Analysis and Design of Algorithms

National University of Singapore

Algorithms

CS3230

Tutorial

Week 6

Question 5

Recall the **universal hashing** assumption discussed in class: for any two distinct keys x and y from universe U,

$$\Pr[h(x) = h(y)] \leq \frac{1}{m}$$

where the probability is over the random choice of the hash function $h: U \to [m]$.

Suppose n elements are hashed using such a hash function h. For $j \in [m]$, let n_j be the number of elements hashing to slot j, and let $\alpha = n/m$ (load factor).

Show that:

$$E\left[\sum_{j \in [m]} n_j^2\right] \leq n(1 + \alpha)$$

1:00

Question 6

- Let A[1..n] be an array of n distinct names. Suppose m of them are male names. We hope to select q male names from A[1..n]. We propose the following algorithm to obtain q male names.

- Since personal data is sensitive, we hope to estimate the expected number of accesses to the array *A*.

- Please compute the expected number of access of Query(A, q).

```
Query(A, q)
Let S=Φ;
for j = 1 to q
    Repeat
        Randomly select k from
            {1, 2, …, n};
        Set B=A[k];
    Until B is a male and k∉S;
    S = {k}∪S;
Report S;
```

1:00