \in \mathcal{E}$ we have $E \subseteq \Omega$
- A **probability** function $Pr : \mathcal{E} \to [0, 1]$

Example:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- The following could be events in $\mathcal{E}$
    - $E_3 = \{3\}$
    - $E_{\text{low}} = \{1, 2\}$
    - $E_{\text{odd}} = \{1, 3, 5\}$
- The natural (uniform) probability function would set
    - $Pr[E_3] = 1/6$
    - $Pr[E_{\text{low}}] = 1/3$
    - $Pr[E_{\text{odd}}] = 1/2$

Definition of Probability

To define probabilities we need:

- A “universe” of events Ω
- A collection of **events** $\mathcal{E}$ such that for $E \in \mathcal{E}$ we have $E \subseteq \Omega$
- A **probability** function $Pr : \mathcal{E} \to [0, 1]$

Example (infinite space):

- $\Omega = [0, 1]$.
- The following could be events in $\mathcal{E}$
  - $E_3 = \{1/3\}$
  - $E_{\text{low}} = [0, 1/2]$
  - $E_{\text{edge}} = [0, 1/4] \cup [3/4, 1]$
- The natural (uniform) probability function would set
  - $Pr[E_3] = 0$ (why?)
  - $Pr[E_{\text{low}}] = 1/2$
  - $Pr[E_{\text{edge}}] = 1/2$

Definition of Probability

To define probabilities we need:

- A "universe" of events Ω
- A collection of **events** $\mathcal{E}$ such that for $E \in \mathcal{E}$ we have $E \subseteq \Omega$
- A **probability** function $Pr : \mathcal{E} \to [0, 1]$

A **valid** probability measure satisfies:

- $Pr[\Omega] = 1$
- If $E_1, E_2, \dots, E_n \in \mathcal{E}$ and for all $i \neq j, E_i \cap E_j = \emptyset$ (mutually disjoint events), then

$$Pr[\cup_{i=1}^{n} E_i] = \sum_{i=1}^{n} Pr[E_i]$$

These are called the Kolmogorov probability axioms.

- When Ω is finite, the distribution which sets for each $i \in \Omega$ $Pr[\{i\}] = \frac{1}{|\Omega|}$ is called the **uniform distribution**.

Probability Basics

Remember: probabilities are **sets** deep down.

- $Pr[\emptyset] = 0$
- If $E_1 \subseteq E_2$ then $Pr[E_1] \leq Pr[E_2]$
- $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$
  - Proof?

The last principle can be generalized to give the so-called inclusion-exclusion formula:

$$\begin{aligned} Pr[A_1 \cup A_2 \cup \dots A_n] &= \sum_{i=1}^{n} Pr[A_i] - \sum_{i_1 \neq i_2 = 1}^{n} Pr[A_{i_1} \cap A_{i_2}] + \\ &\sum_{i_1 \neq i_2 \neq i_3 = 1}^{n} Pr[A_{i_1} \cap A_{i_2} \cap A_{i_3}] - \dots \end{aligned}$$

Union bound

A very basic property that follows for **any** collection of events:

$$Pr[\cup_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} Pr[A_i]$$

- This is called the **union bound**.
- We often use this bound when A_i are “bad” events, and we want to show that the probability of one of them happening is small.
  - Main interest: it might be hard to calculate exactly $Pr[\cup A_i]$. This allows us to upper bound it without worrying about how each event affects the others.
- The bound becomes an equality only when events are disjoint (mutually exclusive).

Independence

- Informally: a set of events is **independent**, if knowing that one happened gives us no additional information about the others.
- Formally: A, B independent if $Pr[A \cap B] = Pr[A] \cdot Pr[B]$.
- Formally: $A_1, \ldots, A_n$ independent if for any $S \subseteq \{1, \ldots n\}$ we have $Pr[\cap_{i \in S} A_i] = \Pi_{i \in S} Pr[A_i]$.

