

# 行政院國家科學委員會專題研究計畫成果報告

## 代數與離散結構之解析與隨機性質之研究之

### 子計畫四：應用性學科中組合結構之計算性研究(2)

#### Computational Studies of Combinatorial Structures in Various Applied Disciplines, Part II

計畫編號: NSC - 88 - 2115 - M - 032 - 001

執行期限: 1998 年 8 月 1 日至 1999 年 7 月 31 日

計畫總主持人: 葉永南

中央研究院數學所

子計畫主持人: 楊柏因

淡江大學數學系

## 1 中文摘要

沿襲前一期計畫的腳步, 在過去一年中, 我們繼續對應用性的問題- 包括特別是演算法, 統計和化學數學上的組合問題, 做了一些計算性的探討; 我們繼續對排序與尋找上幾個常見的演算法的改進方式, 並應用複變數函數的方法做了一些以前沒有的漸近行為分析; 以數年前報告人曾進行的系列對晶格路徑的討論, 處理一些在無母數統計上發生的組合問題; 並繼續尋找一些化學數學上生成函數式的結果。結果並簡述於下。

**關鍵字:** 演算法, 化學數學, 生成函數, 組合結構, 統計學

## Abstract

During the period of study under the sponsorship of the NSC, we followed the tracks of the previous year, making further computational investigations into some applied problems, notably algorithms, statistics and chemical mathematics. We found some practical improvements on some frequently used sorting methods, and analyzed the asymptotic behavior of expected costs using some improvements from complex analysis. Some of these may be expected to

improve real-life results in computing. We also did some extension of combinatorial Lattice Path Counting arguments (pursued several years previously in a series of colloquia) as used in non-parametric statistics. We sketch our results, in preparation for journal submission, as below.

**Keywords:** Algorithms, Chemical Mathematics, Combinatorial Structures, Generating Functions, Statistics

## 2 緣由與目的

在準備展開整個整合型計畫的工作時, 我們曾發表過這樣的看法:

計算機現已廣泛而深遠的被應用於各門學科, 也刺激了近年來離散結構的蓬勃研究與發展。這個趨勢不僅引進了大量新的結構, 同時也給既有的理論開啓了許多新的研究方向。

另一方面傳統的組合學研究在許多的數學家努力耕耘下, 現在也已進入一個更有系統與更抽象層次的階段。例如: 在計數組合學 (Enumerative Combinatorics) 中就已引入了大量代數技巧, 並加以純化萃煉, 使其更加獨立、完整、且更能有系統的在其他學科應用。

在這樣的時代潮流之下, 代數性質的研究逐漸完備再次刺激其他性質, 例

如漸近 (asymptotic) 與隨機 (probabilistic) 性質的進一步探討。這兩類帶有高度相關的研究對於離散結構而言可謂不可或缺 (很可惜地, 卻為大多數的組合學家所忽略), 其主要原因之一是: 大多數的計數組合問題都有所謂的組合爆炸現象, 再加上大多數的計算公式都十分複雜, 雖可借助於如 Lagrange Inversion, 或 Wilf-Zeilberger Summation Method 的技巧, 或電腦代數軟體來幫助, 但是在實用上還是經常有必要尋求漸近或逼近式, 而樣本空間過於龐大, 也使得隨機性質的探討成為必然之勢。

去年, 我們的目標是從應用性學科中取得素材, 並應用日漸多樣化的技巧來探討其組合性質。今年, 我們繼續進行一些同類型的研究, 但特別著眼於尋求領域間的配合。

### 3 結果

完整的結果目前我們尚在整理中, 分為幾點來簡述我們的工作:

#### 3.1 計算機科學上的探討

我們曾說:

據說電腦有四分之一的時間花在處理排序上, 因之排序是一個典型並且重要性毋庸置疑的演算法。

這句話當然仍然成立。我們亦花了一些功夫於簡化並加強之前的工作與發掘新的方向。排序的一種演算法, Hoare 氏的 Quicksort- 也就是在陣列中取樣做為 Key Value, 並以之把原陣列分割為較 Key 高與低的兩段, 的遞迴式運作- 也因之變成所有現代操作系統的系統內建工具副程式 (system library subroutine) 的一部分。但是 Quicksort 在兩個方面都有加強的空間: 它並不穩定, 另外, 雖在分割上可以採取預期其分割較均勻的方式,

