中國統計學報
第33卷第3期
八十四年九月
373–393頁

共同種類數之估計–客雅溪口與中港溪口共同鳥種分析

陳俞成 黃文瀚 趙蓮菊 郭承裕

新竹市清華大學統計研究所 新竹市野鳥學會

摘 要

假設在二個族群(或母體)中分別取了一組多項樣本,並記錄每一個觀察之種類。本文利用樣本涵蓋的觀念估計二個族群之共同種類數,並以統計模擬分析說明所提出之估計量的表現。本文的結果推廣了 Chao and Lee (1992) 單一族群至二個族群的情形,並以台灣新竹客雅溪口與苗栗中港溪口二地之1994年4月至1995年3月之鳥類調查數據為例,以估計此二溪口共同之鳥種數目。

關鍵詞: 多項式模式,差異性,樣本涵蓋率,變異係數。

美國數學會分類索引: 主要62P10;次要62F10。

1. 前言

首先敘述本研究之動機。政府計畫以填海造陸方式將新竹香山區周圍之竹東丘陵,移填香山潮間帶,填土面積高達1025公頃。各界質疑其影響海岸生

態至巨而要求進行嚴格之環境影響評估。新竹市野鳥學會受中華顧問工程公司委託,辦理新竹香山區海埔地造地開發計畫環境影響評估中之有關鳥類調查部分,而於1994年4月至1995年3月,在新竹頭前溪至苗栗中港溪的濱海地區及香山取土區,共進行了237次之鳥類調查,以上細節請參考新竹市野鳥學會所撰寫之有關報告。在所調查的區域中包含了新竹客雅溪口附近地區及苗栗中港溪口地區,尤其是客雅溪口潮間帶,早期一直是台灣重要的溼地之一,面積廣闊,除了提供漁民豐富的蚵、貝產量外,上億隻的螃蟹和豐富的鳥類資源,也使得此區成爲台灣難得倖存的自然景觀區。然而近年來人爲的污染,水泥堤的興建,已使鳥況變差,而計畫中的西濱快速道路預定穿過此景觀區,此將對本區造成嚴重的生態破壞。中港溪污染嚴重,底棲生物較少,鳥類數量亦較客雅溪口少。在新竹野鳥學會所各65次調查中客雅溪口共有155種鳥種出現而中港溪口有140種鳥種。由於此二溪口距離在20公里之內,有許多鳥種均相同,在此次調查數據發現有111種共同鳥種。

一個族群(或母體)之種類數估計在生物學或生態學上是一個古典的問題,在文獻上也有廣泛的討論。它的應用不僅在生物、生態方面,在其它方面諸如語言文字學(估計作者的字彙數)、經濟學(估計古代錢幣鑄模數目),軟體測試(估計軟體的錯誤數目)上,也有一些實際且有趣的應用。可參考Bunge and Fitzpatrick (1993)之回顧論文。

在過去的文獻中處理種類數的估計多半針對單一族群,而估計的方法不論是從頻率論或從貝氏理論的眼光來看都應有盡有,方式繁多。作者之一過去數年來主要採用樣本涵蓋估計法。基本的觀念在於若直接去估計種類數是很難估計準確的,但是“樣本涵蓋”卻可估的極爲精準,因此採用“樣本涵蓋”來間接估計種類數。事實上樣本涵蓋最早的估計可溯至計算機科學之始祖Turing (參考Good (1953)之歷史說明)。在Chao and Lee (1992)之論文中我們將此方法正式應用至種類數之估計。

在生態學上,二個族群的比較,關係生態體系的均衡與競爭,例如,比較不同海拔高度鳥類的分布情形(見Colwell and Coddington (1994), Colwell (1973), Feinsinger (1976), Karr et al. (1990), Coddington et al. (1991)),或比較人爲破壞後自然的恢復情形,例如比較油輪污染海面之前與之後二個時間種類數有無顯著減少(見Grassle and Smith (1976))。

在新竹市野鳥學會所作的一年間鳥類調查中，涵蓋了中港溪口與客雅溪口大部份地區。此二地區生態環境極爲類似，但一般而言，如前所述客雅溪口環境品質較佳。就鳥類棲息而言，如何比較此二溪口的族群？過去對二個族群的比較通常採用"相似指數"(similarity)(見 Grassle and Smith (1976) 或參考 Pielou (1975, 1977)) 或"互補性"(complementarity)(見 Colwell and Coddington (1994))，但這些指數都必須利用到共同種類的數目。過去文獻上的分析都是以觀察到的共同種即認爲是所有二族群之共同種。是否有共同鳥種並未被觀察到？對二族群共同種類數之統計估計，據作者所知在文獻上幾乎尚無討論。本文即利用樣本涵蓋法來估計二個族群之共同種類數。

在第2節中我們介紹模式及符號，在第3節中導出共同種類數之估計式，而其一般表現則在第4節中利用模擬分析來說明，第5節將使用前述新竹野鳥學會所蒐集的數據作爲本文應用之實例，第6節爲後記。

