104下 群試理論

檢視課程綱要 | 編輯課程目標及評量 | 編輯每週進度表 | 編輯單元分配時數表(optional) | 結束編輯

※系統會在您離開編輯框後，自動儲存您編輯的內容。

下載課程綱要Word檔

課程名稱：（中文）群試理論

（英文）Group Testing

開課單位：應數碩

當期課號：5406

永久課號：IAM5813

授課教師：傅恆霖

學分數：3

必/選修：選修

先修科目或先備能力：

Linear Algebra and Discrete Mathematics

課程概述與目標：

Group testing is a special type of search methodology which has various applications in modern technologies such as DNA sequencing, Drug test, Safe communication, Cryptology, Data pattern discovery and Compressed sensing, etc. This course intends to use combinatorial theory to learn more inside of this topic.

教科書（請註明書名、作者、出版社、出版年等資訊）：

Reference Books (by Ding-Zhu Du and Frank K. Hwang)
1. Combinatorial group testing and its applications, and
2.Pooling designs and nonadaptive group testing.

課程大綱		分配時數				備註
單元主題	內容綱要	講授	示範	習作	其他	

教學要點概述

1.學期作業、考試、評量：

There are two term papers to work on, each is of weight 50%.

2.教學方法及教學相關配合事項(如助教、網站或圖書及資料庫等)：

Mainly, I shall use my lecture notes to present in the classes. Students are welcome to share their own opinions, observation and research outcomes.

師生晤談	By appointment. Weekdays are available besides class hours.	SA 200 or classrooms of Science One building.	email address: hlfu@math.nctu.edu.tw or call 0912580590.

每週進度表

週次	上課日期	課程進度、內容、主題

※ 請同學遵守智慧財產權觀念及勿使用非法影印教科書。

備註：

其他欄包含參訪、專題演講等活動。

請同學遵守智慧財產權觀念及勿使用非法影印教科書。

[Top]

上課時間：3B, 4GH

教室：SA 212

2016, 2, 17~18
Week 1 ①

History of Group Testing

大部份研究學者都認為這个概念起源於二次大戰美國為了篩檢梅毒(徵兵)而想出來的一种省錢的方法。

1942~1943

Robert Dorfman and David Rosenblatt

回憶錄中提到，想法来自"?"(碟象)，他把這個想法作進一步"formulate"。

Someone

回憶中說是他想到的

P. sera → syphilitic antigen

單数 Serum 血清　抗原

① If one or more of the sera in the pool contain syphilitic antigen, then a test may be wasted.

② But, if no one in the pool contains syphilitic antigen, then many tests are saved.

Believe or not, this idea of grouping blood samples for syphilis screening was not actually put to use.

也許在某些的疾病篩檢(如AIDS)，會有用！

R. Dorfman, The detection of defective members of large populations, Ann. Math. Statist. 14(1943), 436-440.
(目前引用的論文)

C. H. Li, A sequential method of screening experimental variables, J. Amer. Statist. Assoc. 57 (1962) 455-477.

↓

第一個研究 Combinatorial Group Testing (CGT)

Katona：[4] Combinatorial search problem
(A survey of combinatorial theory)

Consider a set I of n items known to contain exactly d defectives.

(d, n)-problem

(也稱為是 hypergeometric group testing problem)

↓

如果每次都拿 k 個 items 去來測試，它含有 x 個 defectives 的分佈為

hypergeometric distribution → $\dfrac{\binom{d}{x}\cdot\binom{n-d}{k-x}}{\binom{n}{k}}$, probability Group Testing

上面的结果也说明了为什麼 Dorfman 的结果会引起在

Feller 所写的書中： An Introduction to Probability Theory and Its Applications, Vol. 1. John Wiley, NY 1950.

（一个作業）

(*) 當時並不覺得這是一个什麼重要的理論。

接下来相继有些结果让這個概念更加成熟：

A. Sterrett, On the detection of defective members of large populations, Ann. Math. Statist. 28 (1957) (Note 1033-1036).

M. Sobel and P. A Groll, Group testing to eliminate all defectives in a binomial sample, Bell System Tech. J. 28 1959, 1179-1252 (74页)

（测试 Devices 是很痛苦）

condensers
resistors
聖誕樹

Let S be the sample space of a group testing problem.