Independence

- Informally: a set of events is **independent**, if knowing that one happened gives us no additional information about the others.
- Formally: A, B independent if $Pr[A \cap B] = Pr[A] \cdot Pr[B]$.
- Formally: $A_1, \ldots, A_n$ independent if for any $S \subseteq \{1, \ldots n\}$ we have $Pr[\cap_{i \in S} A_i] = \Pi_{i \in S} Pr[A_i]$.
  - What is the difference between independence for two and for more than two events?
- **Pair-wise independence**: $A_1, \ldots, A_n$ are pair-wise independent iff for any $i \neq j \in \{1, \ldots, n\}$ we have $Pr[A_i \cap A_j] = Pr[A_i] \cdot Pr[A_j]$.

Independence

- Informally: a set of events is **independent**, if knowing that one happened gives us no additional information about the others.
- Formally: A, B independent if $Pr[A \cap B] = Pr[A] \cdot Pr[B]$.
- Formally: $A_1, \ldots, A_n$ independent if for any $S \subseteq \{1, \ldots n\}$ we have $Pr[\cap_{i \in S} A_i] = \Pi_{i \in S} Pr[A_i]$.
    - What is the difference between independence for two and for more than two events?
- **Pair-wise independence**: $A_1, \ldots, A_n$ are pair-wise independent iff for any $i \neq j \in \{1, \ldots, n\}$ we have $Pr[A_i \cap A_j] = Pr[A_i] \cdot Pr[A_j]$.

Example: roll a die

- A: result is odd
- B: result is divisible by three
- C: result is ≥ 4

Independence

- Informally: a set of events is **independent**, if knowing that one happened gives us no additional information about the others.
- Formally: A, B independent if $Pr[A \cap B] = Pr[A] \cdot Pr[B]$.
- Formally: $A_1, \ldots, A_n$ independent if for any $S \subseteq \{1, \ldots n\}$ we have $Pr[\cap_{i \in S} A_i] = \Pi_{i \in S} Pr[A_i]$.
  - What is the difference between independence for two and for more than two events?
- **Pair-wise independence**: $A_1, \ldots, A_n$ are pair-wise independent iff for any $i \neq j \in \{1, \ldots, n\}$ we have $Pr[A_i \cap A_j] = Pr[A_i] \cdot Pr[A_j]$.

Example: roll a die

- A: result is odd
- B: result is divisible by three
- C: result is ≥ 4
- A, B are independent; A, C are not; B, C are independent.

Conditional Probabilities

- To define independence we asked “Does A tell us anything about B?”
- This corresponds to the notion of **conditional probabilities**:
- Definition:

$$Pr[A \mid B] = \frac{Pr[A \cap B]}{Pr[B]}$$

- In words: the probability of A, given B.
- Note: only makes sense if $Pr[B] \neq 0$.

Conditional Probabilities

- To define independence we asked "Does A tell us anything about B?"
- This corresponds to the notion of **conditional probabilities**:
- Definition:

$$Pr[A \mid B] = \frac{Pr[A \cap B]}{Pr[B]}$$

- In words: the probability of A, given B.
- Note: only makes sense if $Pr[B] \neq 0$.
- So, if A, B independent, then $Pr[A \mid B] = Pr[A]$.
  - Makes sense!

Conditional Probabilities

- To define independence we asked “Does A tell us anything about B?”
- This corresponds to the notion of **conditional probabilities**:
- Definition:

$$Pr[A \mid B] = \frac{Pr[A \cap B]}{Pr[B]}$$

- In words: the probability of A, given B.
- Note: only makes sense if $Pr[B] \neq 0$.
- So, if A, B independent, then $Pr[A \mid B] = Pr[A]$.
    - Makes sense!
- Important not to confuse $Pr[A \mid B]$ with $Pr[B \mid A]$.
    - $Pr[\text{I sneeze} \mid \text{I have a cold}] \neq Pr[\text{I have a cold} \mid \text{I sneeze}]$
- $Pr[A \mid B]Pr[B] = Pr[B \mid A]Pr[A] = Pr[A \cap B]$.