2. 模式與符號說明

假設有二個族群，此二族群可能是在不同地區的族群，或不同生態環境的二個族群。我們由各自族群中分別抽取一組多項樣本，且此二樣本相互獨立。假設第一族群種類數爲 N_1，而第二族群種類數爲 N_2。令 p_i, p_i^* 分別爲觀察到第一及第二族群中第 i 類之機率，而樣本大小分別爲 n_1 及 n_2。令 X_i, Y_i 分別是在第一及第二族群取樣中觀察到第 i 類的個數，則 $(X_1, X_2, \cdots, X_{N_1})$ 及 $(Y_1, Y_2, \cdots, Y_{N_2})$ 分別爲 $(n_1, p_1, p_2, \cdots, p_{N_1})$ 及 $(n_2, p_1^*, p_2^*, \cdots, p_{N_2}^*)$ 的多項分布，往後我們將簡稱其爲樣本一及樣本二，並令 $P = (p_1, p_2, \cdots, p_{N_1})$, $P^* = (p_1^*, p_2^*, \cdots, p_{N_2}^*)$。假設此二族群之共同種類數爲 N_{12}，不失一般性，我們將假設其爲 $(X_1, X_2, \cdots, X_{N_1})$ 及 $(Y_1, Y_2, \cdots, Y_{N_2})$ 之首 N_{12} 個種類；而於其中觀察到相異種類數爲 M_{12}，即 $M_{12} = \sum_{i=1}^{N_{12}} I[X_i \geq 1, Y_i \geq 1]$，則有下列統計量 $(X_1', X_2', \cdots, X_{M_{12}}')$ 及 $(Y_1', Y_2', \cdots, Y_{M_{12}}')$，其中 X_i', Y_i' 分別表示兩族群共同種類同時出現在樣本一及樣本二中之第 i 類的個數。令 f_1 表共同種類在樣本二至少出現一次，而在樣本一僅出現一次之種類數，同樣定義 f_1^* 爲共同種類在樣本一至少出現一次而在樣本二僅出現一次之種類數，而定義 f_{11} 爲共同種類同時在樣本一及樣本二均出現一次之種類數。對 P 及 P^* 有兩種模式：

(1)固定效應模式(fixed-effects model)

視P及P^*爲固定之參數,$p_1, p_2, \cdots, p_{N_{12}}$具有平均值$\bar{p} = \sum_{i=1}^{N_{12}} p_i/N_{12}$, p_1^*, $p_2^*, \cdots, p_{N_{12}}^*$具有平均值$\bar{p}^* = \sum_{i=1}^{N_{12}} p_i^*/N_{12}$。令$Q = (Q_1, Q_2, \cdots, Q_{N_{12}})$, $Q_i = p_i p_i^*$且定義

$$P\text{ 及 }Q\text{ 之變異係數 } CV_1^* \text{ 爲 } CV_1^* = \frac{\frac{1}{N_{12}}\sum_{i=1}^{N_{12}}(p_i - \bar{p})(Q_i - \bar{Q})}{\bar{p}\bar{Q}},$$

$$P^*\text{ 及 }Q\text{ 之變異係數} CV_2^* \text{ 爲 } CV_2^* = \frac{\frac{1}{N_{12}}\sum_{i=1}^{N_{12}}(p_i^* - \bar{p}^*)(Q_i - \bar{Q})}{\bar{p}^*\bar{Q}},$$

$$P, P^*, Q\text{ 之變異係數 } CV_{12}^* \text{ 爲 } CV_{12}^* = \frac{\frac{1}{N_{12}}\sum_{i=1}^{N_{12}}(p_i - \bar{p})(p_i^* - \bar{p}^*)(Q_i - \bar{Q})}{\bar{p}\bar{p}^*\bar{Q}}\text{。}$$

(2)隨機效應模式(random-effects model)

假設(p_i, p_i^*), $i = 1, 2, \cdots, N_{12}$,具有相同之二維分布$F(p, p^*)$,其邊際分布各爲$F_1(p)$與$F_2(p^*)$,令$p_0 = \int p\,dF_1(p)$, $p_0^* = \int p^* dF_2(p^*)$。令$Q_i = p_i p_i^*$, $Q_0 = \int pp^* dF(p, p^*)$,且定義

$$\begin{aligned}
\gamma_1^2 &= \int (p - p_0)^2 dF_1(p)/p_0^2,\\
\gamma_2^2 &= \int (p^* - p_0^*)^2 dF_2(p^*)/(p_0^*)^2,\\
CV_1^* &= \frac{\int (p - p_0)(pp^* - Q_0)dF(p, p^*)}{p_0 Q_0},\\
CV_2^* &= \frac{\int (p^* - p_0^*)(pp^* - Q_0)dF(p, p^*)}{p_0^* Q_0},\\
CV_{12}^* &= \frac{\int (p - p_0)(p^* - p_0^*)(pp^* - Q_0)dF(p, p^*)}{p_0 p_0^* Q_0}\text{。}
\end{aligned}$$

