Analysis and Design of Algorithms

National University of Singapore

Algorithms

CS3230

Tutorial

Week 6

Question 1

Suppose the birthday of each person is an independent random variable, uniformly distributed between Jan 1 and Dec 31 (365 days).

For what values of m below is it true that if there are m people in a room, the expected number of pairs with a common birthday is < 1?

- $m = 365$
- $m = \lfloor 365/2 \rfloor$
- $m = \lfloor \sqrt{2 \cdot 365} \rfloor$
- $m = \lfloor \sqrt{365} \rfloor$

Question 1

Answer: C and D

Suppose there are m people in the room.

Define an indicator random variable C_{ij} that is 1 iff the i'th and j'th people in the room have the same birthday.

$$E[C_{ij}] = \Pr[C_{ij} = 1] = 1/365$$

So,

$$E[\text{collisions}] = E\left[\sum_{(i,j):i>j} C_{ij}\right] = \binom{m}{2} \cdot \frac{1}{365}$$

The expected number of collisions is < 1 if $m \leq \sqrt{2 \cdot 365}$.

Question 2

The TA would like to encourage everyone in class to prepare before coming to class. To do that, the TA wants to select students who have not prepared to answer questions in class.

The TA selects *k* students in each tutorial to answer the questions and wants to come up with an algorithm for selecting the students. The worst case for the TA would be the case where only the *k* selected students have prepared, but the other *n*-*k* students in the class did not prepare. The TA thinks that an uncooperative class can make the worst case happen every time if the algorithm for selecting the students is made known. Hence there is no choice but to hide the selection algorithm **(source code or pseudocode)** from the class. Is it true that if the TA makes the selection algorithm known, the students will be able to force the worst case to happen?

- Yes
- No

Question 2

Answer: False

If the TA uses a randomized algorithm, it does not matter if he reveals the pseudo code or source code of his algorithm to the class as long as the random numbers are secret.

For example, the TA could toss a coin in class to get the random numbers.

Question 3

The TA would like to encourage everyone in class to prepare before coming to class. To do that, the TA wants to select students who have not prepared to answer questions in class.

The TA decides to select students at random (with replacement) to answer the questions. Assume that there are *n* students and *k* of them have not prepared. What is the expected number of questions required for finding a student who has not prepared?

---

1. n
2. n/k
3. $n/2k$
4. k/n

Question 3

Answer: B

This is a *geometric distribution* with a probability of success of $p = k/n$. The expected value of the number of trials needed to obtain a success is $1/p = n/k$.

Question 4

The TA would like to encourage everyone in class to prepare before coming to class. To do that, the TA wants to select students who have not prepared to answer questions in class.

The TA decides to select students at random to answer questions. However, the TA is worried that not everyone in class would get to participate even though many questions will be asked throughout the semester. Assume that there are n students in the class. What is the expected number of questions that need to be asked before every student has been asked a question?

- $\Theta(\lg n)$
- $\Theta(n)$
- $\Theta(n \lg n)$
- $\Theta(n^2)$

Question 4

Answer: C

This is the coupon collector problem, covered in lecture.

Next few slides from lecture notes.

T_i is the time to collect the i-th coupon after i-1 coupon has been collected.

The prob of collecting a new coupon after i-1 coupon has been collected $p_i = (n-(i-1))/n$.

T_i has a **geometric distribution** with expectation $1/p_i = n/(n-(i-1))$.

Let T_i be the number of draws used to collect the i-th coupon and T be the total number of draws.

$$T = \sum_{i=1}^{n} T_i$$ Total number of draws

$$E[T] = E\left[\sum_{i=1}^{n} T_i\right]$$ Expected value

$$= \sum_{i=1}^{n} E[T_i]$$ Linearity of expectation

$$
\begin{aligned}
E[T] &= \sum_{i=1}^{n} E[T_i] \\
&= \sum_{i=1}^{n} \frac{n}{n-(i-1)} \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} \\
&= n \cdot \left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} \right) \\
&= n \cdot H_n \\
&= O(n \lg n)
\end{aligned}
$$

where H_n is the n-th harmonic number and is of size $O(\lg n)$

Question 5

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

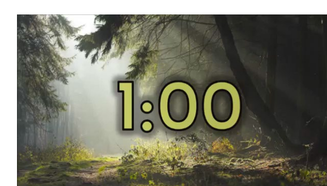
$$\Pr[h(x) = h(y)] \leq \frac{1}{m}$$

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:

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



Question 5: Solution

Let the n elements hashed be $i_1, \ldots, i_n$.

Define $X_{k\ell}$ to be the indicator for the event that $h(i_k) = h(i_\ell)$. By universality, $E[X_{k\ell}] \leq 1/m$. So:

$$E[\#\text{ of }(k,\ell)\text{ s.t. }k > \ell\text{ and }h(i_k) = h(i_\ell)]$$
$$= \sum_{k>\ell} E[X_{k\ell}]$$
$$\leq \binom{n}{2} \cdot \frac{1}{m}$$

Note that number of distinct pairs (k, ℓ) such that $h(i_k) = h(i_\ell)$ is exactly: $\sum_{j \in [m]} \binom{n_j}{2}$. So, $E\left[\sum_{j \in [m]} \binom{n_j}{2}\right] \leq \frac{n^2}{2m} = \alpha n/2$.

Question 5: Solution

Therefore:

$$E\left[\sum_j n_j^2\right] = E\left[\sum_j n_j + 2\binom{n_j}{2}\right] \leq E\left[\sum_j n_j\right] + n\alpha = n(1+\alpha)$$

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 $k \notin S$;
    $S = \{k\} \cup S$;
Report S;

Question 6: Solution

Let X be the number of accesses of Query(A, q)
Let X_j be the number of accesses to obtain the j^{th} male.

$X = X_1 + X_2 + \ldots + X_q$.

For j iteration, we need to find a male from (m-(j-1)) males.
Hence, X_j follows a geometric distribution with probability $\frac{m-j+1}{n}$.
Hence, $E[X_j] = \frac{n}{m-j+1}$.

$$E[X] = E[X_1] + \cdots E[X_q] = \frac{n}{m} + \frac{n}{m-1} + \cdots + \frac{n}{m-q+1}$$
$$= \Theta\left(n \ln \frac{m}{m-q}\right)$$