行政院國家科學委員會專題研究計畫 成果報告

應用於可交換變數之無母數方法
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2118-M-032-007-
執 行 期 間 ：96 年 08 月 01 日至 97 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：碩士班研究生-兼任助理人員：許宏源

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，1 年後可公開查詢

中 華 民 國 97 年 10 月 09 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

應用於可交換變數之無母數方法

計畫類別：□ 個別型計畫　□ 整合型計畫
計畫編號：NSC　96－2118－M－032－007 －
執行期間：　96 年 8 月 1 日至 97 年 7 月 31 日

計畫主持人：鄭惟厚
共同主持人：
計畫參與人員： 兼任助理碩士生許宏源

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 97 年 10 月 9 日

中文摘要：

在做隨機化比較實驗的時候，許多時候是把並非隨機取得的樣本作隨機分組而已，且樣本可能並不大，此時使用傳統的 t 檢定或 F 檢定並不適合。有學者建議使用排列檢定，這當然是一種選擇。但排列檢定有它的弱點，另一種選擇會是無母數方法。但是一但把傳統的 i.i.d.條件減弱成可交換，就有許多工作需要做。比如檢定統計量的確實原始分布如果用到「互相獨立」的條件，就必須重新推導，檢定力也須重新檢視；本報告包括了 Wilcoxon rank sum test 檢定力之討論，並探討了在某種可交換條件下，用順序統計量所構成的信賴區間，其信心水準的變化狀況。

關鍵詞：可交換隨機變數、無母數方法

英文摘要:

When a randomized comparative study is carried out, the experiment groups were often constructed by randomization of nonrandom samples rather than random sampling from a population in which case the use of classical tests like t or F can be very inappropriate . Permutation tests were suggested by some researchers because they are based on exchangeability only. But permutation tests have their shortcomings and another choice might be to use nonparametric methods.But with the weaker condition of exchangeability instead of the usual i.i.d.,lots of work need to be done.For example,we will have to find the null distribution of some nonparametric test statistics under the condition of exchangeable instead of iid random variables, and power can change under this new condition.This report includes discussions of the power of Wilcoxon rank sum test,and the confidence coefficient of confidence intervals formed by order statistics under certain exchangeable conditions .

Key words: exchangeable random variables, nonparametric methods

前言

Ludbrook, J. 及 Dudley, H.[7] 檢視了生命科學領域 5 種常被引用的期刊，發現在 252 項研究中，只有 4%是從母體中隨機抽樣來產生實驗組，這些實驗對象都是動物。其他的研究只是把並非隨機樣本的受試對象(人、動物或其他)隨機分組來產生實驗組及對照組，所以不應假設兩組為獨立之隨機樣本；然而大部份的研究仍然用 t 檢定來判斷母體平均數是否相等，似乎背後的假設條件是否符合，並沒受到足夠的重視。

若考慮在 $X_1, X_2, \cdots, X_n$ 為可交換 (exchangeable)，而並非是 i.i.d. (independent, identically distributed) 的較弱假設下，使用一些傳統檢定，結果必然與 $X_1, X_2, \cdots, X_n$ 為 i.i.d.的假設下會有差異。無母數檢定是另一種選擇，然而在不同的可交換分配之下，檢定統計量的原始分配如何，檢定力又如何，是值得討論的問題。我們將探討在 $X_1, X_2, \cdots, X_n$ 為可交換的條件下， Wilcoxon rank sum 檢定的相關問題。另外，無母數方法常利用順序統計量來構成信賴區間，但少了 i.i.d.的條件，狀況會如何呢？我們將探討在某種可交換條件下，用順序統計量所構成的信賴區間之信心水準的變化狀況。

研究目的

可交換隨機變數的定義如下：設 $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ 為整數 $(1,2,\cdots,n)$ 的任意排列，如果隨機變數 $X_1, X_2, \cdots, X_n$ 滿足

$$(X_1, X_2, \cdots, X_n) \overset{d}{=} (X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_n}),$$