但就必須花更多的前置作業時間來達成這個目標。由於 Quicksort 自 D. Knuth 與 R. Sedgwick 以來也是最為人所分析的演算法之一, 我們曾作了這個結論:

大部份的在 Quicksort 的基本方法上所做的變形..., 都浪費到太多前置作業的時間; ...對於一個經常使用的系統副程式而言, 它們是接受到“報酬遞減律”的制約, 因此是不經濟的。

去年, 採取在我們的論文([22]) “如何經濟的增加已排序度”的方針, 我們探討了兩個變式: Tukey 的準九樣本中位數 (Ninther) 分割法, 和七樣本全排序法 (7-Samplesort), 其與常見的各種 Quicksort 變形的平均比較次數有如下的表示:

$$Q(n) \approx 2 n \ln n - 2.84n \quad (1)$$

$$Q_3(n) \approx 1.72 n \ln n - 2.24n \quad (2)$$

$$Q_{7s}(n) \approx 1.64 n \ln n - 9.03n \quad (3)$$

$$Q_{9p}(n) \approx 1.57 n \ln n - 6.44n \quad (4)$$

去年測試的結果, 發現 7-samplesort 的效能 500 個數據以上就十分顯著, Ninther 則要到 30000 個數據左右才比前者為優。

為了探討實際上的用途, 我們把由 Richard Stallman 創設, 十年以來被認為是程式庫經典之作的 GNU C Library (glibc2 為其第二版, 見於幾乎所有的 Unix 系統中) 中關於排序的部份加以修改來測試我們由 Ninther 與 7-samplesort 得來的方法在實用中的效果, 並分析在現實中, 由於 page fault (即由於 cache 大小有限造成越小的程式執行效果越佳) 造成的影響。我們的結論是 7-samplesort 與 Ninther (此處我們借用了 Jon Bentley 於 [5] 中的種種改進), 分析顯示對於一個實際使用的系統而言, 我們利用 Ninther 可以改動 glibc2 中的一般排序程式 qsort() 而獲得將近 3% 的效果。這是因為 glibc2 中全面改用穩定快速但牽

涉到大量記憶體存取的 Mergesort 的緣故; 由於現在的 Intel CPU 架構, 電腦經常是 Memory bandwidth bound (被記憶體存取的速度限制), 這讓使用 in-place sorting 在 Intel x86 系列具有某些優勢。

我們已將我們的一些改動 (patch file) 送給 GNU C Library 的 glibc project maintainer, 由於關係到系統程式庫, 這些修改茲事體大, 仍需由 GNU Project 的人員進一步測試, 但是或可在不久的將來列入新的 glibc 中。

同時, 我們對同類的 Find Algorithm 等亦以同類型的方法做出修正, 這允許我們以更高速度方式找出一組數據中某一個指定排名的位置。經測試, 實際生活中可以減少大約 5% 的搜尋時間。

### 3.2 其他的探討

延續上期的工作, 我們對數學化學的工作中整理了一些非遞迴性的公式, 可用以計算幾大類化學上有意義的一般偶多邊形環鍊型圖之 Wiener Polynomial (即圖形中頂點間距離的生成函數), 總成兩篇論文 (其中[28] 已被接受); 對 partition 和 composition 總體的漸近性質我們亦延續地作了一些探討。

最後, 我們借用組合學中的 Lattice Path Counting 的一些方法來討論統計學中常見的 non-parametric hypothesis testing, 結果正在整理中。

## 4 討論與計畫成果自評

我去年如是說:

我們的努力方向一直在於: 在一些常見的計算機科學領域上, 使用組合學的概念, 但是引入分析學的技巧, 希望藉由改換觀點來尋找一些問題的新處方 (prescription) ... 我們相信不論是理論或是實用性上來說, 我們的結果都將是相當有用的。

