# Learning Goals

- State the condition Markov inequality
- Understand distributions for which Markov inequality is tight
- Define perfect hashing
- Implementation and proof of perfect hashing
- Understand the method of amplification by independent trials

# Concentration Inqualities

- Often it is not enough to estimate the expectation of a random variable, but to say that with good probability its value is not far from the expectation.

# Concentration Inqualities

- Often it is not enough to estimate the expectation of a random variable, but to say that with good probability its value is not far from the expectation.
- Such a phenomenon is called *concentration*.

# Concentration Inqualities

- Often it is not enough to estimate the expectation of a random variable, but to say that with good probability its value is not far from the expectation.
- Such a phenomenon is called *concentration*.
- Tools that upper bound the probability with which a random variable deviates far from its expectation are known as *concentration inequalities* or *tail bounds*.

# Markov Inequality

**Theorem (Markov Inequality)**

*If $X$ is a random variable that takes nonnegative value with probability $1$, then for any $\alpha > 1$,*

$$\mathbf{Pr}\left[X \geq \alpha\, \mathbf{E}\left[X\right]\right] \leq \frac{1}{\alpha}.$$

# Markov Inequality

**Theorem (Markov Inequality)**

*If $X$ is a random variable that takes nonnegative value with probability 1, then for any $\alpha > 1$,*

$$\mathbf{Pr}\left[X \geq \alpha\, \mathbf{E}\left[X\right]\right] \leq \frac{1}{\alpha}.$$

**Proof.**

Let $Y$ be the indicator variable for $X \geq \alpha\, \mathbf{E}[X]$.

# Markov Inequality

**Theorem (Markov Inequality)**

*If $X$ is a random variable that takes nonnegative value with probability 1, then for any $\alpha > 1$,*

$$\mathbf{Pr}\left[X \geq \alpha\, \mathbf{E}\left[X\right]\right] \leq \frac{1}{\alpha}.$$

**Proof.**

Let $Y$ be the indicator variable for $X \geq \alpha\, \mathbf{E}[X]$. Then

$$\mathbf{Pr}\left[X \geq \mathbf{E}\left[X\right]\right] = \mathbf{Pr}\left[Y = 1\right] = \mathbf{E}\left[Y\right] \leq \mathbf{E}\left[\frac{X}{\alpha\, \mathbf{E}[X]}\right] = \frac{1}{\alpha}.$$

□

# Remarks

- Markov inequality can be understood as: a nonnegative random variable deviates from its expectation by a constant factor with at most constant probability.

# Remarks

- Markov inequality can be understood as: a nonnegative random variable deviates from its expectation by a constant factor with at most constant probability.
- Equivalently, the theorem can be stated as $\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$ for any $a > 0$.

# Remarks

- Markov inequality can be understood as: a nonnegative random variable deviates from its expectation by a constant factor with at most constant probability.
- Equivalently, the theorem can be stated as $\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$ for any $a > 0$.
  - Stated this way, the inequality has bite only for $a > \mathbf{E}[X]$.

# Remarks

- Markov inequality can be understood as: a nonnegative random variable deviates from its expectation by a constant factor with at most constant probability.
- Equivalently, the theorem can be stated as $\mathbf{Pr}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}$ for any $a > 0$.
  - Stated this way, the inequality has bite only for $a > \mathbf{E}[X]$.
- Note the condition that $X$ must be a nonnegative random variable.

# Essence of Markov Inequality

- Essence of the proof: among distributions having the same $\mathbf{Pr}[X > a]$, which one minimizes $\mathbf{E}[X]$?

# Essence of Markov Inequality

- Essence of the proof: among distributions having the same $\mathbf{Pr}[X > a]$, which one minimizes $\mathbf{E}[X]$?
- Answer: when $X < a$, $X$ should be 0; when $X \geq a$, $X$ should be $a$.

# Essence of Markov Inequality

- Essence of the proof: among distributions having the same $\mathbf{Pr}[X > a]$, which one minimizes $\mathbf{E}[X]$?
- Answer: when $X < a$, $X$ should be 0; when $X \geq a$, $X$ should be $a$.
- The distribution for which Markov inequality tight is a two-point distribution.