3. 估計量及其收斂性

3.1. 估計量

在本節中,我們主要以固定效應模式來說明估計導證過程。首先定義樣本涵蓋C_{12}爲

$$C_{12} = \frac{\sum_{i \in N_{12}} p_i p_i^* I[\text{第 } i \text{ 個共同種類在樣本一和樣本二均至少出現一次}]}{\sum_{i \in N_{12}} p_i p_i^*} \text{ 。}$$

若假設 $p_1 p_1^* = p_2 p_2^* = \cdots = p_{N_{12}} p_{N_{12}}^*$，則 $C_{12} = M_{12}/N_{12}$ 。在此等機率的假設下，對 N_{12} 來說，可得一個很自然的估計量

$$\widehat{N}_{12}^o = M_{12}/\widehat{C}_{12}, \tag{3.1}$$

此處 $\widehat{C}_{12}$ 是 C_{12} 的估計量，此式將在 (3.5) 中列出 。

當 $p_i p_i^*$, $i = 1, 2, \cdots, N_{12}$, 間不等時，我們可藉著使用估計樣本涵蓋的方法來估計共同種類數，最主要的關鍵是導出當族群內的種類機率具有變異性時，EM_{12}/EC_{12} 與 N_{12} 間的差別 。在多項式分布下，我們可得

$$EM_{12} = \sum_{i=1}^{N_{12}} [1 - (1 - p_i)^{n_1}][1 - (1 - p_i^*)^{n_2}],$$

及

$$EC_{12} = 1 - \frac{\sum_{i=1}^{N_{12}} p_i p_i^* [(1 - p_i)^{n_1} + (1 - p_i^*)^{n_2} - (1 - p_i)^{n_1}(1 - p_i^*)^{n_2}]}{\sum_{i=1}^{N_{12}} p_i p_i^*} \text{ 。} \tag{3.2}$$

因此進一步至二階展式可得

$$N_{12} \cong \frac{EM_{12}}{EC_{12}} + \frac{1}{EC_{12}} [(Ef_1) \cdot CV_1^* + (Ef_1^*) \cdot CV_2^* + (Ef_{11}) \cdot CV_{12}^*] \text{ 。} \tag{3.3}$$

首先估計 EC_{12}，因為

$$\begin{aligned} EC_{12} &= 1 - \frac{\sum_{i=1}^{N_{12}} p_i p_i^* [(1 - p_i)^{n_1} + (1 - p_i^*)^{n_2} - (1 - p_i)^{n_1}(1 - p_i^*)^{n_2}]}{\sum_{i=1}^{N_{12}} p_i p_i^*} \\ &\approx 1 - \frac{\sum_{i=1}^{N_{12}} p_i p_i^* [(1 - p_i)^{n_1 - 1} + (1 - p_i^*)^{n_2 - 1} - (1 - p_i)^{n_1 - 1}(1 - p_i^*)^{n_2 - 1}]}{\sum_{i=1}^{N_{12}} p_i p_i^*}, \end{aligned} \tag{3.4}$$

且

$$E(\sum_{i=1}^{M_{12}} X_i' Y_i') = n_1 n_2 \sum_{i=1}^{N_{12}} p_i p_i^*,$$

$$E(\sum_{i=1}^{M_{12}} I[X_i' = 1] Y_i') = n_1 n_2 \sum_{i=1}^{N_{12}} p_i (1 - p_i)^{n_1 - 1} p_i^*,$$

$$E(\sum_{i=1}^{M_{12}} I[Y_i' = 1] X_i') = n_1 n_2 \sum_{i=1}^{N_{12}} p_i^* (1 - p_i^*)^{n_2 - 1} p_i,$$

$$E(\sum_{i=1}^{M_{12}} I[X_i' = Y_i' = 1]) = n_1 n_2 \sum_{i=1}^{N_{12}} p_i p_i^* (1 - p_i)^{n_1 - 1} (1 - p_i^*)^{n_2 - 1} \text{ 。}$$

由此得出 EC_{12} 的估計量為

$$\widehat{C}_{12} = 1 - \frac{\sum_{i=1}^{M_{12}} I(X_i' = 1)Y_i' + \sum_{i=1}^{M_{12}} I(Y_i' = 1)X_i' - \sum_{i=1}^{M_{12}} I(X_i' = 1, Y_i' = 1)}{\sum_{i=1}^{M_{12}} X_i'Y_i'} \text{。} \tag{3.5}$$

其次再估計變異係數 CV_1^*, CV_2^* 及 CV_{12}^*。將 CV_1^*, CV_2^* 及 CV_{12}^* 改寫成

$$CV_1^* = \frac{N_{12}\sum_{i=1}^{N_{12}} p_i^2 p_i^*}{(\sum_{i=1}^{N_{12}} p_i)(\sum_{i=1}^{N_{12}} p_i p_i^*)} - 1, \tag{3.6}$$

$$CV_2^* = \frac{N_{12}\sum_{i=1}^{N_{12}} p_i^{*2} p_i}{(\sum_{i=1}^{N_{12}} p_i^*)(\sum_{i=1}^{N_{12}} p_i p_i^*)} - 1, \tag{3.7}$$

$$CV_{12}^* = \frac{N_{12}^2\sum_{i=1}^{N_{12}} p_i^2 p_i^{*2}}{(\sum_{i=1}^{N_{12}} p_i)(\sum_{i=1}^{N_{12}} p_i^*)(\sum_{i=1}^{N_{12}} p_i p_i^*)} - \frac{N_{12}\sum_{i=1}^{N_{12}} p_i p_i^*}{(\sum_{i=1}^{N_{12}} p_i)(\sum_{i=1}^{N_{12}} p_i^*)} - CV_1^* - CV_2^* \text{。} \tag{3.8}$$

