知 識 天 地

關鍵時刻： 組合群試理論揭秘

陳宏賓(數學研究所研究學者)

寶傑：新聞萬象，內幕追擊，歡迎收看<<關鍵時刻>>，我是劉寶傑。我們先來看看底下這則新聞!

哈佛知名政治經濟學家 羅伯特・多夫曼(Robert Dorfman) 於2002年6月24日凌晨逝世了。多夫曼的一生非常精采，在經濟學領域有非常大的貢獻。但大家不知道的是，在轉到經濟領域之前，他其實是學統計出身。從1939年開始為聯邦政府作統計相關的工作，在第二次世界大戰期間，還同時替美國軍方服務，從事運作分析師的後勤工作。戰爭爆發，美國軍方急需徵召大量軍人投入戰場，又深恐當時的致命傳染病梅毒會在軍隊中擴散開來，於是下令全部士兵都必須經過梅毒篩檢，在此機緣下，多夫曼想到了一個絕妙的點子，這個想法間接開創了一個新興領域 [組合群試理論]，1943年在The Annal of Mathematical Statistics發表的一篇論文“The detection of defective members of large populations”，這篇論文.....

大家都非常好奇為什麼這麼多人開始關注這個議題。
到底群試理論如何幫美國打贏這場戰爭? 有沒有甚麼不能說的秘密?
好! 我們今天請到五位來賓。
第一位是大家熟悉的資深媒體人馬西平，西平你好。
西平：寶傑好! 大家好!
寶傑：第二位是軍事專家張友樺，友樺你好。
友樺：寶傑好! 各位觀眾好!
寶傑：好! 第三位是世界文化史專家眭號坪。
號坪：寶傑好! 大家好!
寶傑：好! 第四位是資深歷史文化導遊謝哲輕。
哲輕：寶傑哥，各位觀眾晚安!
寶傑：好! 第五位是媒體工作者黃創下。
創下：寶傑好! 大家好。
寶傑：好! 我們剛剛看到了今天香蕉日報的一則新聞，到底群試理論又是甚麼東西呢? 先請西平來為我們說明。
西平：寶傑! 我跟你講，群試理論基本上是一個很簡單的想法。原本的梅毒篩檢都是抽血然後再一個一個進行檢驗，你知道這有多花錢嗎? 多夫曼的想法就是把好幾份血液混合再一起進行一次檢驗，如果沒有感染就表示這群血液都是安全的，如果有感染則表示至少有一份血液是受感染的，那再個別進行檢驗就好了，他用這個做法節省了很多檢驗的次數，也替美國軍方省下不少錢。但是，你千萬不要以為群試理論很簡單啊，隨便弄一弄就可以節省檢驗次數。上次我在台大醫院檢驗科進行篩檢就....
寶傑：好! 西平你先等等。就我所知，典型群試理論的模型是假設n個血液樣本之中剛好有d個是受感染的。友樺你知道為什麼沒辦法隨便弄一弄就節省檢驗次數?
友樺：寶傑，這件事全世界只有三個人知道，一個是多夫曼的好朋友羅伯特・索羅（1987年諾貝爾經濟學獎得主），他曾經公開稱讚多夫曼1943年發表在The Annal of Mathematical Statistics的那篇著作[1]“The detection of defective members of large populations”是組合群試理論的一個重要里程碑，另一個是我，最後一個我不能說。你知道嗎? 要能夠利用群試節省檢驗次數是要依賴統計數據的，不只這樣，你還要會演算法、組合設計、編碼理論...等等複雜的工具才行...
寶傑：等等，友樺。你說依賴統計數據是指要事先用統計資料預估感染率嗎?

友樺：對! 寶傑。在台灣有一個非常知名的組合數學家黃光明教授，他在1971年[2]首先證明了感染率高於1/2的話，也就是n≦2d，那麼群試理論就派不上用場了，也就是說一個一個檢驗就是最好的策略。後來他和他的合作者[3]還猜測「群試理論能夠省下檢驗次數的話，感染率必須低於1/3」，而目前最好的結果8/21是他和堵丁柱提出的[4]。現在這個猜測還沒有被證明出來，不過你知道嗎? 其實我已經有一個證明方法，只是節目時間太短我怕.....

寶傑：好! 創下。友樺剛才說到演算法和編碼理論，這方面你有沒有甚麼八卦?

