5CCS2FC2: Foundations of Computing II

Probability & Average Time Complexity

Week 10

Dr Christopher Hampson

Department of Informatics

King's College London

Warm-up : The Birthday Paradox

- What is the **minimum group size** needed so that it is **more likely than not** that two people in the group share a birthday?
- How many **pairs of students** would you **expect** to share a birthday in this class? **(**$\sim$ 287 **students enrolled for FC2)**

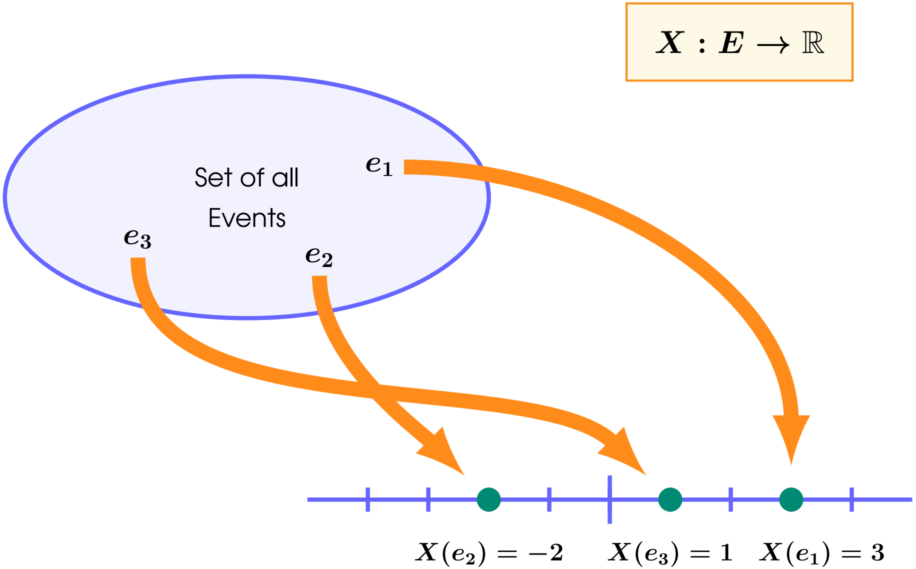
Enter your Birthday below to play along!

https://goo.gl/forms/PIWuTpLTj8kznoSb2

Random Variables

Random Variables

- **Random Variables**
  - A random variable on is a *measurment* of a random event $e \in E$



Random Variables

- **Probability Mass Function (p.m.f.)**
  - The **probability mass function** of a random variable X is a function which says how likely a given value is to appear as the measurement of a random event.

$$\boxed{p_X(k) \;=\; P(\{\text{ all events } e \text{ such that } X_e = k\;\})}$$

Expectation and Variance

Expectation of a Random Variable

- **Expectation**
  - The expectation of a random variable X is the *weighted average* of the possible values of X

$$\mathbf{E}[X] = \sum_{k} k \, p_X(k)$$

(where p_X is the probability mass function for X)

Example Let X be the value thrown on a biased six-sided dice where the probability of rolling a 6 is five times as likely as rolling any other value.

$$\begin{aligned}\mathbf{E}[X] &= \sum_{k=1}^{6} k \times P(\text{rolling value } k) \\ &= 1\left(\frac{1}{10}\right) + 2\left(\frac{1}{10}\right) + 3\left(\frac{1}{10}\right) + 4\left(\frac{1}{10}\right) + 5\left(\frac{1}{10}\right) + 6\left(\frac{5}{10}\right) = \frac{9}{2}\end{aligned}$$

Expectation of a Random Variable

- **Variance**
  - The **variance** of a random variable X is a measure of the *expected deviation* from the average $\mu = \mathbf{E}[X]$ (Greek letter 'mu')

Attempt 1

$$\text{Var}(X) \overset{?}{=} \mathbf{E}[\text{differences between } X \text{ and } \mu]$$
$$= \mathbf{E}[X - \mu] = \text{always zero!} \quad ✗$$

Attempt 2

$$\text{Var}(X) \overset{?}{=} \mathbf{E}[\textit{squared} \text{ differences between } X \text{ and } \mu]$$
$$= \mathbf{E}[(X - \mu)^2]$$
$$= \text{a better measure of deviation} \quad ✓$$

Expectation of a Random Variable

Example

Consider the same biased dice whose expectation is $\mathbf{E}[X] = 9/2$.

