行政院國家科學委員會專題研究計畫　成果報告

無母數檢定與排列檢定之系統性研究

計畫類別：個別型計畫
計畫編號：NSC94-2118-M-032-012-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：淡江大學數學系

計畫主持人：鄭惟厚

計畫參與人員：劉家豪　柯舜傑

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 26 日

行政院國家科學委員會補助專題研究計畫 ☐ 成 果 報 告
☐期中進度報告

無母數檢定與排列檢定之系統性研究

計畫類別：☐ 個別型計畫 ☐ 整合型計畫
計畫編號：NSC 94－2118－M－032－012 －
執行期間： 94 年 8 月 1 日至 95 年 7 月 31 日

計畫主持人：鄭惟厚
共同主持人：
計畫參與人員： 兼任助理碩士生劉家豪、柯舜傑

成果報告類型(依經費核定清單規定繳交)：☐精簡報告 ☐完整報告

本成果報告包括以下應繳交之附件：
☐赴國外出差或研習心得報告一份
☐赴大陸地區出差或研習心得報告一份
☐出席國際學術會議心得報告及發表之論文各一份
☐國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
☐涉及專利或其他智慧財產權，☐一年☐二年後可公開查詢

執行單位：淡江大學

中 華 民 國 95 年 10 月 26 日

中文摘要：

排列檢定由於計算繁複，到了最近電腦普遍且威力強大之後，才受到愈來愈多的重視。而計算排列檢定 p-value 的軟體，在種類和效率方面也持續進步當中，使用排列檢定將會愈來愈方便。在本報告中檢討了排列檢定和無母數檢定的關係；比如無母數檢定中哪些可視為排列檢定、哪些不是。Rank tests 的考慮基礎，就是所有 ranks 的可能排列機率相等，和排列檢定的精神一樣，所以在此不需考慮。在經過比較討論之後，我們得到以下結論。

無母數檢定中與 binomial test 相關的檢定，基本上都可以視作排列檢定的一種，只要找出與檢定問題相關的 equally likely 性質，就可以找出相對的排列分布。如果檢定本身與卡方檢定相關，例如列聯表的問題，基本上就會與格機率 p 的值關係密切。如果知道確實的 p 值，就可以利用討論 binomial test 的方式，將這類卡方檢定相關的問題與排列檢定建立關係，也就是可以找到某個方式的排列檢定與之對應。但如果 p 值未知，就需要視列總和及行總和是否固定來做不同的討論。另外，如 Kolmogorov goodness-of-fit test 之類的檢定，只要統計量是根據順序統計量算出的，因為任一組數據無論怎樣重新排列，其順序統計量都不變，所以無法找到可以判別檢定問題的排列分布。

總結來說，大部分的無母數檢定可以視為排列檢定，關鍵點在於給予所謂「排列」一個適當的定義。所以應該可以這樣說：大部分常見的無母數檢定都可以視為廣義的排列檢定。

關鍵詞：排列檢定、無母數檢定

英文摘要:

Permutation tests go back a long way. But since they involve incredible amount of computations, it is until fairly recently when computers are more and more powerful and softwares are more readily available that they receive new attention. What makes permutation tests special is minimal assumptions about the data, hence much wider range of applications.

Since nonparametric tests also have relatively wide applications, it will be interesting to find out how nonparametric tests and permutation tests are related to each other. It is obvious that rank tests are permutation tests of the ranks. But can other nonparametric tests be viewed as permutation tests as well? In this report we look at common nonparametric tests and decide which of them can or cannot be considered as permutation tests.

