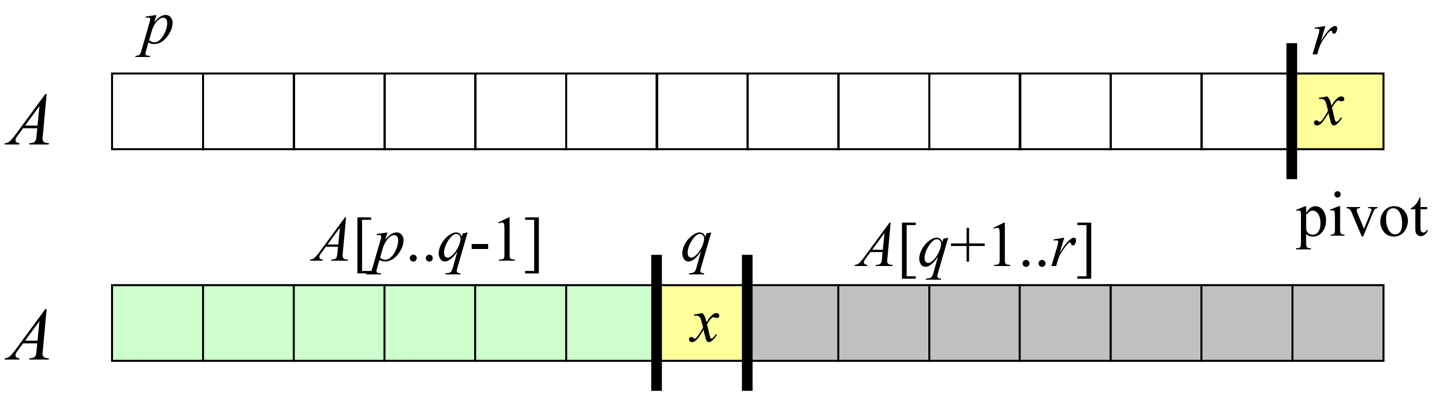
Quicksort

7.1 Quicksort

***Quicksort*(*A*[*p*..*r*])**

Divide: 把 $A[p..r]$ 分成 $A[p..q\text{-}1]$ 和 $A[q+1..r]$



Conquer: 遞迴將 $A[p..q\text{-}1]$ 和 $A[q+1..r]$ 排序

Combine: 不需要作任何事

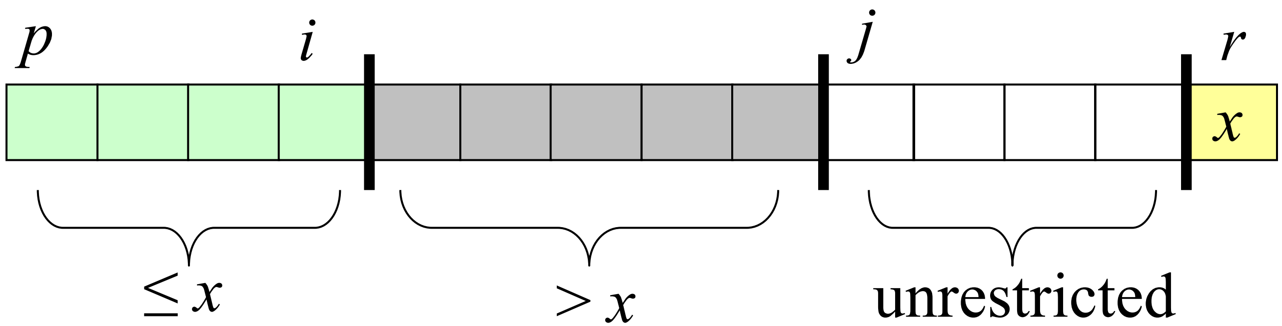
Quicksort(A, p, r)

```
If p < r then
  q ← Partition(A, p, r)      /* divide */
  Quicksort(A, p, q-1)        /* conquer */
  Quicksort(A, q+1, r)        /* conquer */
```