而我們有

$$\begin{aligned}
E\sum_{i=1}^{M_{12}} X_i' &= n_1 \sum_{1}^{N_{12}} p_i[1 - (1 - p_i^*)^{n_2}] \approx n_1 \sum_{i=1}^{N_{12}} p_i, \\
E\sum_{i=1}^{M_{12}} Y_i' &= n_2 \sum_{1}^{N_{12}} p_i^*[1 - (1 - p_i)^{n_1}] \approx n_2 \sum_{i=1}^{N_{12}} p_i^*, \\
E\sum_{i=1}^{M_{12}} X_i'Y_i' &= n_1 n_2 \sum_{i=1}^{N_{12}} p_i p_i^*, \\
E\sum_{i=1}^{M_{12}} X_i'(X_i' - 1)Y_i' &= n_1(n_1 - 1)n_2 \sum_{i=1}^{N_{12}} p_i^2 p_i^*, \\
E\sum_{i=1}^{M_{12}} Y_i'(Y_i' - 1)X_i' &= n_2(n_2 - 1)n_1 \sum_{i=1}^{N_{12}} p_i^{*2} p_i, \\
E\sum_{i=1}^{M_{12}} X_i'(X_i' - 1)Y_i'(Y_i' - 1) &= n_1(n_1 - 1)n_2(n_2 - 1) \sum_{i=1}^{N_{12}} p_i^2 p_i^{*2} \text{。}
\end{aligned}$$

在得到 CV_1^*、CV_2^* 及 CV_{12}^* 的估計量時,須先給 (3.6)、(3.7) 和 (3.8) 中的 N_{12} 一個粗略估計值,我們利用 M_{12} 作為此估計值。因此,我們有下列的變異係數估計量

$$\widehat{CV}_1^* = \frac{M_{12} n_1 \sum_{i=1}^{M_{12}} X_i'(X_i' - 1)Y_i'}{(n_1 - 1)(\sum_{i=1}^{M_{12}} X_i')(\sum_{i=1}^{M_{12}} X_i'Y_i')} - 1, \tag{3.9}$$

$$\widehat{CV}_2^* = \frac{M_{12}n_2\sum_{i=1}^{M_{12}} Y_i'(Y_i'-1)X_i'}{(n_2-1)(\sum_{i=1}^{M_{12}} Y_i')((\sum_{i=1}^{M_{12}} X_i'Y_i')} - 1, \tag{3.10}$$

$$\widehat{CV}_{12}^* = \frac{M_{12}^2 n_1 n_2 \sum_{i=1}^{M_{12}} X_i'(X_i'-1)Y_i'(Y_i'-1)}{(n_1-1)(n_2-1)(\sum_{i=1}^{M_{12}} X_i')(\sum_{i=1}^{M_{12}} Y_i')(\sum_{i=1}^{M_{12}} X_i'Y_i')} - \frac{M_{12}\sum_{i=1}^{M_{12}} X_i'Y_i'}{(\sum_{i=1}^{M_{12}} X_i')(\sum_{i=1}^{M_{12}} Y_i')} - \widehat{CV}_1^* - \widehat{CV}_2^* \text{。} \tag{3.11}$$

在(3.3)中,各項變異係數以(3.9)、(3.10)及(3.11)代入,樣本涵蓋以(3.5)代入,並將各個期望值分別以觀察值代入,我們提出的估計量如下:

$$\widehat{N}_{12} = \frac{M_{12}}{\widehat{C}_{12}} + \frac{1}{\widehat{C}_{12}}[f_1\widehat{CV}_1^* + f_1^*\widehat{CV}_2^* + f_{11}\widehat{CV}_{12}^*] \text{。} \tag{3.12}$$

在隨機效應模式中,所有導證的過程均很類似,只將過去之和的形式改爲積分形式,惟在(3.2)式之等式需換成近似式,即同樣的方法可以得到與固定模式完全相同之估計量 $\widehat{N}_{12}$ 。

在特別的情況下,假設兩群體之種類機率分佈彼此互相獨立,亦即 $F(p,p^*) = F_1(p)F_2(p^*)$ 時,估計量 $\hat{N}_{12}$ 可簡化爲:

$$\widehat{N}_{12}^* = \frac{M_{12}}{\widehat{C}_{12}} + \frac{1}{\widehat{C}_{12}}\{f_1\hat{\gamma}_1^2 + f_1^*\hat{\gamma}_2^2 + f_{11}\hat{\gamma}_1^2\hat{\gamma}_2^2\}, \tag{3.13}$$