則稱 $X_1, X_2, \cdots, X_n$ 為可交換隨機變數。

“$\overset{d}{=}$” 表示分配相等。

隨機樣本因為符合 i.i.d.(independent, identically distributed)條件，必為可交換隨機變數；然而可交換隨機變數未必符合 i.i.d.假設，我們由下面的例子就可以看出這點。

例：考慮隨機變數 X_1, X_2，假設其聯合分配如下表所示：

x_1 \ x_2	0	1	$P(X_1 = x_1)$
0	1/3	1/3	2/3
1	1/3	0	1/3
$P(X_2 = x_2)$	2/3	1/3	

$$\because P\big((X_1, X_2) = (0,0)\big) = \frac{1}{3} \neq \frac{2}{3} \cdot \frac{2}{3} = P(X_1 = 0) \cdot P(X_2 = 0)$$

$\therefore X_1, X_2$ 不獨立，但明顯可見 $(X_1, X_2) \overset{d}{=} (X_2, X_1)$，所以

X_1, X_2 為可交換隨機變數，但並非 i.i.d.。

有的實際應用狀況並不符合 i.i.d.，而只是可交換，但若把傳統的 $X_1, X_2, \cdots, X_n$ 為 i.i.d.條件減弱成可交換，一些常用的檢定統計量的原始分布會有改變，檢定力也須重新檢視；本報告包括了 Wilcoxon rank sum 檢定之討論。「可交換」條件下的可能性很多，我們將利用 copula 的觀念來討論。另外就是在某種可交換條件下，考慮信賴區間的問題；希望能提供應用時的參考。

文獻探討

Ludbrook, J. 及 Dudley, H. 在 1998 年的一篇文章[7]中指出，由於大部分的醫學研究所用的樣本都不是隨機樣本，通常都只是把並非隨機取得的樣本作隨機分組而已，因此傳統常用的 t 檢定和 F 檢定並不適用，應該用排列檢定才適合。事實上醫學研究除了動物實驗外，要取得符合 i.i.d.如此嚴格假設的樣本並不容易，例如早期榮總如果大部分的病人都是榮民，相信彼此多少會有些關聯，若仍然當作 i.i.d.樣本來進行傳統檢定，相信會有些問題產生，要拿這些資料去推估整個母體似乎不太恰當。有 Langbehn, D.R. [6]等多位學者對 Ludbrook 及 Dudley 該篇文章作出回應。

在樣本不大又非隨機樣本的情況下，不管 i.i.d. 條件，而使用 t 檢定讓人有很大的疑慮。Ludbrook, J. 及 Dudley, H. [7]在文章中用了一個人造例子，比較了 t 檢定，Wilcoxon rank sum 檢定等的 p-value，發現差別相當大。然而 Box,G.E.P., Hunter,J.S. and W.G.[1]在他們出版的書中，對上述的結果也有意見；認為 Ludbrook 及 Dudley 所得到的結果，與他們所造出的數據之某項特質(有一個離群值)有很大的關係，這點和 Higgins, J.J. 及 Blar, R.C.兩人的意見很接近。

即便例子不理想，但是 Ludbrook, J. 及 Dudley, H.提出的觀點仍然值得重視；也就是說，在樣本並非獨立的隨機樣本、而只是經過隨機分組的情況下，與其在條件不符合的情況下使用 t 檢定等傳統檢定，不如尋求更適合的檢定方法。

因本文將樣本 $X_1, X_2, \cdots, X_n$ 在「可交換」的條件下，探討檢定的相關問題，因此我們需要模擬出不同分配的可交換隨機變數。在 Joe[5]及 Nelsen[8]的書中有提到幾個 copulas，是將邊際機率密度函數透過 copula 函數結合成聯合密度函數。
例如 normal copula、Frank copula 及 Gumbel copula 等，並且這幾個 copula 在假設邊際分配相同下均滿足可交換性。

研究方法

研究的主要目標是小樣本，在缺乏互相獨立條件的狀況下， Wilcoxon rank sum 檢定的檢定統計量在原始假設下的分布以及其檢定力都很難導出，只能依賴統計模擬。我們用到copula的觀念，先介紹如下：