Useful Tools From Probability Theory

Expectation

- Random variable: a function $X : \Omega \to \mathbb{R}$.
- Informally: a variable whose value depends on the outcome of a random event.

Expectation

- Random variable: a function $X : \Omega \to \mathbb{R}$.
- Informally: a variable whose value depends on the outcome of a random event.

  Example: we roll a die
- If X is the number shown, X is a random variable that takes values in $\{1, \ldots, 6\}$.
- $Pr[X = 1] = \frac{1}{6}$
- If we roll three dice, let Y be (a r.v. equal to) their sum
- Y takes values in $\{3, \ldots, 18\}$
- $Pr[Y = 3] = \frac{1}{6^3}$ (why?)

Expectation

- Random variable: a function $X : \Omega \to \mathbb{R}$.
- Informally: a variable whose value depends on the outcome of a random event.

Example: we roll a die

- If X is the number shown, X is a random variable that takes values in $\{1, \ldots, 6\}$.
- $Pr[X = 1] = \frac{1}{6}$
- If we roll three dice, let Y be (a r.v. equal to) their sum
- Y takes values in $\{3, \ldots, 18\}$
- $Pr[Y = 3] = \frac{1}{6^3}$ (why?)

Expectation (discrete variables)

- For a variable $X : \Omega \to \mathbb{Z}$ we define

$$E[X] = \sum_{i \in \mathbb{Z}} i \cdot Pr[X = i]$$

- Informally $E[X]$ is the “average” value of X.

Expectation – Geometric distribution

- We have a coin which comes up heads with probability p. We start flipping it until it comes up heads.
- Let X be the number of times we flipped it.
- X follows a **geometric distribution**.
- What is $E[X]$?

Expectation – Geometric distribution

- We have a coin which comes up heads with probability p. We start flipping it until it comes up heads.
- Let X be the number of times we flipped it.
- X follows a **geometric distribution**.
- What is $E[X]$?

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{\infty} iPr[X=i] = \\
& \sum_{i=1}^{\infty} ip(1-p)^{i-1} = \\
& -p\sum_{i=0}^{\infty}\frac{d}{dp}((1-p)^i) = -p\frac{d}{dp}(\sum_{i=0}^{\infty}(1-p)^i) = \\
& = -p\frac{d}{dp}(\frac{1}{p}) = \frac{1}{p}
\end{aligned}
$$

- Makes sense!

Linearity of Expectations

Why do we like expectations so much?

- Relatively easy to calculate
- Gives a good estimate for value of r.v. with high probability (using Markov, Chebyshev, Chernoff,. . .)

Linearity of Expectations

Why do we like expectations so much?

- Relatively easy to calculate
- Gives a good estimate for value of r.v. with high probability (using Markov, Chebyshev, Chernoff,...)
- Why are they easy to calculate?

  Linearity of expectations
- For random variables $X_1, \ldots, X_n$, constants $a_1, \ldots, a_n \in \mathbb{R}$ we have

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]$$

Linearity of Expectations

Why do we like expectations so much?

- Relatively easy to calculate
- Gives a good estimate for value of r.v. with high probability (using Markov, Chebyshev, Chernoff,...)
- Why are they easy to calculate?

Linearity of expectations

- For random variables $X_1, \ldots, X_n$, constants $a_1, \ldots, a_n \in \mathbb{R}$ we have

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i E[X_i]$$

- **Important** We don't care if the X_i's are independent or not!

An application: Coupon Collector

- Experiment: we throw a die until we have seen all possible numbers as outcomes.
- Let X be the number of throws until we stop.
- $E[X] = ?$ (if the die has k sides)

An application: Coupon Collector

- Experiment: we throw a die until we have seen all possible numbers as outcomes.
- Let X be the number of throws until we stop.
- $E[X] =?$ (if the die has k sides)
- Define X_i, throws needed to see the i-th distinct number, after we have already seen $i-1$ distinct numbers.
- $X_1 = 1$.