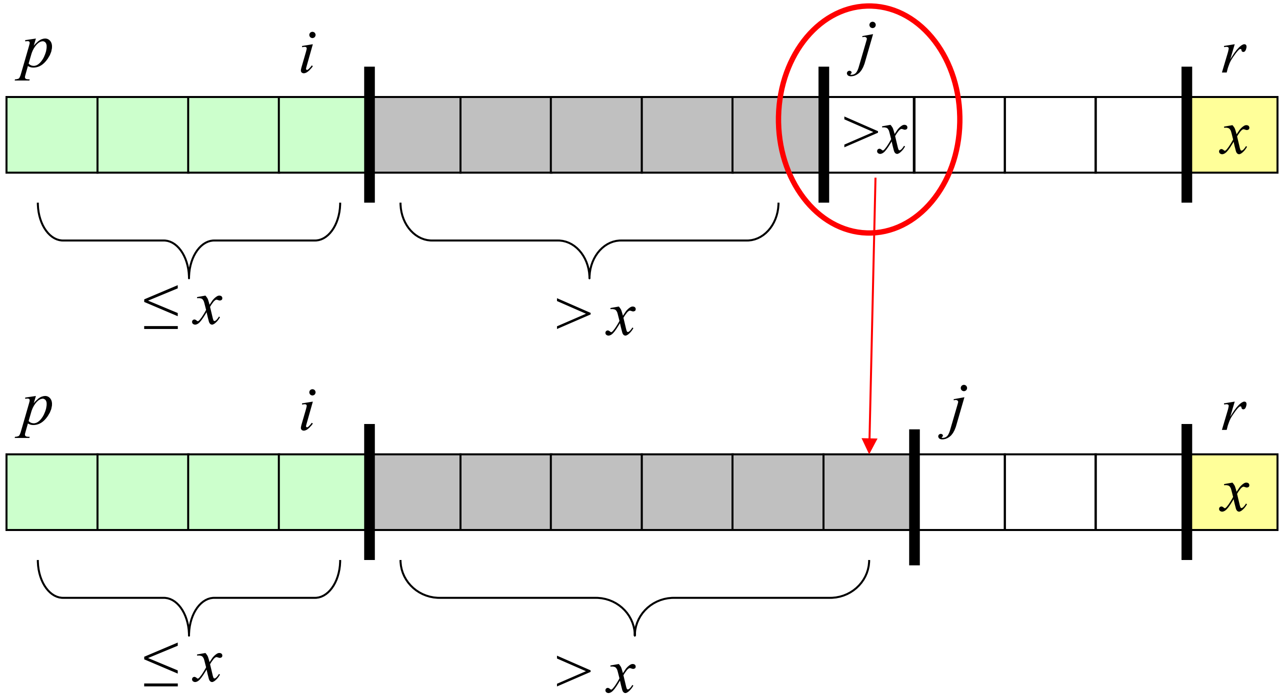
Partition(A, p, r)

```
x ← A[r]
i ← p-1
for j← p to r-1
  do if A[j] ≤ x
  then i ← i+1
        exchange A[i]↔A[j]
exchange A[i+1]↔A[r]
return i+1
```

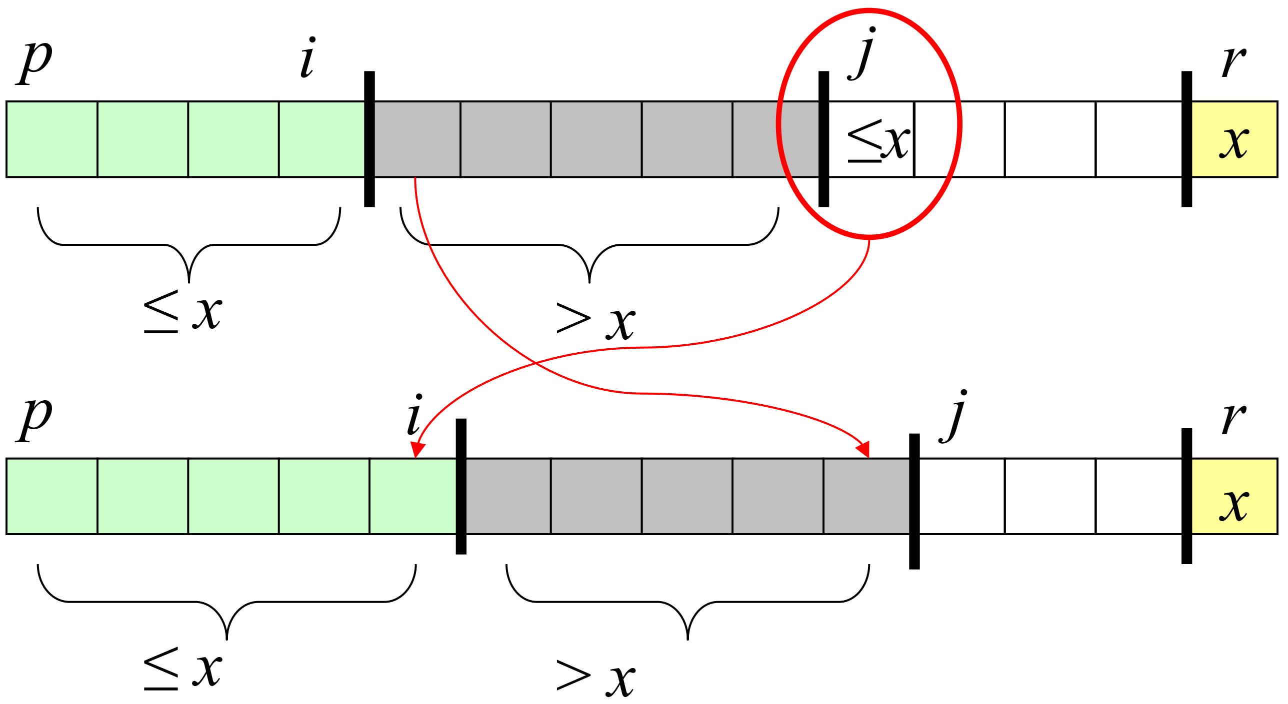
i 和 j 的意義:



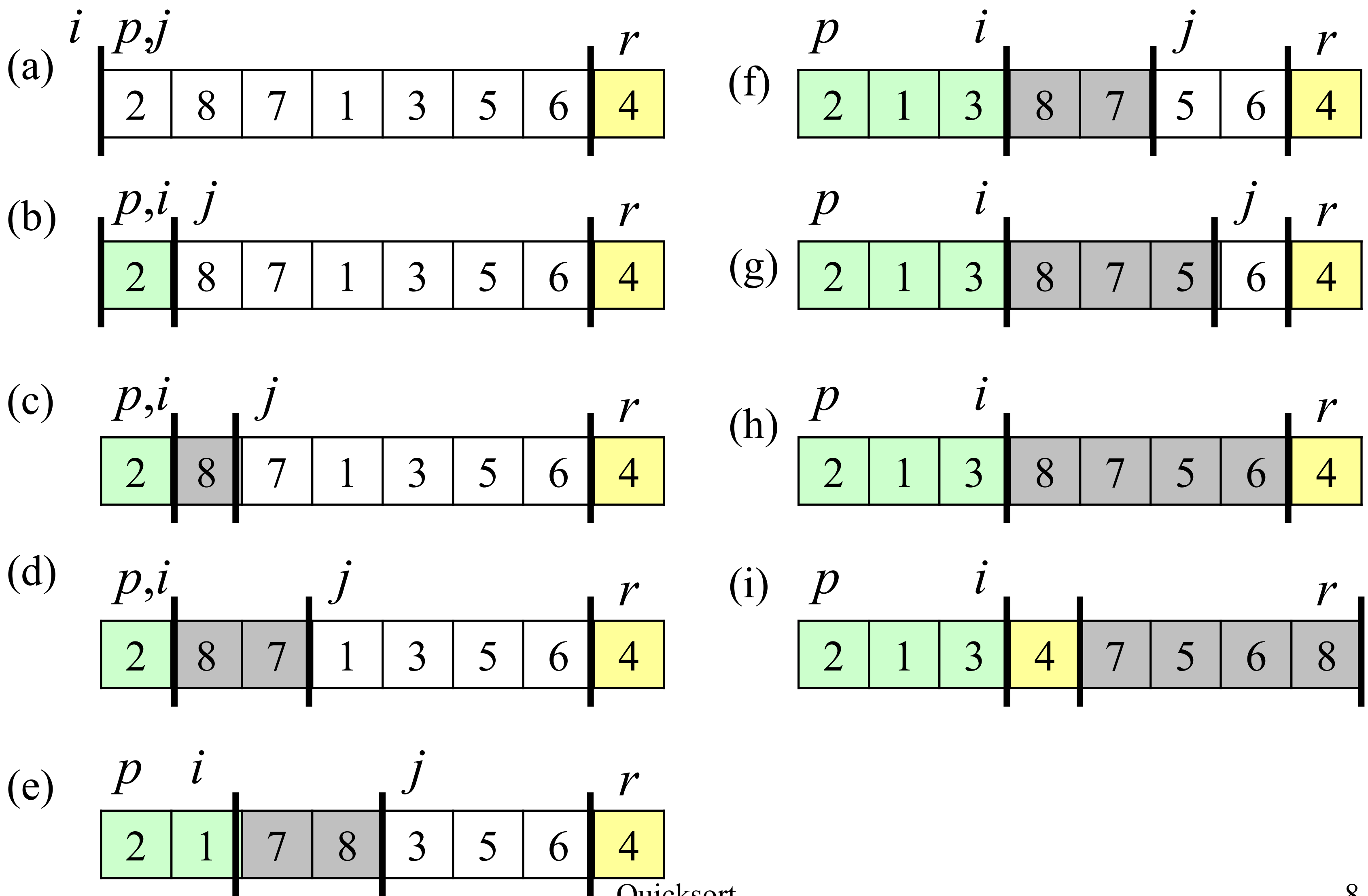
i 和 j 如何改變:



i 和 j 如何改變:



範例: (Partition, $x=A[r]=4$)



7.2 分析

Worst-case: $\Theta(n^2)$ (對於已排序好的輸入)