Copula C 是從 $[0,1]^n$ 到 $[0,1]$ 的函數，假設 F_j 代表第 j個邊際c.d.f.的話，經過copula的連結，會得到以下的n維c.d.f.：

$$F(x_1, x_{2,}..., x_n) = C(F_1(x_1), F_2(x_2),..., F_n(x_n))$$

我們將從不同的多維分布取出可交換樣本，所用的多維分布包括多維常態、Gumbel copula with normal marginals 以及 Frank copula with normal marginals，而 Gumbel copula 及 Frank Copula 的定義如下：

Gumbel〔4〕 For $1 \le \delta < \infty$

$$C(u_1, u_2, \cdots, u_n; \delta) = \exp\left\{-\left(\tilde{u}_1^{\ \delta} + \tilde{u}_2^{\ \delta} + \cdots + \tilde{u}_n^{\ \delta}\right)^{\frac{1}{\delta}}\right\}$$

Frank〔2〕 . For $0 < \delta < \infty$

$$C(u_1, u_2, \cdots, u_n; \delta) = -\delta^{-1} \ln\left\{ \left[\eta^{n-1} - \left(1-e^{-\delta u_1}\right)\left(1-e^{-\delta u_2}\right)\cdots\left(1-e^{-\delta u_n}\right)\right] \Big/ \eta^{n-1} \right\}$$

我們考慮以下檢定問題

$$H_0 : \mu_X - \mu_Y = 0 \quad \text{v.s.} \quad H_1 : \mu_X - \mu_Y > 0$$

用無母數方法中常用的 Wilcoxon rank sum 來做檢定。假設 $X_1, X_2, \cdots, X_{n_1}$ 和

$Y_1, Y_2, \cdots, Y_{n_2}$ 來自同樣的連續分配，只有平均數可能不同。

Wilcoxon rank sum 檢定的檢定統計量定義如下: $W = \sum_{j=1}^{n_1} S_j$

其中 S_j 為 $X_1, X_2, \cdots, X_{n_1}$ 和 $Y_1, Y_2, \cdots, Y_{n_2}$ 混合排序後 X_j 之秩(rank)。

檢定規則：若 $W \ge w_\alpha(n_1, n_2)$ ，則 reject H_0

我們利用模擬來檢視檢定力，所考慮的可交換隨機變數包括多變量常態分配以及兩種 copula(Gumbel 及 Frank)，模擬的過程如下：

先產生維度 10 的樣本 $(X_1, X_2, \cdots, X_{10})$，然後隨機分為兩組，其中一組加上一個定值，當作 **X** 樣本，其餘的構成 **Y** 樣本，即

$$\mathbf{X} = (X_{\alpha_1} + \theta, X_{\alpha_2} + \theta, \cdots, X_{\alpha_5} + \theta)$$

$$\mathbf{Y} = (X_{\alpha_6}, X_{\alpha_7}, \cdots, X_{\alpha_{10}})$$

此處 $(\alpha_1, \alpha_2, \cdots, \alpha_{10})$ 是 $(1, 2, \cdots, 10)$ 的一個排列

如此共進行 10,000 次模擬，計算

$W > w_\alpha(n_1, n_2)$ 發生的比例

當做 θ 下的檢定力之估計 $\hat{\pi}(\theta)$。

同樣最後再比較各情況下的檢定力，假設 $\alpha = 0.048$ 下，模擬結果如下表

表 1：

multivariate normal density dim=10					
	$\theta = 0$	$\theta = 1$	$\theta = 1.5$	$\theta = 2$	$\theta = 2.5$
ρ	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$
0.0	0.0462	0.3929	0.6658	0.8631	0.9585
0.1	0.0485	0.4210	0.7057	0.8981	0.9763
0.3	0.0477	0.5073	0.7961	0.9479	0.9917
0.5	0.0440	0.6206	0.8957	0.9874	0.9991