$M(d,n) \longrightarrow M(S)$

$M_A(d,n) \longrightarrow M_A(S)$

$\log x \longrightarrow \log_2 x$

Fact 1 $M(S) \geq \lceil \log |S| \rceil$.

Fact 2 $S \subseteq S' \Rightarrow M(S) \leq M(S')$

Fact 3 $M(d,n_1) \geq M(d,n_2)$ if $n_1 \geq n_2$.

Fact 4 $M(1,n) = \lceil \log n \rceil$. ✓

Fact 5 $M(2,n) \leq 2\lceil \log n \rceil$.

Fact 6 $M(d,n) \leq d\lceil \log n \rceil$.

Problem $M(2,n) = ?$

如果每一种钢板都有"无穷多个"，有没有好的演算法可以

無意義的 Tests 要避免

例:

Sequential: 知道 {a,b,c} 為 negative 就不要再 test 任何它的 proper subset!

Nonadaptive: {a,b}, {a}, {b} 沒有同時測試的必要, 因為選兩個就可以決定; 如果已知有一個 defective, 則選一個 test 即可, 例如 {a}.

Minimize the number of tests (reasonable)

Let A be a "reasonable" algorithm. We use $M(d,n)$ to denote the minimum number of tests conducted by "any" reasonable algorithm to determine all the defectives, i.e.,

$$M(d,n) = \min_{A} M_A(d,n)$$ over all reasonable algorithm A.

$$M(d,n) = \min_{A} M_A(d,n).$$

Two general types of algorithms:

Sequential: the tests are conducted one by one, and the outcomes of previous tests are assumed "known" at the time of determining the current test.

(Previous infos are helpful!)

Non-adaptive: all tests are specified simultaneously.

(No previous infos can be referred.)

Revised algorithms (Multistage algorithms): the tests are divided into several stages where the stages are considered sequential but all tests in the same stage are treated nonadaptive.

If the number of stages is "s", then it's an s-stage algorithm.

(*) An algorithm is reasonable if it contains no test whose outcome can be predicted from outcomes of ~~the~~ other tests either conducted previously or

A sequential algorithm for group testing is indeed an algorithm. In general, we assume that all the outcomes are error-free. If one or more outcomes are "error outcomes", then the algorithm for finding all the defectives is very difficult to attend "optimum", i.e., to decide $M(d,n)$.

On the other hand, a non-adaptive algorithm can be represented by a matrix (or say array). If the number of tests is t, then we have a $t \times n$ matrix ($(0,1)$-matrix) to represent the algorithm. (in general form)

Let the matrix be $M_{t\times n} = [m_{i,j}]$, then $m_{i,j}=1$ if the j^{th} item is in the i^{th} group (of tests), and $m_{i,j}=0$ otherwise.

$$\begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array}\begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5^{*} & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}_{3\times 7} \end{array}$$

For convenience, we write $t_1 = \{4,5,6,7\}$, $t_2 = \{2,3,6,7\}$ and $t_3 = \{1,3,5,7\}$. This matrix represents a non-adaptive algorithm which uses 3 tests on 7 items.

Now if 5 is a defective item, then $Q(t_1)=1$, $Q(t_2)=0$,

	1	2	3	4	5	6	7	8
0	1	2	3	4	5	6	7	
0	0	0	0	1	1	1	1	
0	0	1	1	0	0	1	1	
0	1	0	1	0	1	0	1	

4	5	6	7

2	3	6	7

1	3	5	7

用磅秤 a b c

重

- a=b=c (0)
- a=b≠c 6
- a=c≠b 5
- b=c≠a 3

重

- a=b≠c 1
- a=c≠b 2
- b=c≠a 4

輕

8個銅幣

① 有一个重量不一樣 (找出一個)

② 有一个ee輕(重)(或輕) (找出一個)

③ 有一个重1g,一个輕1g (找出那兩個)

正常的是10g

磅秤 天平秤

研究的大事

何者較好？

那一种比較合乎 GT 的原理？

Definitions

Consider a set I of n items "known" to contain exactly d defectives. The defectives look exactly like the good items and the only way to identify them is through testing (group), which is "error-free".

A test can be applied to an arbitrary subset of I with "two" possible "outcomes": a negative outcome indicates all items in the subset are good, a positive outcome indicates at least one item of the subset is defective (but not knowing which ones or how many are defective).

A positive group is also said to be contaminated