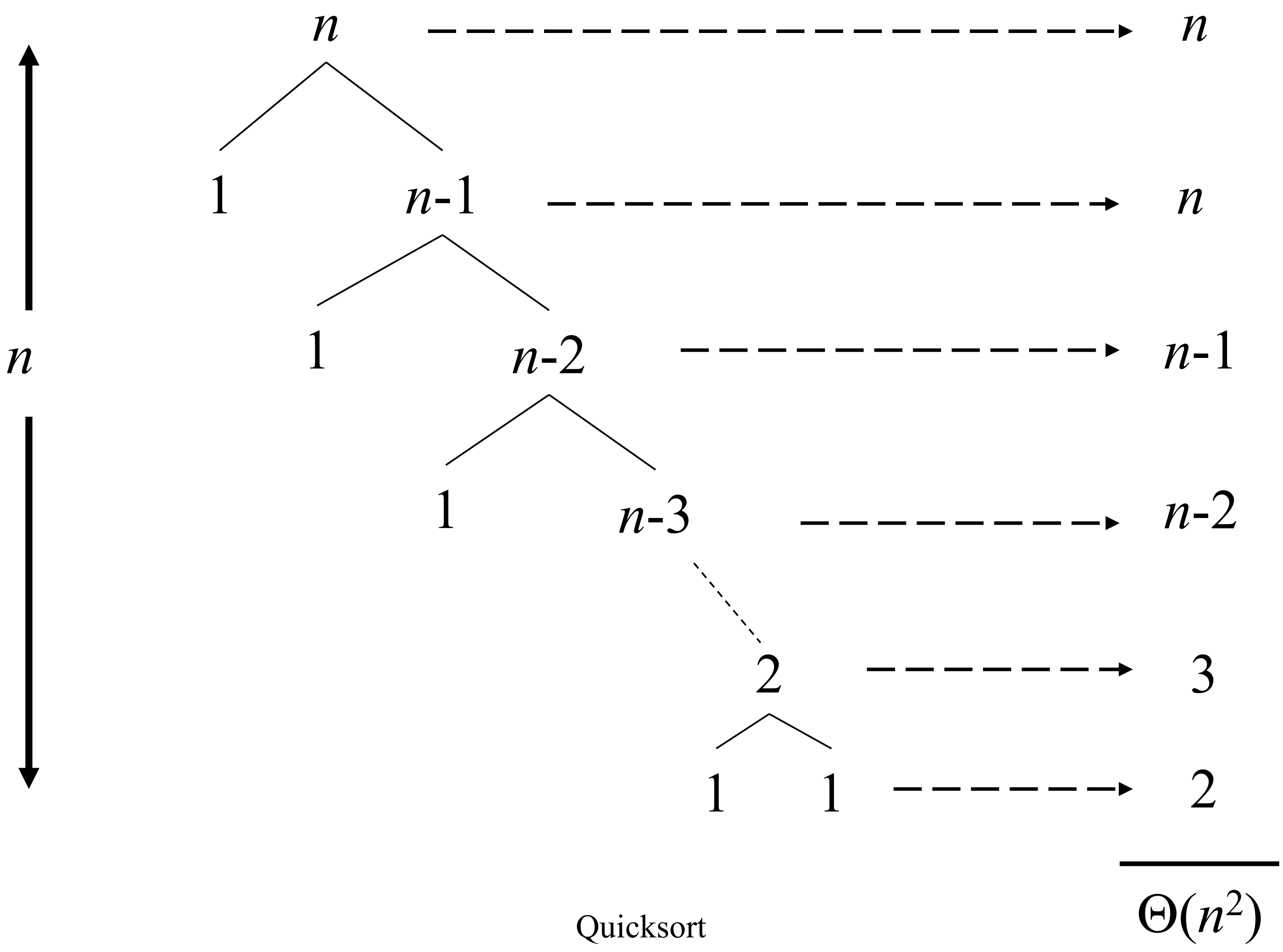
$$\mathrm{T(n)} = \max_{1 \le q \le n}\{T(q-1)+T(n-q)\}+\Theta(n)$$
$$= \max_{0 \le k \le n-1}\{T(k)+T(n-k-1)\}+\Theta(n)$$

猜測: $T(n) \le c\, n^2 = O(n^2)$

Substituting:

$$T(n) \le \max_{0 \le k \le n-1}\{ck^2+c(n-k-1)^2\}+\Theta(n)$$
$$\le c \max_{0 \le k \le n-1}\{k^2+(n-k-1)^2\}+\Theta(n)$$
$$\le c(n-1)^2+\Theta(n)$$
$$\le cn^2-c(2n-1)+\Theta(n)$$
$$\le cn^2 \text{ (挑選夠大的 } c \text{ 即可)}$$

- $T(n)=\Theta(n^2)$



Average-case: $\Theta(n \lg n)$

(假設所有的元素都不相同)

- $T(n)=O(n+X)$，X 是 Partition 中第四行的執行次數。
- 每次呼叫 Partition 的時候，如果 $A[i]<x<A[j]$ 或 $A[j]<x<A[i]$，$A[i]$ 和 $A[j]$ 將來就不會再互相比較。

