# Lecture #21: Tail Bounds

## Lecture Presentation

### Review of Preparatory Material

## Application: Hash Tables with Chaining

Once again, consider a ***hash table with chaining***. As before, suppose that the hash table stores a set

$$S = \{k_1, k_2, \ldots, k_n\}$$

of $n$ values from some "universe" $\mathcal{U}$" — where $n$ is a positive integer. Suppose, as well, that the table size is $m$, so that the hash table is constructed, and accessed, using some ***hash function***

$$h : \mathcal{U} \to \{0, 1, 2, \ldots, m-1\}.$$

For $0 \le i \le m-1$, let $S_i$ be the set of integers $j$ such that $h(k_j) = i$:

$$S_i = \{j \in \mathbb{Z} \mid 1 \le j \le n \text{ and } h(k_j) = i\}.$$

Suppose that we now model the problem using the same kind of ***sample space*** as in recent lectures, and that we make the same ***assumptions*** about hash functions, as well.

**What This Means:**

As in the previous lecture, let us consider the size of the set $S_0$, that is, the number of integers $j$ such that $1 \le j \le n$ and the $j^{\text{th}}$ key, $k_j$, is “hashed” to location $0$.

As noted in the previous lecture presentation this can be modelled by a random variable $X_0 : \Omega \to \mathbb{R}$, where

$$X_0 = X_{0,1} + X_{0,2} + \cdots + X_{0,n}$$

such that, for $1 \le i \le n$, $X_{0,i}$ is an *indicator random variable* whose value is $1$ if the $i^{\text{th}}$ key, $k_i$, is hashed to location‘$0$, and whose value is $1$ otherwise.

- Using the probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$ that we have been using so far,

  $$\mathsf{P}(X_{0,i} = 1) = \frac{1}{m}$$

  for every integer $i$ such that $1 \le i \le n$.

***What is* $\mathsf{E}[X]$*?***

***What is* $\mathsf{var}(X_i)$ *for* $1 \le i \le n$*?***

***What Else Can Be Established About $X_{0,1}, X_{0,2}, \ldots, X_{0,n}$?***

***What Can Be Established about*** **var**$(X)$ ***?***

Let $\alpha$ be a real number such that $\alpha > 0$. Suppose we wish that bound the probability that

$$X \geq (\alpha + 1)\mathsf{E}[X],$$

that is, the probability that the value of the random variable $X$ is greater than or equal to $\frac{(\alpha+1)n}{m}$.

***Using Markov's Inequality:***

***Using Chebyshev's Inequality:***

***Using Cantelli's Inequality:***

***Using the Chernoff Bound:***

## Chebyshev's Inequality

***Theorem (Chebyshev's Inequality):*** *Let $\Omega$ be a **finite** sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, let $X$ be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then*

$$\mathsf{P}(|X| \geq a) \leq \frac{\mathsf{E}[X^2]}{a^2}.$$

*Proof:*