由表 1 中，發現幾個現象：

(i) $\theta = 0$ 時，即使 $\rho > 0$ ，$\pi(\theta)$ 大致都在 0.048 附近。

(ii) 固定 θ 值($\theta > 0$)時，ρ 越大，$\pi(\theta)$越大

可能的解釋是，因為樣本間隨著相關係數 ρ 的增加而使得其差異越小，因此

$\mathbf{X} = (X_{\alpha_1} + \theta, X_{\alpha_2} + \theta, \cdots, X_{\alpha_5} + \theta)$ 因為加上 θ 值使得其中每個分量的秩會偏大，因此

$W = \sum_{j=1}^{n_1} S_j$ 亦會更大，如此一來 W 值大於 $w_\alpha(n_1, n_2)$ 的比例因此會增加。

(iii) 固定 ρ 值時，θ 越大，則 $\pi(\theta)$越大，一如預期。

表 2：

Gumbel normal density dim=10					
	$\theta = 0$	$\theta = 0.5$	$\theta = 1$	$\theta = 1.5$	$\theta = 2$
δ	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$
1.1	0.0467	0.1943	0.4551	0.7307	0.9041
1.5	0.0431	0.2873	0.6563	0.8877	0.9747
2	0.0443	0.4070	0.8133	0.9628	0.9946
2.5	0.0483	0.5434	0.9013	0.9884	0.9982

表 3：

Frank normal density dim=10					
	$\theta = 0$	$\theta = 0.5$	$\theta = 1$	$\theta = 1.5$	$\theta = 2$
δ	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$	$\pi(\theta)$
0.000001	0.0460	0.1710	0.3965	0.6672	0.8679
1	0.0463	0.1871	0.4588	0.7324	0.9083

3	0.0483	0.2538	0.5977	0.8560	0.9608
5	0.0489	0.3426	0.7384	0.9249	0.9862

從表 1、2、3 可看出，在可交換隨機變數的條件下，檢定力都大於獨立隨機樣本的結果，因此否定原始假設的比例會偏高。

考慮 i.i.d.連續型隨機變數 $X_1, X_2, \cdots, X_n$，令 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ 代表其順序統計量，則很容易可以用二個適合的順序統計量造出百分位數 π_p 之信賴區間，因為

$$P(X_{(i)} < \pi_p < X_{(j)}) = \sum_{k=i}^{j-1} \binom{n}{k} p^k (1-p)^{n-k} \text{ 。}$$

當 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ 的聯合機率密度函數為：

$$f_\alpha(\mathbf{x}) = \left[\prod_{i=1}^{n} f(x_i)\right] \cdot \left[1 + \alpha \prod_{i=1}^{n} \left(2F(x_i) - 1\right)\right] \quad , \ \alpha \in [-1,1], \alpha \neq 0$$

其邊際 p.d.f. 為 f，c.d.f. 為 F，則 $X_1, X_2, \cdots, X_n$ 只是可交換而非 i.i.d.，我們想探討在此分配下 $P\left(X_{(1)} < m < X_{(n)}\right)$ 和 α 值的關係，並和 i.i.d.的結果作比較。定義

$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}$，$X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$ 則

$$P\left(X_{(1)} < m < X_{(n)}\right) = \sum_{k=1}^{n-1} P\{\mathbf{X} \text{中有} k \text{個大於} m, (n-k) \text{個小於} m\}$$

$$= \sum_{k=1}^{n-1} C_k^n \underbrace{\int_m^\infty \int_m^\infty \cdots \int_m^\infty}_{k} \underbrace{\int_{-\infty}^m \int_{-\infty}^m \cdots \int_{-\infty}^m}_{n-k} f_\alpha(\mathbf{x}) dx_1 dx_2 \cdots dx_n$$

$$= \sum_{k=1}^{n-1} C_k^n \underbrace{\int_m^\infty \int_m^\infty \cdots \int_m^\infty}_{k} \underbrace{\int_{-\infty}^m \int_{-\infty}^m \cdots \int_{-\infty}^m}_{n-k} \prod_{i=1}^{n} f(x_i) dx_1 dx_2 \cdots dx_n$$

