**Inequality 1: Markov’s Inequality**

If $X$ is a random variable which takes non-negative values, and $a > 0$, then

$$\Pr[X \geq a] \leq \frac{E[X]}{a}$$

**Example:** $n$ tosses of an unbiased coin. $X = \#\text{heads}$

$E[X] = n/2$. Let $a = 3n/4$. By Markov’s Inequality, $\Pr(X \geq a) \leq 2/3$. But what is it really?

$$\Pr[X \geq \frac{3n}{4}] \leq \left( \left( \binom{n}{3n/4} + \binom{n}{3n/4 + 1} + \ldots + \binom{n}{n} \right) \cdot 2^{-n} \leq n \cdot 2^{-n} \cdot \binom{n}{n/4} \right)$$

$$\leq n \cdot 2^{-n} \cdot (4e)^{n/4} \leq n(e/4)^{n/4} < (e/3)^{n/4} \quad \text{for large } n$$

**Fact:** If $n \geq k$

$$\left( \frac{n}{k} \right)^k \leq \binom{n}{k} \leq \left( \frac{ne}{k} \right)^k$$

**Summary:** Markov’s inequality can be weak, but it only requires $E[X]$ to be finite!
Inequality 2: Chebyshev’s Inequality

If $X$ is a random variable and $a > 0$, then

$$
\Pr(|X - E[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}
$$

**Example:** $n$ tosses of an unbiased coin. $X = \#\text{heads}$

$E[X] = n/2$  \hspace{1cm} $\text{Var}[X] = n/4$ (how would you compute this?)

From last slide, $\Pr(X \geq 3n/4) \leq c^{n/4}$ for some constant $c < 1$, and large enough $n$

Let $a = n/4$, so that we compute $\Pr(X \geq 3n/4)$. By Chebyshev, $\Pr(X \geq 3n/4) \leq 4/n$

**Summary:** Chebyshev’s inequality can also be weak, but only requires finite $\text{Var}[X], E[X]$
Inequality 3: Chernoff Bounds

Let $X_1, \ldots, X_n$ be independent 0/1 random variables, and $X = X_1 + \ldots + X_n$. Then, for any $t > 0$,

$$\Pr(X \geq (1 + t)E[X]) \leq \left(\frac{e^t}{(1 + t)^{1+t}}\right)^{E[X]}$$

Moreover, for $t < 1$,

$$\Pr(X \leq (1 - t)E[X]) \leq e^{-\frac{1}{2}t^2E[X]}$$

**Example:** $n$ tosses of an unbiased coin. $X = \#$heads$= X_1 + \ldots + X_n$ where $X_i = 1$ if toss $i$ = head $E[X] = n/2$. $\Pr[X \geq 3n/4] = \Pr[X \geq (1 + 1/2) E[X]]$, so $t = 1/2$

Thus from Chernoff Bounds,

$$\Pr(X \geq 3n/4) \leq \left(e^{1/2} \cdot (2/3)^{3/2}\right)^{n/2} \leq (0.88)^{n/2}$$

**Summary:** Stronger bound, but needs independence!
Chernoff Bounds: Simplified Version

Let $X_1, \ldots, X_n$ be independent $0/1$ random variables, and $X = X_1 + \ldots + X_n$. Then, for any $t > 0$,

$$
\Pr(X \geq (1 + t)E[X]) \leq \left( \frac{e^t}{(1 + t)^{1+t}} \right)^{E[X]}
$$

Moreover, for $t < 1$,

$$
\Pr(X \leq (1 - t)E[X]) \leq e^{-\frac{1}{2}t^2 E[X]}
$$

**Simplified Version:**

Let $X_1, \ldots, X_n$ be independent $0/1$ random variables, and $X = X_1 + \ldots + X_n$. Then, for $t < 2e - 1$,

$$
\Pr(X > (1 + t)E[X]) \leq e^{-t^2 E[X]/4}
$$
Randomized Algorithms

• Contention Resolution
• Some Facts about Random Variables
• Global Minimum Cut Algorithm
• Randomized Selection and Sorting
• Max 3-SAT
• Three Concentration Inequalities
• Hashing and Balls and Bins
**Hashing and Balls-n-Bins**

**Problem:** Given a large set $S$ of elements $x_1, \ldots, x_n$, store them using $O(n)$ space s.t it is easy to determine whether a query item $q$ is in $S$ or not.