Key words: permutation tests, nonparametric tests

前言

對於在統計推論中佔絕大部份的統計檢定，Good[3]將其分成四大類來討論：

(1) 排列檢定(permutation tests)
(2) 秩檢定(rank tests)
(3) Bootstrap
(4) 母數檢定(parametric tests)

其中以母數檢定最廣為人知且受到重視，其次為秩檢定；秩檢定其實被許多教科書歸類為無母數檢定的一部分。Bootstrap 和排列檢定因計算量極大，所以在早些年前，並不是那麼多人使用。然由於近年來電腦科技的發達，大量計算已不成問題，期刊上 bootstrap 的文章相當多，但排列檢定似乎尚未普遍應用。

然而排列檢定可使用的範圍極廣，無論數據屬於區間/比例尺度(interval/ratio scale)、順序尺度(order scale)、還是類別(categorical)資料都可以使用，母體的分布也不要求是否為常態或接近常態。對大部份的母數或無母數(nonparametric)檢定，都可以找到對應的排列檢定，而此排列檢定常有同樣甚至更好的檢力(power)。即使其檢力差不多，排列檢定的優勢在於它的假設最少。比如在單一樣本的問題中，可能只需要假設對稱分布。

而排列檢定和無母數檢定之間有什麼關係呢？比如一般無母數檢定裡面，哪些屬於排列檢定、哪些不是，應是值得探討的題目。

研究目的

Good 聲稱（3, p.9）一般無母數檢定當中，有 99%等於是用在秩上面的排列檢定，但是並未加以說明。這似乎提出了一個有趣的問題，即在常用的無母數檢定當中，有哪些是排列檢定而哪些不是？

初步看來似乎大部分都是，因為無母數檢定有許多都是根據數據的秩來做的，而秩檢定（rank tests）明顯相當於排列檢定。其他的無母數檢定哪些屬於排列檢定、哪些不是，是否能做出歸類？有沒有既不是秩檢定、也非排列檢定的無母數檢定存在呢？本計劃希望對這些有趣的問題作探討，做出結論供大家參考。

文獻探討

排列檢定的概念很早以前就有人用了，比如說 Fisher[2]用在人類學上的研究，也有很多人將其應用在農業、考古、大氣科學、生物、生態、教育、流行病學、遺傳、地理、地質、醫藥、古生物學、社會學…等等上面，可說應有盡有，這些相關的文獻在 Good[3]的書中都有列出。它最大的特色在於假設極少，基本上檢定結果只根據數據本身而決定。從此角度看，可說它是最「自然」的方法。而因為背景假設少，對應用的人來說，應該比較容易理解，也比較不用擔心應用得是否恰當。

排列檢定由於計算繁複、應用較不普遍，許多教科書中也未著墨，但無母數檢定就普遍多了。二者間哪些相異、哪些相同，少見有人討論。Good 聲稱（3, p.9）一般無母數檢定當中，有 99%等於是用在秩上面的排列檢定，但是卻未加以說明。

有的學者認為排列檢定和 randomization tests 是相同的概念，其他學者卻將二者做區分，Sprent [4]對此有討論。

研究方法

我們以 Conover[1]為準，對常用的一些無母數檢定分別討論，探討其是否相當於排列檢定，但是秩檢定(rank test)除外。因為如果將資料轉成秩，而統計量是秩的函數(叫做秩統計量)，則該統計量在原始假設之下的分布基本上就是將所有排列情況下的秩統計量找出，和排列檢定的精神完全符合，所以不需要另外討論。

我們做比較時所用的方法，是檢視無母數檢定的統計量，判斷它在原始假設下的分布，是否可用排列檢定的概念獲得。排列檢定的主要基礎是 equally likely 性質，在原始假設之下，所有「排列」發生之機會均等，根據這個性質可找出統計量的分布，叫做排列分布。所以排列檢定不如說是一種概念，而所謂「排列」在不同的問題之下有不同意義。因次在考量的時候不能只從表面看；即使表面看來不像是排列檢定的無母數檢定，若能把「排列」的意義釐清，還是可能把無母數檢定和排列檢定連結起來。