$$+\alpha \sum_{k=1}^{n-1} C_k^n \underbrace{\int_m^\infty \int_m^\infty \cdots \int_m^\infty}_{k} \underbrace{\int_{-\infty}^m \int_{-\infty}^m \cdots \int_{-\infty}^m}_{n-k} \prod_{i=1}^{n} f(x_i)\left(2F(x_i) - 1\right) dx_1 dx_2 \cdots dx_n$$

$$= \sum_{k=1}^{n-1} C_k^n \left(\frac{1}{2}\right)^n + \alpha \sum_{k=1}^{n-1} C_k^n \left(\frac{1}{4} \cdot 1\right)^k \left(\frac{1}{4} \cdot (-1)\right)^{n-k}$$

$$= \left[1 - C_0^n \left(\frac{1}{2}\right)^n - C_n^n \left(\frac{1}{2}\right)^n\right] + \left(\frac{1}{4}\right)^n \alpha \sum_{k=1}^{n-1} C_k^n (1)^k (-1)^{n-k}$$

$$= \left[1 - \frac{2}{2^n}\right] + \left(\frac{1}{2^{2n}}\right) \alpha \left(\sum_{k=0}^{n} C_k^n (1)^k (-1)^{n-k} - C_0^n (-1)^{n-0} - C_n^n (-1)^{n-n}\right)$$

$$= \left[1 - \frac{2}{2^n}\right] - \left(\frac{1}{2^{2n}}\right) \alpha \left((-1)^n + 1\right)$$

$$=\left[\frac{2^{n-1}-1}{2^{n-1}}\right]-\left(\frac{1}{2^{2n}}\right)\alpha\cdot 2\cdot I_A(n)$$

$$=\left[\frac{2^{n-1}-1}{2^{n-1}}\right]-\left(\frac{\alpha}{2^{2n-1}}\right)\cdot I_A(n)$$

其中

$$\int_m^{\infty} f(x_i)\left(2F(x_i)-1\right)dx_i=\frac{1}{4}\left(F(x_i)-1\right)^2\Big|_m^{\infty}=\frac{1}{4}\cdot 1$$

$$\int_{-\infty}^{m} f(x_i)\left(2F(x_i)-1\right)dx_i=\frac{1}{4}\left(2F(x_i)-1\right)^2\Big|_{-\infty}^{m}=\frac{1}{4}\cdot(-1)$$

I 是 indicator 函數，A 為所有正偶數之集合

$$I_A(n)=\begin{cases}1, n\in A\\ 0, n\notin A\end{cases}$$

一般在 i.i.d.的情況下，$X_{(1)}<m<X_{(n)}$ 的機率為 $\frac{2^{n-1}-1}{2^{n-1}}$；這個例子的聯合機率分配下 $X_{(1)}<m<X_{(n)}$ 的機率為 $\left[\frac{2^{n-1}-1}{2^{n-1}}\right]-\left(\frac{\alpha}{2^{2n-1}}\right)\cdot I_A(n)$，在 n 為偶數時，若 α 值越大很明顯可知 $P\left(X_{(1)}<m<X_{(n)}\right)$ 的值越小；n 為奇數時，$P\left(X_{(1)}<m<X_{(n)}\right)$ 值卻不受 α 值影響。

例：在本節所述的聯合分配假設下，我們考慮 $X_1, X_2, \cdots, X_n \sim U(0,1)$，$m=0.5$，則的聯合機率密度函數為：$f_\alpha(\mathbf{x})=\left[1+\alpha\prod_{i=1}^{n}(2x_i-1)\right], x_i\in(0,1) \quad for\ i=1,2,\cdots,n$