創下：寶傑，你問對人，我當年在跑科技新聞的時候，剛好認識一個搞資訊的專家，他告訴我群試理論主要分為兩類演算法，一類叫做逐次演算法(sequential algorithm)，剛才友樺說得結果就是這類，另一類叫做同步演算法(nonadaptive algorithm)，顧名思義逐次演算法的檢驗是一個做完接著做下一個，因此前一次檢驗的結果就能夠用來做為下一次檢驗設計的參考，同步演算法就不一樣了，它要求所有檢驗都要事先設計好，這樣所有檢驗就可以同時進行，節省檢驗的時間，最後再由全部檢驗結果推導出受感染的血液樣本。資訊學家把這樣的實驗設計想像成一組0-1密碼(code)，每一血液樣本分別在哪幾次檢驗就看成一個碼字(codeword)，我跟你講，要能夠同步進行檢測還能夠找到答案，那可是必須要滿足某種性質才有辦法的阿。滿足這種性質的最低要求被稱為d-可分辨，這種碼可以用來解決 [剛好有d個] 血液受感染的問題，只不過解碼時間需要$O(tn^d)$太長了；後來Kautz和Singleton 1964年[5]提出一種條件較強的碼稱做 d-分離碼，這種碼很厲害哦，除了可以用來解決 [最多有d個] 血液受感染的問題之外，解碼的時間也大幅減少到O(tn)，這裡t指的是每個碼字的長度。

寶傑：等等...你說可以用來解決 [剛好有d個] 的最低要求稱為d-可分辨，意思是說這種碼比起d-分離碼要節省檢驗次數嗎?

創下：是的! 寶傑，就同步演算法來說，檢驗次數跟解碼速度就像是魚與熊掌一樣不可兼得! 想要解碼速度快就得在檢驗次數上付出代價，這中間必須做出取捨。取捨一直都是人生重要的課題啊! 有一次我上小燕姐的百萬小學堂就差點.....

寶傑：好的，號坪，我聽說你很久以前就曾經為了d-分離碼到過俄羅斯的研究中心，是嗎?

號坪：對阿，我用講的你們都不相信，所以我拿照片出來給你們看! 看到了嗎? 中間這個是我，左邊是d-分離碼全球最知名的專家Dỳachkov，右邊是Rykov，他們兩個自從1982年開始發表了好幾篇關於d-分離碼的重要論文[6]。你知道嗎? 俄羅斯數學家非常擅長複雜的計算，雖然他們算出目前d-分離碼最短的碼長，但卻一直沒有找到比這個碼長還要好的d-可分辨碼，這裡的好是指漸近式的比較。一直到2007年才由陳宏賓博士和他的老師黃光明教授[7]解決這個問題，他們證明了這兩種碼在漸近式的比較下是一樣好的，簡單的說就是實際上這兩種碼的碼長只會差一個常數倍而已。我上次在台北車站的地下街，還意外地發現了一個跟d-可分辨碼有關的魔術道具 [孔明神算]。你看，畫面上這7張卡[圖一]列了中國古今的百家姓，有些姓重複出現在好幾張卡上面，這個東西就叫做孔明神算。你知道孔明就是三國時代蜀國的軍師諸葛亮.....

寶傑：好，哲輕，你在中國帶導遊這麼多年，有沒有聽說甚麼孔明神算的事蹟呢? 孔明神算真的是諸葛亮發現的? 是嗎?

哲輕：好的，寶傑哥，我在中國當導遊的時候，有一回帶團到湖北襄樊，這裡就是古時候襄陽城的附近，臥龍諸葛亮年輕時就隱居在西南方15公里的隆中，那次我來到了諸葛亮故居紀念館，一進門忽然就被諸葛亮的神像吸進去了，彷彿一直無窮盡的往下掉，接著就看到一些奇怪的畫面，先是抗戰勝利，然後國父推翻滿清，接著韋小寶、朱元璋、岳飛...然後咻的一聲碰，我就躺在一張草蓆上了，只記得隱約看到屋內有兩個人，大人頭戴綸巾手拿羽扇輕輕搖擺，小童則搬著木板。

「主人，您要我把這七張寫著姓氏的木板兒放在門口做啥阿?」

「正所謂，天機不可洩漏。哈哈...」

約莫經過一柱香的時間，門外傳來一陣呼喊。

「請問，諸葛先生在嗎? 有人在嗎?」

「大哥，您看這門口擺了七塊奇怪的木板兒，上面希哩呼嚕不知道畫了些甚麼，要俺說這木板兒拿來切豬肉剛好。」

「三弟，叫你多讀點書你就不聽，這是百家姓。你瞧，大哥的姓氏不就出現在第四塊板子上，你的在第三塊，我的在第一塊和第七塊板子上。」
「俺...只是不想說而已，哼。裡面的人聽著，再不出來俺要殺進去啦~」
只見屋內那人在小童耳邊不知說了甚麼後，小童轉身出去應門。
「主人今日不見客，請劉先生回去吧。」
黑面人聽了之後大吃一驚：「這! 這臭小子怎麼知道咱大哥姓劉!」
「主人不只知道您大哥姓劉，二哥姓關，還知道您老爸姓張呢!」
三人面面相覷，無不震驚，帶頭者拱手：「煩請回告您家主人，劉備擇日再訪，告辭。」
童子一轉進門說：「主人您神機....」
「你想說我神機妙算，怎麼得知三人姓氏是吧? 告訴你也無妨，這是我近日研究數術發現的，這玩意兒叫做1-可分辨碼，只出現在第四塊板上的就只有劉字，只出現在第三塊板上的是張字，而同時出現在第一和第七塊板上的就只有關字了，不信你去瞧一瞧。」
「哇~ 主人您真是太厲害了，不只懂得替馬接生，沒想到您還懂可分辨碼呀!」
「哈哈....略懂，略懂。」
接著，我就迷迷糊糊的醒來了，醒來後發現自己還站在諸葛亮故居紀念館裡，也不曉得剛才是怎麼了....
寶傑：好的! 聽完哲輕在諸葛亮故居的奇遇之後，時間也差不多了，我在這裡做個總結。群試理論發展至今已經應用在很多地方了，除了實驗設計之外，還有產品檢驗、藥物篩選、DNA定序、壓縮感測以及最近的網路安全防護。因應真實應用的需求，群試理論有很多種變型，舉例來說，有抑制物型、門檻型、複合物型...等等，目前關於群試理論的專書是堵丁柱和黃光明寫的，第一本[8]主要是逐次演算法理論，第三本[10]主要是同步演算法理論和分子生物應用，第二本[9]則介於兩者之間，有興趣的觀眾可以參考看看。今天的節目就到此告一段落，我是劉寶傑，<<關鍵時刻>>下次再會。
參考資料：