An application: Coupon Collector

- Experiment: we throw a die until we have seen all possible numbers as outcomes.
- Let X be the number of throws until we stop.
- $E[X] =?$ (if the die has k sides)
- Define X_i, throws needed to see the i-th distinct number, after we have already seen $i-1$ distinct numbers.
- $X_1 = 1$.
- X_2 follows a geom. dist. with probability $p_2 = \frac{n-1}{n}$
- X_i follows a geom. dist. with probability $p_i = \frac{n-i+1}{n}$
- $X = \sum_{i=1}^{n} X_i$

An application: Coupon Collector

- Experiment: we throw a die until we have seen all possible numbers as outcomes.
- Let X be the number of throws until we stop.
- $E[X] =?$ (if the die has k sides)
- Define X_i, throws needed to see the i-th distinct number, after we have already seen $i-1$ distinct numbers.
- $X_1 = 1$.
- X_2 follows a geom. dist. with probability $p_2 = \frac{n-1}{n}$
- X_i follows a geom. dist. with probability $p_i = \frac{n-i+1}{n}$
- $X = \sum_{i=1}^{n} X_i$

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n} E[X_i] = \\
&\sum_{i=1}^{n} \frac{1}{p_i} = n \sum_{i=1}^{n} \frac{1}{n-i+1} \approx n \ln n
\end{aligned}
$$

When average is not enough

- Calculating $E[X]$ is usually only a first step.
- We want to show that X is "good" (close to $E[X]$) with high probability.
- For this, we need to use various helpful inequalities.

When average is not enough

- Calculating $E[X]$ is usually only a first step.
- We want to show that X is "good" (close to $E[X]$) with high probability.
- For this, we need to use various helpful inequalities.
- **Markov's inequality**
- Assumes that X is always ≥ 0 ($Pr[X < 0] = 0$)

$$Pr[X > \alpha E[X]] \leq \frac{1}{\alpha}$$

When average is not enough

- Calculating $E[X]$ is usually only a first step.
- We want to show that X is "good" (close to $E[X]$) with high probability.
- For this, we need to use various helpful inequalities.
- **Markov's inequality**
- Assumes that X is always ≥ 0 ($Pr[X < 0] = 0$)

$$Pr[X > \alpha E[X]] \leq \frac{1}{\alpha}$$

- Proof:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} iPr[X = i] \geq \\ & \sum_{i=\alpha E[X]}^{\infty} iPr[X = i] \geq \alpha E[X] Pr[X \geq \alpha E[X]] \end{aligned}$$

When average is not enough

- Calculating $E[X]$ is usually only a first step.
- We want to show that X is "good" (close to $E[X]$) with high probability.
- For this, we need to use various helpful inequalities.
- **Markov's inequality**
- Assumes that X is always ≥ 0 ($Pr[X < 0] = 0$)

$$Pr[X > \alpha E[X]] \leq \frac{1}{\alpha}$$

- Proof:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} iPr[X = i] \geq \\ & \sum_{i=\alpha E[X]}^{\infty} iPr[X = i] \geq \alpha E[X] Pr[X \geq \alpha E[X]] \end{aligned}$$

- Makes sense!

Markov collects coupons

Connecting the two previous slides:

- If X is the number of repetitions until we see all numbers, $E[X] = n \ln n$
- For all $\alpha > 0$, $Pr[X > \alpha E[X]] \leq \frac{1}{\alpha}$
- $\Rightarrow Pr[X > 100n \ln n] \leq \frac{1}{100}$
- With high probability $X = O(n \log n)$
- Note: we use the fact that $X \geq 0$

Using Variance

Variance

- A basic way to bound the distance of X from $E[X]$ is to calculate $Var[X]$
- Definition:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- Reminder: we often write $\sigma = \sqrt{Var[X]}$ to denote the **standard deviation** of X.

Variance

- A basic way to bound the distance of X from $E[X]$ is to calculate $Var[X]$
- Definition:

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- Reminder: we often write $\sigma = \sqrt{Var[X]}$ to denote the **standard deviation** of X.
- Reminder: variance is **not** as nice as expectation.
- Example: in general $Var[X + Y] \neq Var[X] + Var[Y]$
  - However, $Var[X + Y] = Var[X] + Var[Y]$ if X, Y independent.