其中

$$\hat{\gamma}_1^2 = \max\{\frac{M_{12}n_1\sum X_i'(X_i'-1)}{(n_1-1)(\sum X_i')^2} - 1, 0\}, \tag{3.14}$$

$$\hat{\gamma}_2^2 = \max\{\frac{M_{12}n_2\sum Y_i'(Y_i'-1)}{(n_2-1)(\sum Y_i')^2} - 1, 0\} \text{。} \tag{3.15}$$

在研究中發現所提出之估計量中的變異係數項($CV_i^*, i = 1,2,12$)不容易得到準確的估計 。一般情況下, CV_i^* 的估計多半會低估 。但若某些較爲普通之種類出現許多次時, CV_i^* 之估計值又變成高估太多,而造成了估計值表現很差 。因此對有出現頻率次數很高的資料,我們使用一個調整估計 CV_i^* 的方法,就是利用出現次數爲1到 k 次的資料作估計,根據(3.1)、(3.12)及(3.13)得到 $\widehat{N}_{12}^o$, $\widehat{N}_{12}$, $\widehat{N}_{12}^*$,再將重覆捕取次數至少有一樣本大於 k 次的種類加入估計量而得:

$$\widetilde{N}_{12}^o = \sum_{i=1}^{M_{12}} I(X_i' > k \text{ or } Y_i' > k) + \widehat{N}_{12}^o, \tag{3.16}$$

$$\widetilde{N}_{12} = \sum_{i=1}^{M_{12}} I(X_i' > k \text{ or } Y_i' > k) + \widehat{N}_{12}, \tag{3.17}$$

$$\widetilde{N}_{12}^{*} = \sum_{i=1}^{M_{12}} I(X_i' > k \text{ or } Y_i' > k) + \widehat{N}_{12}^{*} \text{ 。} \tag{3.18}$$

k 值大小的擇定對結果有些影響,經驗告訴我們 $k = 10$ 是一個合理的選擇(參考 Chao et al. (1993)) 。

我們採用重抽法(bootstrap resampling)來估計各估計量的變異數 。以估計量 $\widehat{N}_{12}$ 爲例,首先假設每一組 (X_i', Y_i'), $i = 1, \cdots, M_{12}$, 中每對出現的機率爲 $1/\widehat{N}_{12}$, (0,0) 出現的機率爲 $1 - M_{12}/\widehat{N}_{12}$, 重新抽取一組大小爲 M_{12} 的歸還重抽樣本(a bootstrap sample), 得到 $(X_i'^{*}, Y_i'^{*})$, $i = 1, \cdots, M_{12}^{*}$, 其中 $X_i'^{*} Y_i'^{*} \neq 0$, 據此計算出 $\widehat{N}^{*b}$ 。對每一組資料重複生成 B 次,得到 B 個重抽法估計值 $\widehat{N}^{*1}$, $\widehat{N}^{*2}, \cdots, \widehat{N}^{*B}$ 。此 B 個 $\widehat{N}^{*i}$ 的樣本變異數即爲 $\widehat{N}_{12}$ 的重抽法變異數,亦即

$$\widehat{\text{Var}}(\widehat{N}_{12}) = [\sum_{i=1}^{B} (\widehat{N}^{*i})^2 - (\sum_{i=1}^{B} \widehat{N}^{*i})^2/B]^2/(B-1) \text{ 。}$$

在第4節之模擬分析中可知,由重抽法計算出的標準差之平均估計值,在大部分的情況下都很接近樣本標準差 。

3.2. 估計量之收斂性

根據 Chen (1980) 的結果,所提出的估計量在某些情況下具一致性 。首先我們需要下列引理:

<u>引理 3.1:</u> 在隨機效應模式下,假如種類機率分布 $(p_1, p_2, \cdots, p_{N_1})$ 及 $(p_1^*, p_2^*, \cdots, p_{N_2}^*)$ 分別服從一含參數 α_1 及 α_2 且對稱的 Dirichlet 分配 。當 $\lim_{N_1 \to \infty} n_1/N_1 = \lambda_1 > 0$, $\lim_{N_2 \to \infty} n_2/N_2 = \lambda_2 > 0$, 時,並假設上述二分配彼此互相獨立 。令 $a_1 = \alpha_1(\lambda_1 + \alpha_1)^{-1}$, $b_1 = 1 - a_1$, $a_2 = \alpha_2(\lambda_2 + \alpha_2)^{-1}$, $b_2 = 1 - a_2$ 。則對 j, $k = 0, 1, \cdots$,

$$\frac{\sum_{i=1}^{N_{12}} I[X_i = j] I[Y_i = k]}{N_{12}} \xrightarrow{P} \frac{\Gamma(j + \alpha_1)}{\Gamma(j+1)\Gamma(\alpha_1)} b_1^j a_1^{\alpha_1} \frac{\Gamma(k + \alpha_2)}{\Gamma(k+1)\Gamma(\alpha_2)} b_2^k a_2^{\alpha_2} \text{ 。}$$