**Popular Data Structure:** A Hash table

**Algorithm:**
1. Pick a completely random function $h : U \rightarrow \{1, \ldots, n\}$
2. Create a table of size $n$, initialize it to null
3. Store $x_i$ in the linked list at position $h(x_i)$ of table
4. For a query $q$, look at the linked list at location $h(q)$ of table to see if $q$ is there

What is the query time of the algorithm?
**Problem:** Given a large set $S$ of elements $x_1, \ldots, x_n$, store them using $O(n)$ space s.t. it is easy to determine whether a query item $q$ is in $S$ or not.

**Algorithm:**
1. Pick a completely random function $h$
2. Create a table of size $n$, initialize it to null
3. Store $x_i$ in the linked list at position $h(x_i)$ of table
4. For a query $q$, check the linked list at location $h(q)$

**Average Query Time:** Suppose $q$ is picked at random s.t. it is equally likely to hash to $1, \ldots, n$. What is the expected query time?

Expected Query Time $= \sum_{i=1}^{n} \Pr[q \text{ hashes to location } i] \cdot (\text{length of list at } T[i])$

$= \frac{1}{n} \sum_{i} \text{(length of list at } T[i]) = \frac{1}{n} \cdot n = 1$
Hashing and Balls-n-Bins

**Problem:** Given a large set $S$ of elements $x_1, ..., x_n$, store them using $O(n)$ space s.t it is easy to determine whether a query item $q$ is in $S$ or not.

**Algorithm:**
1. Pick a completely random function $h$
2. Create a table of size $n$, initialize it to null
3. Store $x_i$ in the linked list at position $h(x_i)$ of table
4. For a query $q$, check the linked list at location $h(q)$

**Worst Case Query Time:** For any $q$, what is the query time? (with high probability over the choice of hash functions)

**Equivalent to the following Balls and bins Problem:**
Suppose we toss $n$ balls u.a.r into $n$ bins. What is the max #balls in a bin with high probability?

With high probability (w.h.p) = With probability $1 - 1/poly(n)$
Balls and Bins, again

Suppose we toss n balls u.a.r into n bins. What is the max load of a bin with high probability?

Some Facts:

1. The expected load of each bin is 1

2. What is the probability that each bin has load 1?

\[
\text{Probability} = \frac{\# \text{ permutations}}{\# \text{ ways of tossing n balls to n bins}} = \frac{n!}{n^n}
\]

3. What is the expected \#empty bins?

\[
\text{Pr[Bin } i \text{ is empty]} = \left(1 - \frac{1}{n}\right)^n
\]

\[
E[\# \text{ empty bins}] = n \left(1 - \frac{1}{n}\right)^n = \Theta(n) \quad ( (1-1/n)^n \text{ lies between } 1/4 \text{ and } 1/e \text{ for } n \geq 2 )
\]
Balls and Bins

Suppose we toss $n$ balls u.a.r into $n$ bins. What is the max load of a bin with high probability?

Let $X_i =$ #balls in bin $i$

$$\Pr(X_i \geq t) \leq \left( \frac{n}{t} \right)^{\frac{1}{n^t}} \leq \left( \frac{ne}{t} \right)^{t} \frac{1}{n^t} \leq \left( \frac{e}{t} \right)^{t} \leq \frac{1}{n^2}$$

From Fact: If $n \geq k$

$$\left( \frac{n}{k} \right)^{k} \leq \binom{n}{k} \leq \left( \frac{ne}{k} \right)^{k}$$

Would like this for whp condition

Let $t = \frac{c \log n}{\log \log n}$ for constant $c$

$$\log \left( \frac{t}{e} \right)^{t} = t \log t - t = \frac{c \log n}{\log \log n} \cdot (\log c + \log \log n - \log \log \log n) \geq \frac{c}{2} \log n \geq 2 \log n, \text{ for } c \geq 4$$

For large $n$, this is

$$\geq \frac{1}{2} \log \log n$$

Therefore, w.p. $1/n^2$, there are at least $t$ balls in Bin $i$. What is $\Pr(\text{All bins have } \leq t \text{ balls})$?