[1] R. Dorfman, The detection of defective men=mbers of large populations, Ann. Math. Statist. 14 (1943) 436-440.
[2] F.K. Hwang, A minimax procedure on group testing problems, Tamkang J. Math. 2 (1971) 39-44.
[3] M.C. Hu, F.K. Hwang and J.K. Wang, A boundary problem for group testing, SIAM J. Alg. Disc. Methods 2 (1981) 81-87.
[4] D.Z. Du and F.K. Hwang, Minimizing a combinatorial function, SIAM J. Alg. Disc. Methods 3 (1982) 523-528.
[5] W.H. Kautz and R.R. Singleton, Nonrandom binary superimposed codes, IEEE Trans. Inform. Thy. 10 (1964) 363-377.
[6] A.G. Dỳachkov and V.V. Rykov, A survey of superimposed code theory, Problems. Control Inform. Thy. 12 (1983) 1-13.
[7] H.B. Chen and F.K. Hwang, Exploring the missing link among d-separable, dbar-separable and d-disjunct matrices, Disc. Appl. Math. 133 (2007) 662-664.
[8] D.Z. Du and F.K. Hwang, Combinatorial Group Testing and Its Applications, World Scientific, 1993.
[9] D.Z. Du and F.K. Hwang, Combinatorial Group Testing and Its Applications, 2nd ed., World Scientific, 2000.
[10] D.Z. Du and F.K. Hwang, Pooling Designs and Nonadaptive Group Testing - Important Tools for DNA Sequencing, World Scientific, 2006.

圖一：孔明神算

陳	黃	李	吳	蔡
許	謝	洪	邱	賴
徐	葉	呂	何	羅
簡	鍾	游	沈	胡
盧	顏	趙	翁	方
張簡	范	宋	杜	曹
傅	溫	關	歐	連
馬	石	程	康	古
湯	白	涂	巫	鐘
嚴	黎	袁	陸	錢

(一)

林	黃	王	吳	楊
許	郭	洪	廖	賴
蘇	葉	江	何	高
簡	施	游	彭	胡
潘	顏	柯	翁	孫
張簡	歐陽	宋	侯	曹
丁	溫	蔣	歐	唐
馬	卓	程	馮	古
汪	白	鄒	巫	鞏
嚴	阮	袁		

(二)

張	李	王	吳	鄭
謝	郭	洪	周	徐
蘇	葉	蕭	羅	高
簡	詹	沈	彭	胡
梁	趙	柯	翁	戴
范	歐陽	宋	薛	傅
丁	溫	藍	連	唐
馬	姚	康	馮	古
田	涂	鄒	巫	韓
黎	阮	袁	邵	

(三)

劉	蔡	楊	許	鄭
謝	郭	洪	莊	呂
江	何	蕭	羅	高
簡	余	盧	潘	顏
梁	趙	柯	翁	鄧
杜	侯	曹	薛	傅
丁	溫	董	石	卓
程	姚	康	馮	古
尤	鐘	鞏	嚴	韓
黎	阮	袁		

(四)

曾	邱	廖	賴	周
徐	蘇	葉	莊	呂
江	何	蕭	羅	高
簡	魏	方	孫	張簡
戴	范	歐陽	宋	鄧
杜	侯	曹	薛	傅
丁	溫	姜	湯	汪
白	田	涂	鄒	巫
尤	鐘	鞏	嚴	韓
黎	阮	袁		

(五)

朱	鍾	施	游	詹
沈	彭	胡	余	盧
潘	顏	梁	趙	柯
翁	魏	方	孫	張簡
戴	范	歐陽	宋	鄧
杜	侯	曹	薛	傅
丁	溫	童	陸	金
錢	邵			

(六)

紀	關	蔣	歐	藍
連	唐	馬	董	石
卓	程	姚	康	馮
古	姜	湯	汪	白
田	涂	鄒	巫	尤
鐘	鞏	嚴	韓	黎
阮	袁	童	陸	金
錢	邵			

(七)