此引理之證明係利用 Chen (1980) 文章中的 Lemma 2, 可以得到

$$E\left(\frac{\sum_{i=1}^{N_{12}} I[X_i=j]I[Y_i=k]}{N_{12}} - \frac{\Gamma(j+\alpha_1)}{\Gamma(j+1)\Gamma(\alpha_1)} b_1^j a_1^{\alpha_1} \frac{\Gamma(k+\alpha_2)}{\Gamma(k+1)\Gamma(\alpha_2)} b_2^k a_2^{\alpha_2}\right)^2 \to 0,$$

而上述引理即得證。利用此引理，可得下列之定理（證明省略）。

定理 3.1: 在引理 3.1 及 α_1, α_2 已知的條件下，則我們有 $CV_1^* = \gamma_1^2 \approx \frac{1}{\alpha_1}$, $CV_2^* = \gamma_2^2 \approx \alpha_2^{-1}$ 且

$$N_{12}^{-1}\left[\frac{M_{12}}{\widehat{C}_{12}} + \frac{1}{\widehat{C}_{12}}(f_1 CV_1^* + f_1^* CV_2^* + f_{11} CV_{12}^*)\right] \xrightarrow{P} 1, \tag{3.19}$$

$$N_{12}^{-1}\left[\frac{M_{12}}{\widehat{C}_{12}} + \frac{1}{\widehat{C}_{12}}(f_1 \gamma_1^2 + f_1^* \gamma_2^2 + f_{11}\gamma_1^2\gamma_2^2)\right] \xrightarrow{P} 1 \text{ 。} \tag{3.20}$$

若自 Dirichlet 分配中取出機率後，即加以固定，成爲固定效應模式，我們可以證明在固定效應下之 CV_1^*, CV_2^*, CV_{12}^* 分別收斂至 α_1^{-1}, α_2^{-1}, $(\alpha_1\alpha_2)^{-1}$。因此，不論是在固定或隨機效應模式下，利用引理 3.1 的結果，可以證得所提出的估計式均具有一致性。而對於不考慮 CV_i^* 之值的估計量，在定理 3.1 的條件下，則有下列之結果：

$$N_{12}^{-1}\left[\frac{M_{12}}{\widehat{C}_{12}}\right] \xrightarrow{P} \frac{(1-a_1^{\alpha_1})(1-a_2^{\alpha_2})}{(1-a_1^{\alpha_1+1})(1-a_2^{\alpha_2+1})} \text{ 。} \tag{3.21}$$

4. 模擬分析

本節將利用電腦模擬研究來探討所提出之估計式的表現。在此實驗中我們取 $N_1 = N_2 = 100$, $N_{12} = 40, 80$ 等情形。由於 Zipf's 法則的形式在族群種類分配的模型上被廣泛使用，在此我們亦考慮種類機率分配 p_i 具 Zipf's 法則形式的分佈，其中 CV 代表該母體之變異係數 $CV^2 = [\sum_{i=1}^N (p_i - \bar{p})^2/N]/\bar{p}^2$, $\bar{p} = \sum_{i=1}^N p_i/N$，而討論下列四種固定效應模式情形。

I: $CV = 0, P_i = 0.01, i = 1, 2, \cdots, 100$ 。

II: $CV = 0.75, P_i = c/(i+10), i = 1, 2, \cdots, 100, c = 1/(\sum_{i=1}^{100} 1/(i+10))$ 。

III: $CV = 0.75, P_i = c/(111-i), i = 1, 2, \cdots, 100, c = 1/(\sum_{i=1}^{100} 1/(111-i))$ 。

IV: $CV = 2.25, P_i = c/i, i = 1, 2, \cdots, 100, c = 1/(\sum_{i=1}^{100} 1/i)$ 。

母體 II 與母體 III 有相同之 p_i 值及 CV 值，但順序不同。由於我們假設共同之種類數爲前 N_{12} 個，故 p_i 的順序會影響共同種類數的結構。例如，表中

'I v.s. II' 表示兩母體分別爲 I 與 II, 但共同種類爲母體 II 中較容易出現的種類。而表中 'I v.s. III' 則表示兩母體分別爲 I 與 III, 但共同種爲母體 II 中較稀有的種類。我們考慮所有此四種母體的組合, 共有 I v.s. I, I v.s. II, I v.s. III, I v.s. IV, II v.s. II, $\cdots$ 等十種情形。再對每一種母體組合分別作 $(n_1,\ n_2) = (200, 200)$, 及 $(n_1,\ n_2) = (300, 300)$ 二種多項取樣。對固定之二母體及樣本數, 分別作了 200 次之模擬, 每次模擬均計算四種估計量, M_{12}, $\tilde{N}^o_{12}$, $\tilde{N}_{12}$ 及 $\tilde{N}^*_{12}$, 及後三者之重抽法標準差估計值, 此 200 組估計量之平均、標準差之平均及偏差平均記錄之。最後再利用此 200 組估計值求得樣本標準差及樣本均方根差 (root mean squared error, 簡稱 RMSE)。對 $(n_1, n_2) = (200, 200)$ 之模擬結果列於表 1 中, 而對 $(n_1, n_2) = (300, 300)$ 之結果列於表 2 中。