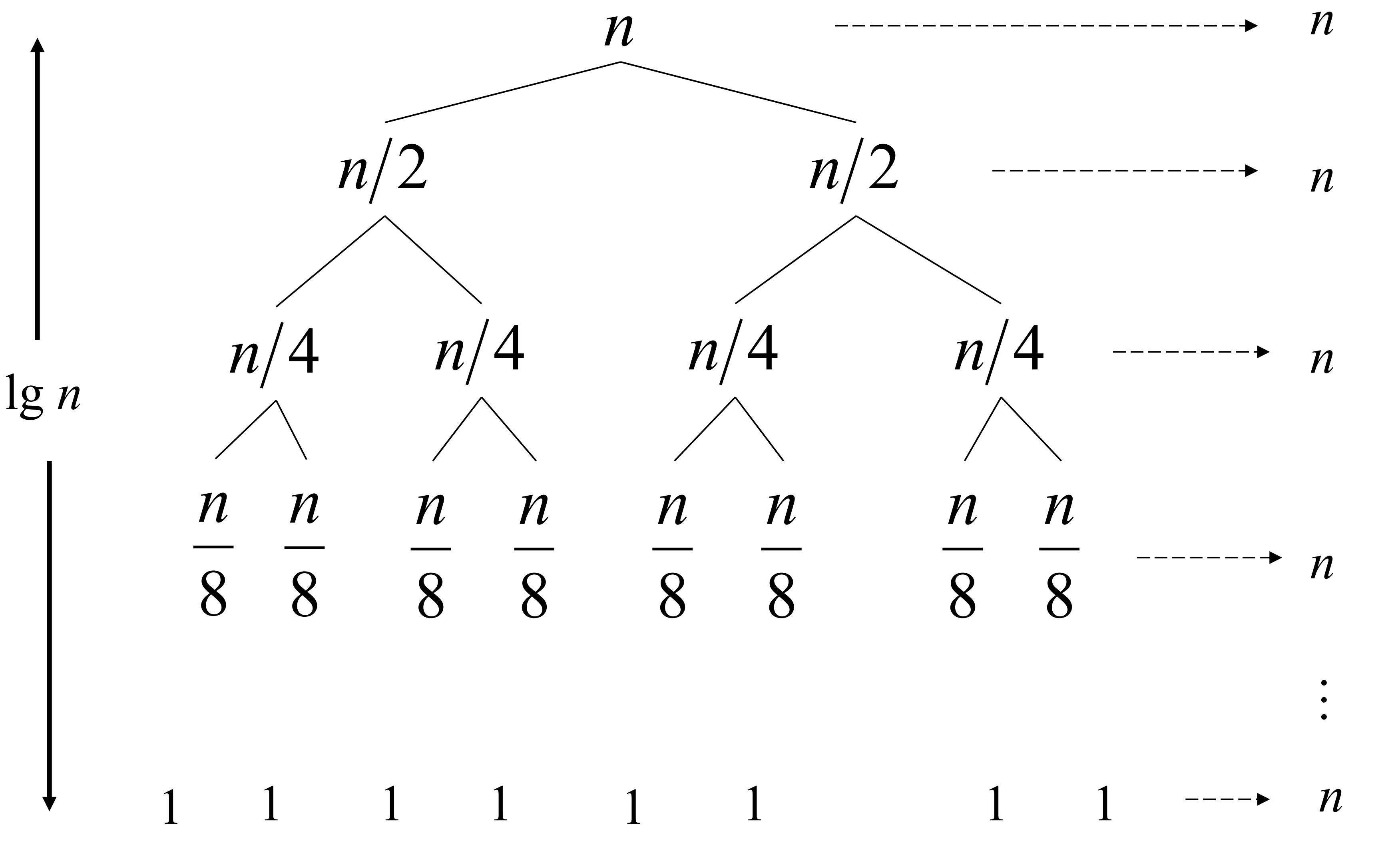
範例: 令 $A=\{3，9，2，7，5\}$。 第一個回合後，$A=\{3，2，5，9，7\}$。 之後 $\{3，2\}$ 再也不會和 $\{9，7\}$ 比較了。

- 將 A 的元素重新命名為 z_1，z_2，...，z_n，其中 z_i 是第 i 小的元素。且定義 $Z_{ij}=\{z_i，z_{i+1}，...，z_j\}$。