# Essence of Markov Inequality

- Essence of the proof: among distributions having the same $\mathbf{Pr}[X > a]$, which one minimizes $\mathbf{E}[X]$?
- Answer: when $X < a$, $X$ should be 0; when $X \geq a$, $X$ should be $a$.
- The distribution for which Markov inequality tight is a two-point distribution.
- With this intuition, it is not difficult to prove the following corollary:

**Corollary (Reverse Markov Inequality)**

*If $X$ is a random variable that is never larger than $a$, then for any $b < a$,*

$$\mathbf{Pr}\left[X \leq b\right] \leq \frac{a - \mathbf{E}[X]}{a - b}.$$

# Application: Perfect Hashing

**Definition**

A hash function $h : U \to \{0, \ldots, m-1\}$ is *perfect* on $S \subseteq U$ if $\textsc{Find}(x)$ for every $x \in S$ takes $O(1)$ time.

# Application: Perfect Hashing

**Definition**

A hash function $h : U \to \{0, \ldots, m-1\}$ is *perfect* on $S \subseteq U$ if $\textsc{Find}(x)$ for every $x \in S$ takes $O(1)$ time.

- Recall: to store a dataset of $n$ entries, if we sample from a universal hash family, it suffices to have a hash table of size $m = \Theta(n)$, so that each element has $O(1)$ collisions in expectation.

# Application: Perfect Hashing

**Definition**

A hash function $h : U \to \{0, \ldots, m-1\}$ is *perfect* on $S \subseteq U$ if $\textsc{Find}(x)$ for every $x \in S$ takes $O(1)$ time.

- Recall: to store a dataset of $n$ entries, if we sample from a universal hash family, it suffices to have a hash table of size $m = \Theta(n)$, so that each element has $O(1)$ collisions in expectation.
- It does not follow immediately that there exists an $h \in H$ under which every element has only $O(1)$ collisions.

# Application: Perfect Hashing

**Definition**

A hash function $h : U \to \{0, \ldots, m-1\}$ is *perfect* on $S \subseteq U$ if $\textsc{Find}(x)$ for every $x \in S$ takes $O(1)$ time.

- Recall: to store a dataset of $n$ entries, if we sample from a universal hash family, it suffices to have a hash table of size $m = \Theta(n)$, so that each element has $O(1)$ collisions in expectation.
- It does not follow immediately that there exists an $h \in H$ under which every element has only $O(1)$ collisions.
  - In fact, with an "ideal hash", i.e., that sends every element in $U$ uniformly at random to $\{0, \ldots, m-1\}$, for $m = n$, w.h.p. the worst bucket has $\Theta(\log n / \log \log n)$ collisions.

# Low Collision with More Space

- First let's relax the problem: with how much space (not necessarily $O(n)$) do we know how to hash with low collision?

# Low Collision with More Space

- First let's relax the problem: with how much space (not necessarily $O(n)$) do we know how to hash with low collision?

**Claim**

Let $H$ be a universal hash family from $U$ to $\{0, \ldots, m\}$, then for any $S \subseteq U$ with $|S| = n \leq \sqrt{m}$, for a random $h$ from $H$, with probability at least $\frac{1}{2}$, there is no collision under $h$.

# Low Collision with More Space

- First let's relax the problem: with how much space (not necessarily $O(n)$) do we know how to hash with low collision?

**Claim**

Let $H$ be a universal hash family from $U$ to $\{0, \ldots, m\}$, then for any $S \subseteq U$ with $|S| = n \leq \sqrt{m}$, for a random $h$ from $H$, with probability at least $\frac{1}{2}$, there is no collision under $h$.

**Proof.**

By definition of universal hashing, for every $x \neq y$ in $S$, $\mathbf{Pr}_{h \sim H}[h(x) = h(y)] \leq \frac{1}{m}$.

# Low Collision with More Space

- First let's relax the problem: with how much space (not necessarily $O(n)$) do we know how to hash with low collision?

**Claim**

Let $H$ be a universal hash family from $U$ to $\{0, \ldots, m\}$, then for any $S \subseteq U$ with $|S| = n \leq \sqrt{m}$, for a random $h$ from $H$, with probability at least $\frac{1}{2}$, there is no collision under $h$.

**Proof.**

By definition of universal hashing, for every $x \neq y$ in $S$, $\mathbf{Pr}_{h \sim H}[h(x) = h(y)] \leq \frac{1}{m}$.
By the union bound, the probability that any collision happens is at most $\sum_{x \neq y \in S} \frac{1}{m} < \frac{n^2}{2} \cdot \frac{1}{m} \leq \frac{1}{2}$. □

# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?

# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?
- This is not an easy question, and remained open for many years. We present the first solution, given by Fredman, Komlós and Szemerédi (1982).

# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?
- This is not an easy question, and remained open for many years. We present the first solution, given by Fredman, Komlós and Szemerédi (1982).
- Main idea: use two levels of hashing.

# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?
- This is not an easy question, and remained open for many years. We present the first solution, given by Fredman, Komlós and Szemerédi (1982).
- Main idea: use two levels of hashing.
  - Let $A[\cdot]$ be the array for the first level hash, and $h$ be a hash function from $U$ to $\{0, \ldots, n-1\}$.

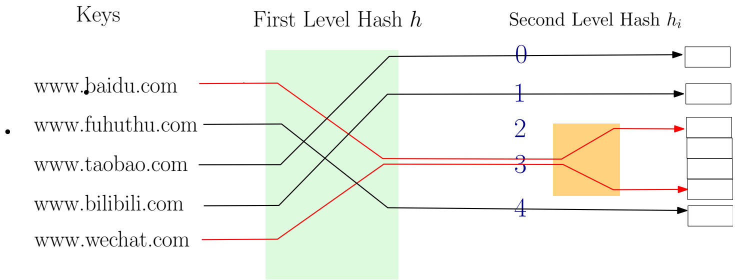
# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?
- This is not an easy question, and remained open for many years. We present the first solution, given by Fredman, Komlós and Szemerédi (1982).
- Main idea: use two levels of hashing.
    - Let $A[\cdot]$ be the array for the first level hash, and $h$ be a hash function from $U$ to $\{0, \ldots, n-1\}$.
    - For each $i = 0, \ldots, n-1$, let $n_i$ be the number of collisions in that bucket. Set up a hash table $B_i$ of size $n_i^2$, and a *perfect* hash function mapping $U$ to $\{0, \ldots, n_i^2 - 1\}$.

# Perfect Hashing in $O(n)$ Space

- Is it possible to have perfect hashing with $m = O(n)$?
- This is not an easy question, and remained open for many years. We present the first solution, given by Fredman, Komlós and Szemerédi (1982).
- Main idea: use two levels of hashing.
  - Let $A[\cdot]$ be the array for the first level hash, and $h$ be a hash function from $U$ to $\{0, \ldots, n-1\}$.
  - For each $i = 0, \ldots, n-1$, let $n_i$ be the number of collisions in that bucket. Set up a hash table $B_i$ of size $n_i^2$, and a *perfect* hash function mapping $U$ to $\{0, \ldots, n_i^2 - 1\}$.
  - When looking up $x$, we first find its position in the first level. Let $j$ be $h(x)$. Then we look up $B_j[h_j(x)]$.

# Illustration: Perfect Hashing



# Space Calculation

- The resulting hash function is obviously perfect. The remaining question is whether we satisfy the space constraint.

# Space Calculation

- The resulting hash function is obviously perfect. The remaining question is whether we satisfy the space constraint.
- We need $h$ to satisfy $\sum_i n_i^2 = O(n)$.

# Space Calculation

- The resulting hash function is obviously perfect. The remaining question is whether we satisfy the space constraint.
- We need $h$ to satisfy $\sum_i n_i^2 = O(n)$.

**Lemma**

*Let $h$ be sampled uniformly at random from a universal hash function family mapping $U$ to $\{0, \ldots, n-1\}$. Let $n_i$ be $|h^{-1}(i)|$, the number of elements mapped to $i$ by $h$. Then $\mathbf{Pr}[\sum_i n_i^2 \leq 4n] \geq \frac{1}{2}$.*

# Space Calculation

- The resulting hash function is obviously perfect. The remaining question is whether we satisfy the space constraint.
- We need $h$ to satisfy $\sum_i n_i^2 = O(n)$.

**Lemma**

*Let $h$ be sampled uniformly at random from a universal hash function family mapping $U$ to $\{0, \ldots, n-1\}$. Let $n_i$ be $|h^{-1}(i)|$, the number of elements mapped to $i$ by $h$. Then $\mathbf{Pr}[\sum_i n_i^2 \leq 4n] \geq \frac{1}{2}$.*

**Proof.**

Game plan: we first show that $\mathbf{E}[\sum_i n_i^2]$ is no more than $2n$. Then the conclusion follows from Markov inequality.

# Space Calculation

- The resulting hash function is obviously perfect. The remaining question is whether we satisfy the space constraint.
- We need $h$ to satisfy $\sum_i n_i^2 = O(n)$.