假如我們僅以觀察到的種類數直接作爲估計量, 則其將嚴重低估真正的共同種類數是可以預期的。因此在表 1 及表 2 中, M_{12} 有最大的偏差及均方根差, 故此種傳統直觀的作法必須加以改正。如果我們不考慮捕取機率間之差異, 而以 $\tilde{N}^o_{12}$ 作爲估計量, 其在 CV_1^*, CV_2^*, CV_{12}^* 值均較小時例如, I v.s. I, I v.s. II, 及 I v.s. III 之情況。如表 1 中當資料較少以致無法得到 CV_i^* 之穩定估計量時, $\tilde{N}^o_{12}$ 有較佳的表現。此情況下假設機率相等與假設機率爲相互獨立而得到的 $\tilde{N}^*_{12}$ 表現相仿。但若 CV_i^* 中至少有一個值較大時, 我們所提出之估計量 $\tilde{N}_{12}$ 即有最佳的表現。在表 2 中, 當資料足夠充分以估計 CV_i^* 時, $\tilde{N}_{12}$ 一般有最小的偏差及均方根差。同時在表 2 中, $\tilde{N}^*_{12}$ 雖均較 $\tilde{N}^o_{12}$ 爲佳, 但不如 $\tilde{N}_{12}$ 之表現。因此若資料充足且種類機率間之差異夠大時, 必須考慮估計 CV_i^* 而採用 $\tilde{N}_{12}$。

在表 3 中, 我們探討兩組機率分布爲獨立情況之隨機效應模式, 並且對第 3 節中之收斂性以數值結果來驗證。假設二組母體機率 $(p_1, p_2, \cdots, p_{N_1})$ 與 $(p_1^*, p_2^*, \cdots, p_{N_2}^*)$ 分別服從參數 α_1 及 α_2 之對稱的 Dirichlet 分布, $\alpha_1 = \alpha_2 = 0.5$ 或 1, N_1, N_2 之值從 100 增加到 1000, N_{12} 固定爲 N_1, N_2 之 40% 及 80%, 觀察總數 (n_1, n_2) 分別爲 N_1, N_2 之兩倍 (即 $\lambda_1 = \lambda_2 = 2$)。如表 1、表 2 中之模擬, 在表 3 的實驗中亦重復 200 次, 分別計算四個估計量, 同時並列出一個 "理論值", 即假設兩組機率之變異係數 γ_1^2 及 γ_2^2 已知的理想情況下所計算出之估計值 $\tilde{N}^t_{12}$, 比值 "ratio" 爲估計值與真正 N_{12} 之比值。由表 3 $\tilde{N}^t_{12}$ 之數值比值即可了解第 3 節所討論的收斂性質。但若 γ_1^2 及 γ_2^2 未知而必須估計時, 則所有估計值均低估。特別當 $\gamma_1^2 = \gamma_2^2 \cong 2$ 時 (即 $\alpha_1 = \alpha_2 = 0.5$), 由於觀察到的 M_{12} 共同種類數很少, 故

偏差很大。但表3顯示了一個很重要的訊息，即是當二組機率爲獨立時，在獨立條件所導出之估計量 $\tilde{N}_{12}^{*}$，與一般未假設獨立情況下之 $\tilde{N}_{12}$ 表現差不多，因此即使在獨立情況下，亦可採用 $\tilde{N}_{12}$。由以上的討論，可知不論固定效應或隨機效應下，只要資料充足且至少一組捕取機率變異夠大時，$\tilde{N}_{12}$ 是最佳的選擇。

5. 實例

新竹野鳥學會於1994年4月至1995年3月，在新竹頭前－客雅區(以下簡稱客雅溪口)及崎頂－中港區(以下簡稱中港溪口)，調查記錄相異鳥種各有155及140種，在客雅溪口出現超過10次有96種，不超過10次有59種，在中港溪口有81種超過10次，不超過10次有59種。若以客雅溪口單一族群來分析，其出現次數1次至10次的數目依序爲25, 6, 10, 3, 4, 4, 1, 1, 4, 1。對此一單一族群，我們根據Chao and Lee (1992)之分析，可得約有26種類未觀察到，即全區的種類數估計約爲181±11。而在中港溪口出現次數1次至10次的數目依序爲22, 12, 4, 4, 6, 2, 1, 2, 3, 3。約有23種類未被觀察到，因此全區種類數之估計爲163±10。