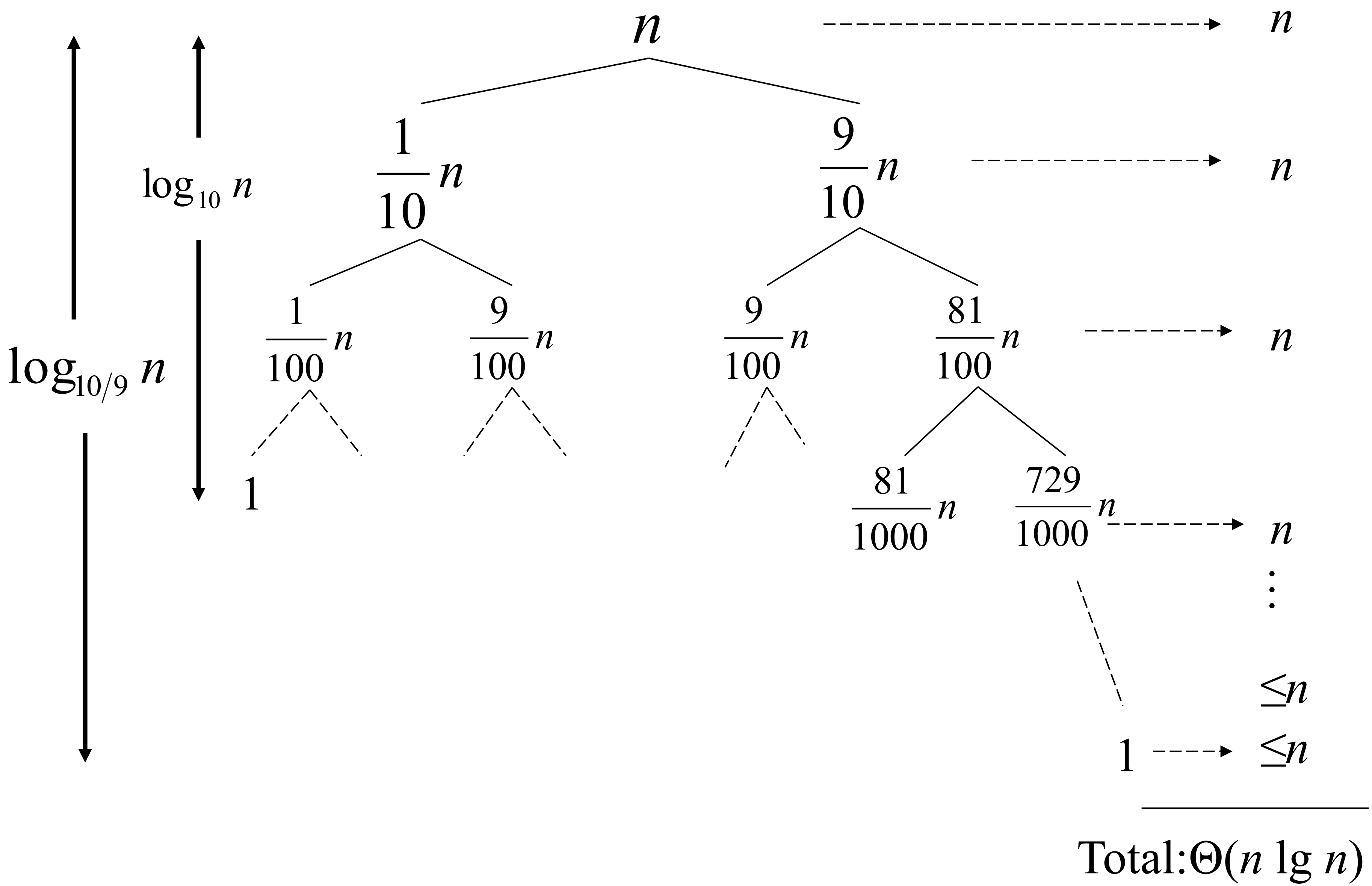
- 定義$z_i : z_j$ ：若且唯若第一個從 Z_{ij} 選出來的 pivot 是 z_i 或 z_j。

- 對於任意的 i 和 j，發生 $z_i : z_j$ 的機率為 $2/(j-i+1)$，因此，

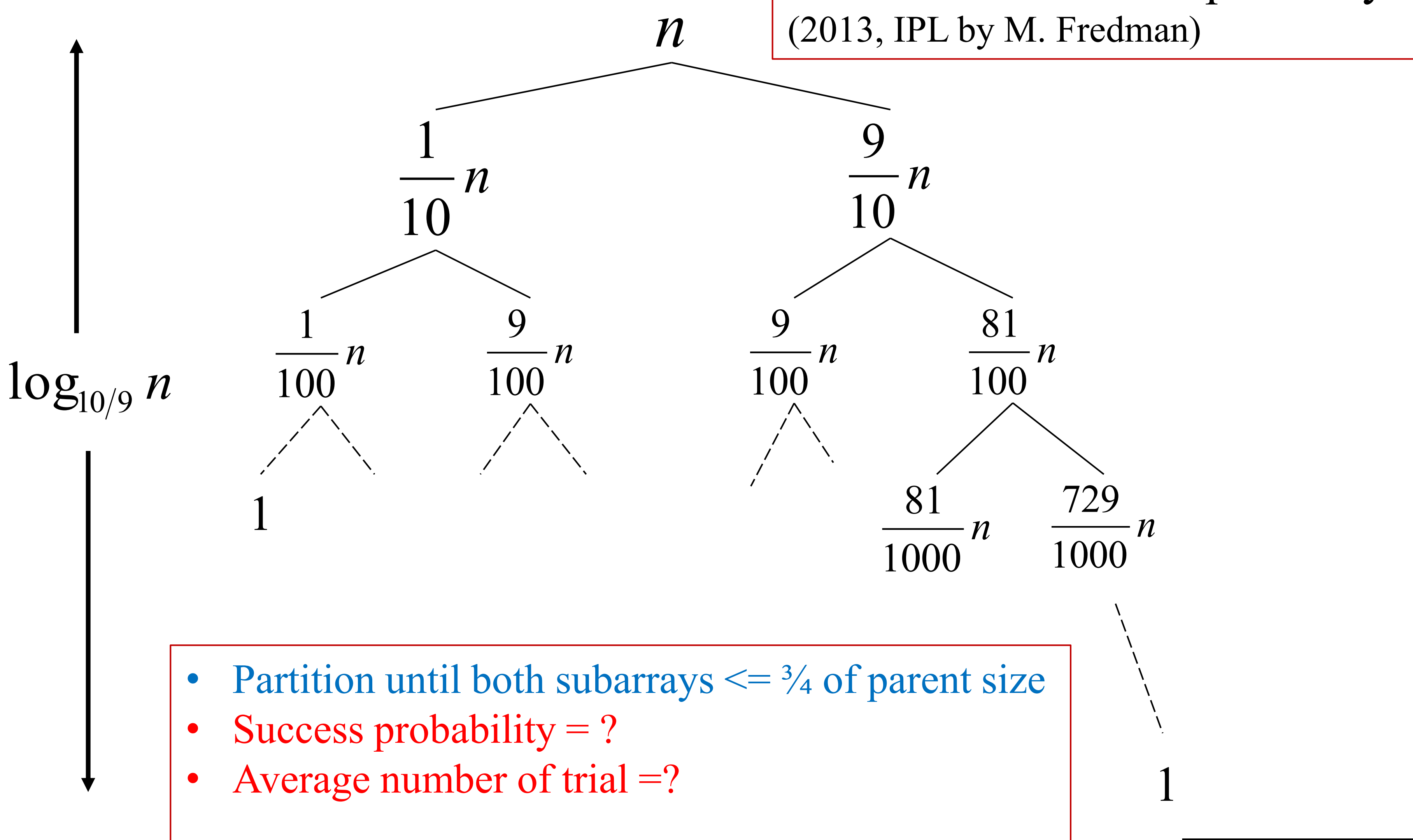
$$
\begin{aligned}
X &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&= \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \frac{2}{k+1} \\
&< \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \qquad \text{(套用 Harmonic Series)} \\
&= \sum_{i=1}^{n-1} O(\lg n) \\
&= O(n \lg n)
\end{aligned}
$$