**Lemma**

*Let $h$ be sampled uniformly at random from a universal hash function family mapping $U$ to $\{0, \ldots, n-1\}$. Let $n_i$ be $|h^{-1}(i)|$, the number of elements mapped to $i$ by $h$. Then $\mathbf{Pr}[\sum_i n_i^2 \leq 4n] \geq \frac{1}{2}$.*

**Proof.**

Game plan: we first show that $\mathbf{E}[\sum_i n_i^2]$ is no more than $2n$. Then the conclusion follows from Markov inequality.
For $x \neq y$ in $S$, let $C_{xy}$ be the indicator variable for the event that $x$ clashes with $y$ under $h$, then $\mathbf{E}[C_{xy}] \leq \frac{1}{n}$ by universality.

# Proof of Lemma (Cont.)

**Proof.**

Key observation: $\sum_i n_i^2 = n + \sum_{x \in S} \sum_{y \in S \setminus \{x\}} C_{xy}$.

# Proof of Lemma (Cont.)

**Proof.**

Key observation: $\sum_i n_i^2 = n + \sum_{x \in S} \sum_{y \in S \setminus \{x\}} C_{xy}$.
To see this, let $S_i$ be $h^{-1}(i)$, then $\sum_{x \in S_i} \sum_{y \in S \setminus \{x\}} C_{xy} = n_i(n_i - 1)$.

# Proof of Lemma (Cont.)

**Proof.**

Key observation: $\sum_i n_i^2 = n + \sum_{x \in S} \sum_{y \in S \setminus \{x\}} C_{xy}$.
To see this, let $S_i$ be $h^{-1}(i)$, then $\sum_{x \in S_i} \sum_{y \in S \setminus \{x\}} C_{xy} = n_i(n_i - 1)$.
$\Rightarrow \sum_x \sum_{y \neq x} C_{xy} = \sum_i \sum_{x \in S_i} \sum_{y \neq x} C_{xy} = \sum_i n_i(n_i - 1)$.

# Proof of Lemma (Cont.)

**Proof.**

Key observation: $\sum_i n_i^2 = n + \sum_{x \in S} \sum_{y \in S \setminus \{x\}} C_{xy}$.
To see this, let $S_i$ be $h^{-1}(i)$, then $\sum_{x \in S_i} \sum_{y \in S \setminus \{x\}} C_{xy} = n_i(n_i - 1)$.
$\Rightarrow \sum_x \sum_{y \neq x} C_{xy} = \sum_i \sum_{x \in S_i} \sum_{y \neq x} C_{xy} = \sum_i n_i(n_i - 1)$.
Now we can bound

$$\mathbf{E}\left[\sum_{x \in S} \sum_{y \in S \setminus \{x\}}\right] C_{xy} \leq n(n-1) \cdot \frac{1}{n} \leq n.$$

# Proof of Lemma (Cont.)

**Proof.**

Key observation: $\sum_i n_i^2 = n + \sum_{x \in S} \sum_{y \in S \setminus \{x\}} C_{xy}$.
To see this, let $S_i$ be $h^{-1}(i)$, then $\sum_{x \in S_i} \sum_{y \in S \setminus \{x\}} C_{xy} = n_i(n_i - 1)$.
$\Rightarrow \sum_x \sum_{y \neq x} C_{xy} = \sum_i \sum_{x \in S_i} \sum_{y \neq x} C_{xy} = \sum_i n_i(n_i - 1)$.
Now we can bound

$$\mathbf{E}\left[\sum_{x \in S} \sum_{y \in S \setminus \{x\}}\right] C_{xy} \leq n(n-1) \cdot \frac{1}{n} \leq n.$$

Therefore $\mathbf{E}[\sum_i n_i^2] \leq 2n$. □

# Amplification by Repeated Trial

- How do we make use of the lemma?

# Amplification by Repeated Trial

- How do we make use of the lemma?
- Each time we sample an $h$, we satisfy the space requirement with probability at least $\frac{1}{2}$.

# Amplification by Repeated Trial

- How do we make use of the lemma?
- Each time we sample an $h$, we satisfy the space requirement with probability at least $\frac{1}{2}$.
- We can check if we succeed in polynomial time. If not, we simply try again.

# Amplification by Repeated Trial

- How do we make use of the lemma?
- Each time we sample an $h$, we satisfy the space requirement with probability at least $\frac{1}{2}$.
- We can check if we succeed in polynomial time. If not, we simply try again.
- After $k$ trials, we succeed with probability $1 - \frac{1}{2^k}$.