結果與討論

二項檢定(binomial test)之討論

在符合 binomial test 的假設下，樣本大小為 n ：考慮檢定

$$H_0 : p = p^*$$

在此原始假設下，我們利用排列檢定的性質，equally likely，找出統計量對應的排列分布，此處的統計量等於 n 個樣本資料中「成功」的個數，即屬於 class1 的個數。將所有可能的結果列出，再找出可能的方法數。

class 1	*class 2*
x	$n-x$

1) $p^* = \frac{1}{2}$

這 n 個樣本資料分成兩類，且每個樣本資料被歸類到 class 1 與 class 2 的機會相同，也就是由實驗得到單一組結果的機率相同，均為 $\frac{1}{2^n}$，利用 equally likely 這個性質找出排列分布。

結果	(0, n)	(1, $n-1$)	(2, $n-2$)	$\cdots$	(n,0)
方法數	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$		$\binom{n}{n}$

所以得到($x, n-x$)的機率為

$$P(X = x) = \frac{\#(X = x)}{\sum_{x'=0}^{n} \#(X = x')} = \frac{\binom{n}{x}}{2^n} = \frac{(x, n-x)\text{的方法數}}{\text{所有可能方法數總和}}$$

2) $p^*=\frac{q}{w}\neq\frac{1}{2}$，$q$ 與 w 為正整數且最大公因數為 1

每個樣本點被分配到 class 1 與 class 2 的機會不同，所以我們先將 class 1 與 class 2 作分割。將 class 1 分成 q 格，class 2 分成 $w-q$ 格，使任一樣本點落入 w 格中任一格的機會相同，如此可得單一組 n 個樣本點落入 w 格數的結果機率皆相同，為 $\frac{1}{w^n}$。因此這裡也運用相同的原理，只是現在不只有兩類的區別，如符號的正或是負。所以換一個想法看此問題，可以視為 class 1 為正號，但是這些正號有 q 類，class 2 視作負號，而這些負號有 $w-q$ 類。一樣作這些不同類型的排列，只是在計算統計量時，需要加總所有 q 類的正號數值，因此統計量算法也比較不同。

class 1		*class* 2		
x_1	$\dots$	x_{q+1}	$\dots$	x_w
x_q				

$x_1+x_2+\ldots+x_q=x$， $x_{q+1}+x_{q+2}+\ldots+x_w=n-x$

其中 x_i 為分割後的第 i 格中樣本的個數

例：$p^*=\frac{1}{3}$，為了符合 $\frac{1}{3}$ 的機率，我們共分三格，其中第一格屬於 class 1，用 a_1 表示，而 class 2 再分成 2 格，分別用 a_2，a_3 表示。任一觀測值會落入 a_1，a_2 或是 a_3 任一格的機率都相同。

class 1		*class* 2
a_1	,	a_2
	a_3	

因此在 $p^*=\frac{1}{3}$ 的情況下，可能的結果如下：

結果	$(0,n)$	$(1,n-1)$	$(2,n-2)$	$\dots$	$(n,0)$
	$(0,[a,n-a])$	$(1,[a,n-1-a])$	$(2,[a,n-2-a])$	$\dots$	$(n,[0,0])$
	$a=0,\ldots,n$	$a=0,\ldots,n-1$	$a=0,\ldots,n-2$	$\dots$	
方法數	$\sum_{a=0}^{n}\binom{n}{a}$	$\binom{n}{1}\sum_{a=0}^{n-1}\binom{n-1}{a}$	$\binom{n}{2}\sum_{a=0}^{n-2}\binom{n-2}{a}$	$\dots$	$\binom{n}{n}$

所以得到 $(x,n-x)$ 的結果機率為

$$P(X=x)=\frac{\#(X=x)}{\sum_{x'=0}^{n}\#(X=x')}=\frac{(x,n-x)\text{的方法數}}{\text{所有可能方法數總和}}$$

這部分的所有可能方法數總和為3^n

$$P(X=x)=\frac{\binom{n}{x}\sum_{a=0}^{n-x}\binom{n-x}{a}}{3^n}=\frac{\binom{n}{x}2^{n-x}}{3^n}=\binom{n}{x}\left(\frac{2}{3}\right)^{n-x}\left(\frac{1}{3}\right)^{x}$$