Total: $\Theta(n \lg n)$


n
$\frac{1}{10}n$
$\frac{9}{10}n$
$\log_{10} n$
$\frac{1}{100}n$
$\frac{9}{100}n$
$\frac{9}{100}n$
$\frac{81}{100}n$
$\log_{10/9} n$
1
$\frac{81}{1000}n$
$\frac{729}{1000}n$
1
n
n
n
n
$\leq n$
$\leq n$
Total: $\Theta(n \lg n)$

An Intuitive and Simple analysis:
(2013, IPL by M. Fredman)



- Partition until both subarrays <= ¾ of parent size
- Success probability = ?
- Average number of trial =?

Avg: $\Theta(n \lg n)$

L_Quicksort(A, p, r) /* Lazy Quicksort */

```
If p < r then
Repeat i ← Random(p, r)
    exchange(A[r], A[i])
    q ← Partition(A, p, r) /* divide */
Until (r-p)/4 <= (q - p) <= 3(r-p)/4
L_Quicksort(A, p, q-1)        /* conquer */
L_Quicksort(A, q+1, r)        /* conquer */
```

- 其他分析

$$E(n) = (n-1) + \frac{1}{n}\sum_{q=1}^{n}\{E(q-1)+E(n-q)\}$$
$$= (n-1) + \frac{2}{n}\sum_{k=1}^{n-1}E(k)$$

為了簡單起見，假設

$$E(n) = n+1 \quad + \quad \frac{2}{n}\sum_{k=1}^{n-1}E(k)$$

$$\Rightarrow nE(n) = n^2+n \quad + \quad 2\sum_{k=1}^{n-1}E(k) \qquad \text{-------}(1)$$

$$\Rightarrow (n\text{-}1)E(n\text{-}1) = (n-1)^2+(n-1)+2\sum_{k=1}^{n-2}E(k) \qquad \text{-------}(2)$$

(用 n-1 替換掉 (1) 裡面的 n)

(1)-(2), 可得

$$nE(n) = (n+1)E(n-1) + 2n$$

$$\Rightarrow E(n) = \frac{n+1}{n}E(n-1) + 2$$ (套用 iteration method)

$$= \frac{n+1}{n}\left\{\frac{n}{n-1}E(n-2) + 2\right\} + 2 = \frac{n+1}{n-1}E(n-2) + 2\frac{n+1}{n} + 2$$

$$= \frac{n+1}{n-2}E(n-3) + 2\frac{n+1}{n-1} + 2\frac{n+1}{n} + 2 = \bullet \quad \bullet \quad \bullet \quad \bullet$$

$$= \frac{n+1}{2}E(1) + 2(n+1)\left(\frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n}\right) + 2 = \Theta(n) + \Theta(n)\sum_{k=3}^{n}\frac{1}{k} + 2$$

$= \Theta(n) + \Theta(n) \times \Theta(\lg n) + 2$ (套用 Harmonic Series)