k	$p(k)$	$k - \mathbf{E}[X]$	$(k - \mathbf{E}[X])^2$	$(k - \mathbf{E}[X])^2 \times p(k)$
1	$\frac{1}{10}$	$-\frac{7}{2}$	$\frac{49}{4}$	$\frac{49}{40}$
2	$\frac{1}{10}$	$-\frac{5}{2}$	$\frac{25}{4}$	$\frac{25}{40}$
3	$\frac{1}{10}$	$-\frac{3}{2}$	$\frac{9}{4}$	$\frac{9}{40}$
4	$\frac{1}{10}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{40}$
5	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{9}{40}$
6	$\frac{5}{10}$	$\frac{3}{2}$	$\frac{9}{4}$	$\frac{35}{40}$

$$\Longrightarrow \quad \mathrm{Var}(X) = \frac{128}{40} = 3.2$$

Properties of Expectation and Variance

Theorem Let X and Y be any two random variable and let $a, b \in \mathbb{R}$ be real numbers. Then:

(i) $\mathrm{E}[aX + b] \;=\; a\,\mathrm{E}[X] + b$,

(ii) $\mathrm{E}[X + Y] \;=\; \mathrm{E}[X] + \mathrm{E}[Y]$,

(iii) $\mathrm{Var}(aX + b) \;=\; a^2\,\mathrm{Var}(X)$,

(iv) $\mathrm{Var}(X + Y) \;=\; \mathrm{Var}(X) + \mathrm{Var}(Y)$,

(if X and Y are independent)

(v) $\mathrm{Var}(X) \;=\; \mathrm{E}[X^2] - \mathrm{E}[X]^2$.

Proof:

- Not given in this course!

Q.E.D

Bernoulli Distribution

- **Bernoulli Random Variable**
  - A random variable is said to be **Bernoulli distributed** if:
    - it only takes the values **1 (true)** or **0 (false)**
    - Has the following probability mass function (p.m.f)

$$p_X(1) = p \quad \text{and} \quad p_X(0) = (1 - p)$$

(in which case we write $X \sim \text{Ber}(p)$)

Theorem Let $X \sim \textbf{Ber}(p)$ be Bernoulli distributed. Then we have that

$$\mathrm{E}[X] = p \quad \text{and} \quad \textbf{Var}(X) = p(1 - p)$$

Average-Time Complexity

Average-Time Complexity

- **Worst-Case Time Complexity**
  - The **(worst-case) time complexity** of an algorithm is the *maximum* time that the algorithm requires to run on *any* input of size n.
    i.e. an upper bound on $T(n)$

$$\boxed{T(n) \;=\; \text{number of steps for input of size } n}$$

- **Average-Time Complexity**
  - The **average-time complexity** of an algorithm is the *expectation* of the (random) variable $T(n)$, which may vary randomly for different inputs of size n,

$$\boxed{\text{average-time complexity} \;=\; \mathbf{E}[\,T(n)\,]}$$

The Bucket Sort Algorithm

- The **Bucket Sort algorithm** can be used to arrays whose items are **uniformly distributed** throughout a given interval (usually the unit interval [0, 1]).

The Bucket Sort Algorithm

Step 1) For each $k = 1, \ldots, n$, create an empty bucket $B_1, \ldots B_n$,

Step 2) For each item $\mathbf{item}_k$ in the list, for $k = 1, \ldots, n$:

- If $\mathbf{item}_k$ is between $(i-1)/n$ and i/n, then put $\mathbf{item}_k$ into bucket B_i.

Step 3) For each bucket $B_1, \ldots, B_n$:

- Sort B_i using *(naïve) insertion sort*.

Step 4 Concatenate each of the sorted bucket lists.

The Bucket Sort Algorithm

Theorem The *worst-time* complexity of the Bucket Sort algorithm is quadratic $O(n^2)$ — bad!.

Proof:

Step 1) Suppose we have n items whose values are all $< 1/n$.

Step 2) Every item gets placed into the first bucket B_1 in **Step 2**.

Step 3) The insertion sort on B_1 in **Step 3** takes $\Theta(n^2)$ steps.

$$T(n) = \underbrace{\Theta(n)}_{\text{Step 1}} + \underbrace{\Theta(n)}_{\text{Step 2}} + \underbrace{\Theta(n^2)}_{\text{Step 3}} + \underbrace{\Theta(n)}_{\text{Step 4}} = \Theta(n^2)$$

Q.E.D

The Bucket Sort Algorithm

Theorem The *average-time* complexity of the Bucket Sort algorithm is linear $O(n)$.