(i) 若 $n=2$，則 $f_\alpha(\mathbf{x})=\left[1+\alpha\prod_{i=1}^{2}(2x_i-1)\right], x_i\in(0,1) \quad for\ i=1,2$

$$P\left(X_{(1)}<m<X_{(2)}\right)=\sum_{k=1}^{2-1}P\{\mathbf{X}\text{中有}k\text{個大於}m,(2-k)\text{個小於}m\}$$

$$=\sum_{k=1}^{2-1}C_k^2\int_{0.5}^{1}\int_{0}^{0.5}f_\alpha(\boldsymbol{x})dx_1dx_2$$

$$=C_1^2\int_{0.5}^{1}\int_{0}^{0.5}f_\alpha(\boldsymbol{x})dx_1dx_2$$

$$=C_1^2\left[\int_{0.5}^{1}\int_{0}^{0.5}1dx_1dx_2+\alpha\int_{0.5}^{1}\int_{0}^{0.5}(2x_1-1)(2x_2-1)dx_1dx_2\right]$$

$$=2\left[\frac{1}{4}+\alpha\int_{0}^{0.5}(2x_1-1)dx_1\int_{0.5}^{1}(2x_2-1)dx_2\right]$$

$$= 2\left[\frac{1}{4} + \alpha\left(\frac{1}{4}(2x_2 - 1)^2\Big|_{0.5}^{1}\right)\left(\frac{1}{4}(2x_1 - 1)^2\Big|_{0}^{0.5}\right)\right]$$

$$= 2\left[\frac{1}{4} + \alpha\left(\frac{1}{4}(1)\right)\left(\frac{1}{4}(-1)\right)\right]$$

$$= \frac{2^{2-1} - 1}{2^{2-1}} - \left(\frac{\alpha}{2^{2\cdot 2-1}}\right)$$

則 α 值越大， $P\left(X_{(1)} < m < X_{(2)}\right)$ 越小

(ii) 若 $n = 3$ ，則 $f_\alpha(\mathbf{x}) = \left[1 + \alpha\prod_{i=1}^{3}(2x_i - 1)\right], x_i \in (0,1) \quad for\ i = 1,2,3$

$$P\left(X_{(1)} < m < X_{(3)}\right) = \sum_{k=1}^{3-1} P\{\mathbf{X}\text{中有}k\text{個大於}m, (3-k)\text{個小於}m\}$$

$$= C_1^3 \int_{0.5}^{1}\int_{0}^{0.5}\int_{0}^{0.5} f_\alpha(\boldsymbol{x})dx_1dx_2dx_3 + C_2^3 \int_{0.5}^{1}\int_{0.5}^{1}\int_{0}^{0.5} f_\alpha(\boldsymbol{x})dx_1dx_2dx_3$$

$$= C_1^3\left[\int_{0.5}^{1}\int_{0}^{0.5}\int_{0}^{0.5} 1dx_1dx_2dx_3 + \alpha\int_{0.5}^{1}\int_{0}^{0.5}\int_{0}^{0.5}(2x_1 - 1)(2x_2 - 1)(2x_3 - 1)dx_1dx_2dx_3\right]$$

$$+ C_2^3\left[\int_{0.5}^{1}\int_{0.5}^{1}\int_{0}^{0.5} 1dx_1dx_2dx_3 + \alpha\int_{0.5}^{1}\int_{0.5}^{1}\int_{0}^{0.5}(2x_1 - 1)(2x_2 - 1)(2x_3 - 1)dx_1dx_2dx_3\right]$$

$$= 3\left[\left(\frac{1}{2}\right)^3 + \alpha\left(\frac{1}{4}(1)\right)\left(\frac{1}{4}(-1)\right)\left(\frac{1}{4}(-1)\right)\right] + 3\left[\left(\frac{1}{2}\right)^3 + \alpha\left(\frac{1}{4}(1)\right)\left(\frac{1}{4}(1)\right)\left(\frac{1}{4}(-1)\right)\right]$$

$$= \frac{2^{3-1} - 1}{2^{3-1}}$$

則 $P\left(X_{(1)} < m < X_{(3)}\right)$ 不受 α 值影響。

綜上述 $P\left(X_{(1)} < m < X_{(n)}\right) = \left[\frac{2^{n-1} - 1}{2^{n-1}}\right] - \left(\frac{\alpha}{2^{2n-1}}\right)\cdot I_A(n)$ ， 在 n 為偶數時，若 α 值越大很明顯可知 $P\left(X_{(1)} < m < X_{(n)}\right)$ 的值越小，而 α 值越大表示 X_i 之間的相關性較大，此時 $X_1, X_2, \cdots, X_n$ 的數值越接近，因此 $X_{(1)} < m < X_{(n)}$ 的機會就變小了。