$= \Theta(n \lg n)$

7.3 Randomized version of quicksort

Randomized Algorithm:

使用*亂數產生器*的演算法。

Pseudorandom-number generator:

一個傳回在統計上看似隨機數字的 deterministic algorithm 。

```
Randomized-Partition(A, p, r)
  i ← Random(p, r);
  exchange(A[r], A[i]);
  return Partition(A, p, r)
```

Exercises

Problem 1:

企業喜歡有個好記的電話號碼。用一個好記的詞或片語來拼電話號碼是一個方法。例如，您撥打 TUT-GLOP 就能打到滑鐵盧大學。有的時候號碼中只有一部分的數字被拿來拼字。當您回到您的旅館時，您能撥打 310-GINO 到 Gino’s 點一個披薩。另一個讓電話號碼好記的方法是把數字編組。例如您能撥打”三個十”3-10-10-10 到 Pizza Hut 點您的披薩。

電話號碼的標準格式由七個十進位的數字所構成，其中有一個連字號在三和第四個數字之間(例如：888-1200)。電話的按鍵提供了字母與數字的對應，如下所示：

Exercises

A、 B 和 C 對應到 2； D、 E 和 F 對應到 3
G、 H 和 I 對應到 4； J、 K 和 L 對應到 5
M、 N 和 O 對應到 6； P、 R 和 S 對應到 7
T、 U 和 V 對應到 8； W、 X 和 Y 對應到 9

Q 和 Z 沒有對應的數字。連字號不用撥，而且可以視情況增加或刪除。TUT-GLOP 的標準格式是 888-4567，310-GINO 的標準格式是 310-4466，而 3-10-10-10 的是 310-1010。

二個電話號碼如果有同樣標準格式表示他們是相同的。(他們撥同樣的數字。)您的公司正要整理地方企業的電話號碼。為了控制品質，您想要檢查有沒有兩家(或更多)的企業有同樣電話號碼。

Exercises

輸入：第一行包含了一個整數，代表總共有幾筆資料。隨後會接著一個空行。之後的第一行是一個正整數(最大到100000)，代表要處理的電話號碼的數目。接下來的每一行都有一組電話號碼，由十進位的數字、大寫字母(Q 和 Z 除外)以及連字號所構成。其中剛好有七個字元是字母或數字。每筆資料中間都有一個空行。

輸出：列出所有出現兩次以上的電話號碼。每一行必須是電話號碼的標準格式，接著是該電話號碼出現的次數(兩者用一個空白隔開)。這些電話號碼必須要由小到大排好。如果沒有電話號碼是重複的，那就輸出一行：

No duplicate.

在兩筆資料之間要印一個空行。

以下是一個輸出入的實例:

Sample Input	Sample Output
1 12 4873279 ITS-EASY 888-4567 3-10-10-10 888-GLOP TUT-GLOP 967-11-11 310-GINO F101010 888-1200 -4-8-7-3-2-7-9- 487-3279	310-1010 2 487-3279 4 888-4567 3

Exercises

Problem 2:

某家公司正在尋求一個簡單的方法來編寫員工的資訊。目前的做法是各個部門分別列出自己的員工，然後再統一把資料集中送到公司負責人那邊，負責人再將蒐集好的名單根據姓氏排序。但是人力太花時間了，所以負責人希望能有一個程式幫忙做這件事情：輸入所有員工的資料，合併整理之後再輸出。

Exercises

輸入：輸入的第一行包含了公司部門的數目(介於 2 和 12 之間)，接下來會有各個部門的員工資料。每一個部門提供的資料中，第一行是部門的名稱，之後的每一行都有一個員工的資料(直到空行或是檔案結尾為止)。資料格式如下：

稱謂, 名字, 姓氏, 地址, 家中電話, 辦公電話, 校園信箱號碼（用逗號隔開）

最多只有 20000 個員工。

輸出：以下是輸出的格式：(請參閱輸出實例)

```
----------------------
<稱謂> <名字> <姓氏>
<地址>
Department: <部門名稱>
Home Phone: <家中電話>
Work Phone: <辦公電話>
Campus Box: <校園信箱號碼>
```

以下是一個輸出入的實例:

<table>
<tr><th>Sample Input</th><th>Sample Output</th></tr>
<tr><td><pre>
2
English Department
Dr.,Tom,Davis,Anystreet USA,555-2832,555-2423,823
Mrs.,Jessica,Lembeck,Center Street,555-2543,555-8584,928

Computer Science
Mr.,John,Euler,East Pleasure,555-1432,555-2343,126
</pre></td><td><pre>

Dr. Tom Davis
Anystreet USA
Department: English Department
Home Phone: 555-2832
Work Phone: 555-2423
Campus Box: 823

Mr. John Euler
East Pleasure
Department: Computer Science
Home Phone: 555-1432
Work Phone: 555-2343
Campus Box: 126

Mrs. Jessica
Lembeck Center Street
Department: English Department
Home Phone: 555-2543
Work Phone: 555-8584
Campus Box: 928
</pre></td></tr>
</table>