我仍相信這個判斷, 從各自從各自的方向蒐集的一些資料, 已開始建構一些比

較統合性的問題處理模式。現在我們的目標是能整合出一些跨領域都有用的漸近結果。

## 5 參考文獻

### References

- [1] G. E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and Its Applications, Volume 2, Edited by Gian-Carlo Rota, Addison-Wesley Publishing Company, 1976.
- [2] E. A. Bender (1973). Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory, Series A* **15** 91-111.
- [3] E. A. Bender and L. B. Richmond (1983). Central and local limit theorems applied to asymptotic enumeration. II. Multivariate generating functions. *Journal of Combinatorial Theory, Series A* **34** 255-265.
- [4] E. A. Bender, L. B. Richmond and S. G. Williamson (1983). Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions. *Journal of Combinatorial Theory, Series A* **35** 263-278.
- [5] J. Bentley, *Programming Pearls*, 2nd ed. Addison-Wesley, 1999.
- [6] P. Erdős and M. Szalay, *On the statistical theory of partitions*, Colloquia Mathematica Societatis János Bolyai, 34, Topics in classical number theory, Budapest, 397-450, 1981.
- [7] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, **3**, 216-240 (1990).

- [8] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *Journal of Combinatorial Theory, Series A*, 53, 165–182 (1990).
- [9] Foata, D. (1974). *La série génératrice exponentielle dans les problèmes d'énumération*. Les Presses de l'Université de Montréal, Montréal.
- [10] B. Fristedt, The structure of random partitions of large integers, *Transactions of the American Mathematical Society*, 337 (1993), 703–735.
- [11] W. M. Y. Goh and E. Schmutz, The number of distinct part sizes in a random integer partitions, *Journal of Combinatorial Theory, series A*, 69 (1995), 149–158.
- [12] I. P. Goulden and D. M. Jackson, *Combinatorial enumeration*, John Wiley & Sons, New York, 1983.
- [13] D. H. Green and D. E. Knuth, *Mathematics for the Analysis of Algorithms*
- [14] J. C. Hansen, Order statistics for decomposable combinatorial structures, *Random Structures and Algorithms*, 5, 517–533 (1994).
- [15] C. B. Haselgrove and H. N. V. Temperley, Asymptotic formulae in the theory of partitions, *Proceedings of the Cambridge Philosophical Society*, 50 (1954), 225–241.
- [16] P. Henrici, *Applied and computational complex analysis*, three volumes, John Wiley & Sons, New York, 1977.
- [17] J. Herzog, Weak asymptotic formulas for partitions free of small summands, II, *Acta Mathematica Hungarica*, 62 (1993), 173–188.
- [18] H.-K. Hwang, Distributions of integer partitions with large number of summands, *Acta Arithmetica*, accepted for publication.
- [19] H.-K. Hwang, A Poisson \* geometric law for the number of components in unlabelled combinatorial structures, *Combinatorics, Probability and Computing*, accepted for publication.
- [20] H.-K. Hwang, *Théorèmes limites pour les structures combinatoires et les fonctions arithmétiques*, Thèse, Ecole polytechnique, 1994.
- [21] H.-K. Hwang and Y.-N. Yeh, Measures of distinctness for integer partitions and compositions, preprint, 1996.
- [22] H.-K. Hwang, B.-Y. Yang and Y.-N. Yeh, *Presorting algorithms: an average-case point of view*, Theoretical Computer Science, (accepted, to appear).
- [23] D. E. Knuth, *The art of computer programming, volume II, seminumerical algorithms*, Second edition, Addison Wesley, Reading, MA, 1981.
- [24] R. P. Stanley, *Enumerative combinatorics*, Monterey, CA, Wadsworth, 1986.
- [25] G. Szekeres, Asymptotic distributions of the number and size of parts in unequal partitions, *Bulletin of the Australian Mathematical Society*, 36 (1987), 89–97.
- [26] R. Wong, *Asymptotic approximations of integrals*, Academic Press, Inc., Boston, 1989.
- [27] B.-Y. Yang, and Y. Yeh, *Zigging and Zagging in Pentachains*, *Advances in Applied Mathematics* 16(1995) 72-94.
- [28] B.-Y. Yang, and Y. Yeh, *About Wiener Numbers and Polynomials*, to appear in *Proc. of second Tainan-Moscow Algebra Workshop*, Fong et al, eds., to be published by Springer-Verlag.