結果與討論

不論是用 Gumbel copula with normal marginals，或是用Frank copula with normal marginals 來產生只是可交換而並不互相獨立的隨機變數，再經過隨機分組後產生實驗組和對照組、用Wilcoxon rank-sum test來檢定二組的 mean 是否相同時，基本上

α值和設定的值很接近，但其檢定力和獨立隨機樣本的情況比起來都偏大。這種情況下所做出的結論，是否適用於母體，應有很大的討論空間。

當 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的聯合機率密度函數為：

$$f_\alpha(\mathbf{x})=\left[\prod_{i=1}^{n} f(x_i)\right]\cdot\left[1+\alpha\prod_{i=1}^{n}\left(2F(x_i)-1\right)\right] \quad ，\alpha\in[-1,1],\alpha\neq 0$$

其邊際 p.d.f. 為 f ，c.d.f. 為 F ，則 $X_1,X_2,\cdots,X_n$ 只是可交換而非 i.i.d. ，我們探討了在此分配下 $P\left(X_{(1)}<m<X_{(n)}\right)$ 和 α 值的關係，並和 i.i.d.的結果作比較。

$P\left(X_{(1)}<m<X_{(n)}\right)=\left[\frac{2^{n-1}-1}{2^{n-1}}\right]-\left(\frac{\alpha}{2^{2n-1}}\right)\cdot I_A(n)$，$I$ 是 indicator 函數，A 為所有正偶數之集合

$$I_A(n)=\begin{cases}1, n\in A\\ 0, n\notin A\end{cases}$$

在i.i.d.的情況下，$X_{(1)}<m<X_{(n)}$ 的機率為 $\frac{2^{n-1}-1}{2^{n-1}}$ ；所以 n 為偶數時，若 α 值越大很明顯可知 $P\left(X_{(1)}<m<X_{(n)}\right)$ 的值越小，而 α 值越大表示 X_i 之間的相關性較大，此時 $X_1,X_2,\cdots,X_n$ 的數值越接近，因此 $X_{(1)}<m<X_{(n)}$ 的機會就比i.i.d.時要小。

參考文獻

[1] Box,G.E.P., Hunter, J.S. and Hunter, W.G.(2005), Statistics for experimenters, 2nd edi., Wiley-Interscience,pp88-91

[2] Frees, Edward W. and Valdez, Emiliano A.(1998), Understanding relationships using copulas, North American Actuarial Journal, vol. 2, No. 1, pp1-25

[3]Gumbel, E.J.(1960), Bivariate exponential distributions, Journal of the American statistical association, Vol. 55, No. 292, pp698-707

[4]Hollander, M. and Wolfe, D.(1999), Nonparametric statistical methods , 2nd edi., Wiley-Interscience,pp364-377

[5]Joe H.(1997),Multivariate Models and Dependence Concepts, (Chapman & Hall/CRC).

[6] Langbehn,D.R., Berger,V., Higgins,J.J., Blair,R.C., Mallows,C.L., Ludbrook,J. and Dudley,H.(2000), Letter to the editor:"Ludbrook, J. and Dudley, H.(1998), Why

permutation tests are superior to t and F tests in biomedical research, The American Statistician, 52, 127-132 “, The American Statistician, vol. 54, Iss. 1, pp85-87

[7]Ludbrook, J. and Dudley, H.(1998), Why permutation tests are superior to t and F tests in biomedical research, The American Statistician, vol. 52, No. 2, pp127-132

[8]Nelsen, Roger B.(1999), An Introduction to Copulas, (Springer).

計畫成果自評

所做出的結果與原計畫內容大致相符，但有些想做的問題做不出來，結果不盡理想。