Using Variance

Chebyshev's inequality:

$$Pr[|X - E[X]| \geq \alpha] \leq \frac{Var[X]}{\alpha^2}$$

- In other words, probability that we fall more than $\alpha\sigma(X)$ away from $E[X]$ is at most $\frac{1}{\alpha^2}$.
- This is why $\sigma(X) = \sqrt{Var[X]}$ is called "standard" deviation.

Using Variance

Chebyshev's inequality:

$$Pr[|X - E[X]| \geq \alpha] \leq \frac{Var[X]}{\alpha^2}$$

- In other words, probability that we fall more than $\alpha\sigma(X)$ away from $E[X]$ is at most $\frac{1}{\alpha^2}$.
- This is why $\sigma(X) = \sqrt{Var[X]}$ is called "standard" deviation.
- Proof:

$$\begin{aligned} Pr[|X - E[X]| \geq \alpha] &= Pr[(X - E[X])^2 \geq \alpha^2] \leq \\ &\leq \frac{E[(X - E[X])^2]}{\alpha^2} = \frac{Var[X]}{\alpha^2} \end{aligned}$$

Application: Coupon collector again

- Recall Coupon Collector problem: X is the number of repetitions until we see all outcomes
- $E[X] = n \ln n$
- By Markov, $Pr[X > 2n \ln n] \leq \frac{1}{2}$

Application: Coupon collector again

- Recall Coupon Collector problem: X is the number of repetitions until we see all outcomes
- $E[X] = n \ln n$
- By Markov, $Pr[X > 2n \ln n] \leq \frac{1}{2}$
- Recall that X_i is repetitions in phase i
- $X = \sum X_i$, and the X_i's are independent
- $Var[X] = \sum Var[X_i]$
- Variance of a geometrically distributed random variable?
  - $Var[Y] = \frac{1-p}{p^2}$, for Y geom. with parameter p

Application: Coupon collector again

- Recall Coupon Collector problem: X is the number of repetitions until we see all outcomes
- $E[X] = n \ln n$
- By Markov, $Pr[X > 2n \ln n] \leq \frac{1}{2}$
- Recall that X_i is repetitions in phase i
- $X = \sum X_i$, and the X_i's are independent
- $Var[X] = \sum Var[X_i]$
- Variance of a geometrically distributed random variable?
  - $Var[Y] = \frac{1-p}{p^2}$, for Y geom. with parameter p

$$
\begin{aligned}
Var[X] &= \sum_{i=1}^{n} Var[X_i] \leq \sum_{i=1}^{n} \left(\frac{n}{n-i+1} \right)^2 = \\
&= n^2 \sum_{i=1}^{n} \frac{1}{i^2} \leq \frac{\pi^2 n^2}{6}
\end{aligned}
$$

Application: Coupon collector again

$$Var[X] \leq \frac{\pi^2 n^2}{6}$$

We then use Chebyshev's inequality which gives

$$\begin{aligned} Pr[X > 2n \ln n] &\leq Pr[|X - n \ln n| > n \ln n] \leq \\ &\leq \frac{n^2 \pi^2 / 6}{(n \ln n)^2} = O(\frac{1}{\log^2 n}) \end{aligned}$$

Note: Markov's inequality only gives that this probability is at most $1/2$.

Summary

Important lessons to remember.

- Inclusion-Exclusion: $Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B]$
- Union bound: $Pr[A \cup B] \leq Pr[A] + Pr[B]$
- Linearity of Expectation: $E[X_1 + X_2] = E[X_1] + E[X_2]$
- Markov's inequality: $Pr[X > a] \leq \frac{E[X]}{a}$
- Variance: $Var[X] = E[X^2] - E[X]^2$
- Variance only linear for independent variables!
- Chebyshev's inequality: $Pr[|X - E[X]| > a] \leq \frac{Var[X]}{a^2}$