Proof:

Step 1) For a given (uniformly distributed) array of size n, we have that

$$T(n) = \underbrace{\Theta(n)}_{\text{Step 1}} + \underbrace{\Theta(n)}_{\text{Step 2}} + \underbrace{O(B_1^2) + \cdots + O(B_n^2)}_{\text{Step 3}} + \underbrace{\Theta(n)}_{\text{Step 4}}$$

(where B_i is a random variable denoting the number of items in bucket i)

The Bucket Sort Algorithm

Step 2) The **expected** termination time is, therefore, given by

$$
\begin{aligned}
\mathbf{E}[T(n)] &= \mathbf{E}[\,\Theta(n) + O(B_1^2) + \cdots + O(B_n^2)\,] \\
&= \Theta(n) \;+\; \mathbf{E}[\,O(B_1^2)\,] + \cdots + \mathbf{E}[\,O(B_1^2)\,] \\
&= \Theta(n) \;+\; O\big(\,\mathbf{E}[B_1^2]\,\big) + \cdots + O\big(\,\mathbf{E}[B_n^2]\,\big)
\end{aligned}
$$

(since $\mathrm{E}[\cdot]$ is linear are $\Theta(n)$ is not a constant / non-random)

Step 3) Let X_{ij} be the following **Bernoulli random variable**

$$
X_{ij} \;=\; \text{Item } j \text{ is placed into the } i\text{th Bucket}
$$

(so that $B_i = X_{i1} + X_{i2} + \cdots + X_{in}$ for each $i = 1, \ldots, n$)

The Bucket Sort Algorithm

Step 4) Suppose we have only **two buckets**, then

$$
\begin{aligned}
\mathbf{E}[B_i^2] &= \mathbf{E}[\ (X_{i1} + X_{i2})(X_{i1} + X_{i2})\ ] \\
&= \mathbf{E}[\ X_{i1}^2 \ + \ X_{i1}X_{i2} \ + \ X_{i2}X_{i1} \ + \ X_{i2}^2\ ] \\
&= \mathbf{E}[X_{i1}^2] \ + \ \mathbf{E}[X_{i2}^2] \ + \ \mathbf{E}[X_{i2}X_{i1}] \ + \ \mathbf{E}[X_{i2}X_{i1}]
\end{aligned}
$$

or the same calculation in general for more buckets

$$
\begin{aligned}
\mathbf{E}[B_i^2] &= \mathbf{E}[\ (X_{i1} + \cdots + X_{in})(X_{i1} + \cdots + X_{in})\ ] \\
&= \sum_{j=1}^{n} \mathbf{E}[X_{ij}^2] \ + \ 2\sum_{j=1}^{n}\sum_{k=1}^{j-1} \mathbf{E}[X_{ij}X_{ik}]
\end{aligned}
$$

The Bucket Sort Algorithm

Step 5) We can then use the fact that

$$\mathbf{E}[X_{ij}^2] = \frac{1}{n}$$

and

$$\mathbf{E}[X_{ij}X_{ik}] = \frac{1}{n^2}$$

to find that

$$\begin{aligned}
\mathbf{E}[B_i^2] &= \sum_{j=1}^{n}\left(\frac{1}{n}\right) + 2\sum_{j=1}^{n}\sum_{k=1}^{j-1}\left(\frac{1}{n^2}\right) \\
&= n\left(\frac{1}{n}\right) + n(n-1)\left(\frac{1}{n^2}\right) \\
&= 1 + \frac{n-1}{n} \quad = \quad 2 - \frac{1}{n}
\end{aligned}$$

The Bucket Sort Algorithm

Step 6) Hence we have that

$$
\begin{aligned}
\mathbf{E}[T(n)] &= \Theta(n) + O(\mathbf{E}[B_1^2]) + \ldots + O(\mathbf{E}[B_n^2]) \\
&= \Theta(n) + O\left(2 - \frac{1}{n}\right) + \cdots + O\left(2 - \frac{1}{n}\right) \\
&= \Theta(n) + O(2n - 1) \quad = \quad O(n)
\end{aligned}
$$

Q.E.D

The reason for this linear average-time upper bound, is that we have assumed something about additional about the input. Namely that the items are **uniformly distributed** in the unit interval. Otherwise we cannot improve upon the $O(n \log n)$ upper bound.

End of Module!