Applying Union Bound, $\Pr(\text{All bins have } \leq t \text{ balls}) \geq 1 - 1/n$
Suppose we toss $n$ balls u.a.r into $n$ bins. What is the max load of a bin with high probability?

**Fact:** W.p. $1 - 1/n$, the maximum load of each bin is at most $O(\log n/\log \log n)$

**Fact:** The max loaded bin has $(\log n/3 \log \log n)$ balls with probability at least $1 - \text{const.}/n^{(1/3)}$

Let $X_i = \#\text{balls in bin } i$

$$\Pr(X_i \geq t) \geq \left(\frac{n}{t}\right) \frac{1}{n^t} \left(1 - \frac{1}{n}\right)^{n-t} \geq \left(\frac{n}{t}\right)^t \cdot \frac{1}{n^t} \cdot e^{-1} \geq \frac{1}{et^t}$$

At least $1/en^{1/3}$ for $t = \log n/3 \log \log n$

Let $Y_i = 1$ if bin $i$ has load $t$ or more, $= 0$ otherwise

$$Y = Y_1 + Y_2 + .. + Y_n$$

$$\Pr(Y = 0) = \Pr(\text{No bin has load } t \text{ or more}) \leq \Pr(|Y - E[Y]| \geq E[Y])$$

Using Chebyshev, $\Pr(|Y - E[Y]| \geq E[Y]) \leq \text{Var}(Y)/E(Y)^2$ Which concentration bound to use?
Balls and Bins

Suppose we toss n balls u.a.r into n bins. What is the max load of a bin with high probability?

Fact: W.p. 1 - 1/n, the maximum load of each bin is at most O(log n/log log n)

Fact: The max loaded bin has (log n/3log log n) balls with probability at least 1 - const./n^{1/3}

Let $Y_i = 1$ if bin i has load t or more,
= 0 otherwise

$Y = Y_1 + Y_2 + .. + Y_n$

$Pr(Y = 0) = Pr(\text{No bin has load } \geq t) \leq Pr(|Y - E[Y]| \geq E[Y]) \leq \frac{\text{Var}(Y)}{E(Y)^2}$

$\text{Var}[Y] = \text{Var}[(Y_1 + .. + Y_n)^2] = \sum_i \text{Var}(Y_i) + 2 \sum_{i \neq j} (E[Y_i Y_j] - E[Y_i] E[Y_j])$

Now if i is not j, $Y_i$ and $Y_j$ are negatively correlated, which means that $E[Y_i Y_j] < E[Y_i] E[Y_j]$

Thus,

$\text{Var}(Y) \leq \sum_{i=1}^{n} \text{Var}(Y_i) \leq n \cdot 1$

$Pr(Y = 0) \leq \frac{\text{Var}(Y)}{E(Y)^2} \leq \frac{ne^2}{n^{4/3}} \leq \frac{e^2}{n^{1/3}}$
**The Power of Two Choices**

**Problem:** Given a large set $S$ of elements $x_1, .., x_n$, store them using $O(n)$ space s.t it is easy to determine whether a query item $q$ is in $S$ or not.

*Table*

```
 1 2 3 n
```

Linked list of all $x_i$ s.t $h(x_i) = 2$

**Algorithm:**
1. Pick two completely random functions $h_1: U \rightarrow \{1, \ldots, n\}$, and $h_2: U \rightarrow \{1, \ldots, n\}$
2. Create a table of size $n$, initialize it to null
3. Store $x_i$ at linked list at position $h_1(x_i)$ or $h_2(x_i)$, whichever is shorter
4. For a query $q$, look at the linked list at location $h_1(q)$ and $h_2(q)$ of table to see if $q$ is there

**Equivalent to the following Balls and Bins Problem:** Toss $n$ balls into $n$ bins. For each ball, pick two bins u.a.r and put the ball into the lighter of the two bins.

What is the worst case query time? Answer: $O(\log \log n)$ (proof not in this class)