現在考慮一般的情況 $p^*=\frac{q}{w}\neq\frac{1}{2}$ ，q與w為正整數且最大公因數為 1：

(i) 排列檢定

n個樣本點任一個落在w格中任一格的機會相同，只要找出排列組合數，即可得知機率：得到$(x,n-x)$結果的可能方法為：在n個樣本點先取出x個，這x個樣本點放在前q格中，剩下的$n-x$點則放入剩下的$w-q$格中，即$\binom{n}{x}q^x(w-q)^{n-x}$；將$n$個點放入$w$格中，總共有$w^n$可能，所以得到 $(x,n-x)$的機率為

$$P(no.\_in\_class\_1=x,no.\_in\_class\_2=n-x)=\frac{\binom{n}{x}q^x(w-q)^{n-x}}{w^n}$$

(ii) binomial test

$$P(no.\_in\_class\_1=x,no.\_in\_class\_2=n-x)=\binom{n}{x}p^x(1-p)^{n-x}$$

$$=\binom{n}{x}\left(\frac{q}{w}\right)^x\left(1-\frac{q}{w}\right)^{n-x}=\binom{n}{x}\frac{q^x(w-q)^{n-x}}{w^n}$$

所以 binomial test 也可以看成是某種排列撿定。

符號檢定 (sign test)：

兩成對的隨機變數資料中(X_i,Y_i)，後者大於前者稱作"＋"，後者小於前者稱作"－"。

我們考慮檢定 $H_0:P(+)=P(-)$

在原始假設條件下，兩變數之間大於或是小於的機會相同，也就是正負兩種符號出現的機會相同。這個訊息就可以利用到排列檢定的特性，equally likely。與 binomial test 的 $p=\frac{1}{2}$ 作法相同，即可以找出相對的排列檢定分布。

結論

無母數檢定中與 binomial test 相關的檢定（包括 binomial test，quantile test，sign test，McNemar test 等），基本上都可以視作排列檢定的一種，只要找出與檢定問題相關的 equally likely 性質，就可以找出相對的排列分布。

如果檢定本身與卡方檢定相關，例如列聯表的問題，基本上就會與格機率p值的關係相當密切；如果在原始假設下p值已知，就可以利用討論 binomial test 的方式，將這類卡方檢

定相關的問題與排列檢定建立關係，也就是可以找到某個方式的排列檢定與之對應。但是如果 p 值未知，就需要視列總和及行總和是否固定而有不同的結論。行、列總和都固定時，就是 Fisher’s exact test，當然是排列檢定，否則就不是。

另外，如 Kolmogorov goodness-of-fit test 之類的檢定，由於統計量植基於 empirical distribution function ，是根據順序統計量算出的。而任一組數據無論怎樣重新排列，其順序統計量都不變，所以無法找到可以判別檢定問題的排列分布。

總結來看，大部分的無母數檢定可以視為排列檢定，關鍵點在於給予所謂「排列」一個適當的定義。所以應該可以這樣說：大部分常見的無母數檢定都可以視為廣義的排列檢定。

參考文獻

[1] Conover, W.J. (1999), Practical Nonparametric Statistics, 3rd edi. (Wiley)
[2] Fisher, R.A (1936), Coefficient of racial likeness and the future of craniometry J.Roy.Anthrop .Soc 66, pp57–63.
[3] Good, Phillip (2000), Permutation tests (Springer-Verlag)
[4] Sprent, Peter (1998), Data Driven Statistical Methods (Chapman & Hall)

計畫成果自評

所做出的結果與原計畫內容相符，可說大致達成原來目標。對於無母數檢定和排列檢定之間關係的釐清有適度貢獻，且對Good所聲稱的「一般無母數檢定當中，有99%等於是用在秩上面的排列檢定」等於做出澄清，證實他的說法和事實有差距。 若要將此結果在期刊發表，尚有細節需補充。