此二地區觀察到之共同種類數共有111種，在此111種中至少在一地區出現10次以上的共同種有90種。因此我們針對在二樣本出現均不高於10次的21種鳥種來分析。此21種鳥種在客雅溪口之觀察總數 $n_1 = 76$，而在中港溪口觀察總數 $n_2 = 64$。共同種中在客雅溪口僅出現1次且在中港溪至少出現1次有8種，在中港溪口僅出現1次且在客雅溪至少出現1次亦有9種，而在兩地各出現1次的共同種有4種。由(3.5)式可得樣本涵蓋率的估計值爲 $\hat{C}_{12} = 0.85965$，利用(3.9)-(3.11)可得 $\widehat{CV}_1^* = 0.49$, $\widehat{CV}_2^* = 0.72$, $\widehat{CV}_{12}^* = 0.23$，而由(3.12)得到 $\tilde{N}_{12}$ 之估計量及200次重抽法所得標準差估計量約爲 $\tilde{N}_{12} = 128 \pm 11.6$。若我們假設二組種類機率分布爲互相獨立，由(3.13)至(3.15)可得 $\hat{\gamma}_1^2 = 0.36$, $\hat{\gamma}_2^2 = 0.45$，及 $\tilde{N}_{12}^* = 123 \pm 7.0$。此與不假設獨立的結果相差不遠。由以上的分析，我們的結論是尚有約17種的共同種沒有觀察到。

6. 後記

本文所提出的方法雖限於二個族群的比較,但應可推廣至超過二個族群之情形。又所提估計量之收斂性在異於 Dirichlet 分配時之表現仍有待探討。附帶一提的是,新竹沿海地帶鳥種豐富,填海造陸的規劃將使生態環境完全改觀,原棲息在樹林和草叢的鳥類消失,過境猛禽失去一個停棲的區域。而生活在潮間帶的生物,則完全沒有其他緩衝帶可讓其生存,對鳥類和潮間帶生物而言,是一種無法回復的徹底破壞。

表 1. 各估計量之比較, $(n_1, n_2) = (200, 200)$

est. 代表模擬 200 次各估計量之平均
$\widehat{s.e.}$ 代表模擬 200 次各估計量 bootstrap 標準差估計值之平均
s.e. 代表模擬 200 次各估計量之樣本標準差
bias 代表模擬 200 次各估計量之偏差平均
rmse 代表模擬 200 次各估計量樣本均方根差

I vs. I			$N_{12}=40$		
$CV_1^*=0.000$		$CV_2^*=0.000$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	30.09		2.32	-9.90	10.17
$\widetilde{N}_{12}^{o}$	40.76	6.51	5.55	0.76	5.59
$\widetilde{N}_{12}$	36.91	4.53	4.45	-3.09	5.41
$\widetilde{N}_{12}^{*}$	40.79	6.68	5.56	0.79	5.60

I vs. I			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=0.000$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	59.83		3.49	-20.17	20.47
$\widetilde{N}_{12}^{o}$	80.79	7.63	7.70	0.79	7.72
$\widetilde{N}_{12}$	73.77	6.05	6.79	-6.23	9.20
$\widetilde{N}_{12}^{*}$	80.85	7.74	7.72	0.85	7.75

I vs. II			$N_{12}=40$		
$CV_1^*=0.000$		$CV_2^*=0.223$		$CV_{12}^*=-0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	32.19		2.49	-7.82	8.20
$\widetilde{N}_{12}^{o}$	39.69	4.48	4.73	-0.31	4.73
$\widetilde{N}_{12}$	38.05	3.71	4.37	-1.95	4.78
$\widetilde{N}_{12}^{*}$	40.32	4.70	5.01	0.32	5.01

I vs. II			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=0.451$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	56.67		3.59	-23.32	23.60
$\widetilde{N}_{12}^{o}$	74.81	7.27	8.53	-5.19	9.97
$\widetilde{N}_{12}$	74.04	7.40	8.52	-5.96	10.38
$\widetilde{N}_{12}^{*}$	77.74	8.00	9.54	-2.26	9.79

I vs. III			$N_{12}=40$		
$CV_1^*=0.000$		$CV_2^*=0.017$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	21.22		2.96	-18.78	19.01
$\widetilde{N}_{12}^{o}$	43.55	12.64	16.81	3.55	17.14
$\widetilde{N}_{12}$	29.85	6.05	7.34	-10.15	12.51
$\widetilde{N}_{12}^{*}$	43.62	12.50	16.82	3.62	17.16

I vs. III			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=0.146$		$CV_{12}^*=-0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	49.58		3.76	-30.42	30.65
$\widetilde{N}_{12}^{o}$	76.93	11.93	11.48	-3.07	11.85
$\widetilde{N}_{12}$	66.64	7.80	8.63	-13.36	15.89
$\widetilde{N}_{12}^{*}$	77.09	11.96	11.52	-2.91	11.85

表1.（續）

I vs. IV			$N_{12}=40$		
$CV_1^*=0.000$		$CV_2^*=2.540$		$CV_{12}^*=-0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	28.80		2.64	-11.20	11.50
$\widetilde{N}_{12}^o$	36.79	6.23	5.33	-3.21	6.21
$\widetilde{N}_{12}$	36.47	5.33	5.21	-3.53	6.28
$\widetilde{N}_{12}^*$	38.70	6.94	6.08	-1.30	6.20

I vs. IV			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=4.297$		$CV_{12}^*=-0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	45.81		3.64	-34.19	34.38
$\widetilde{N}_{12}^o$	65.74	10.79	9.30	-14.26	17.01
$\widetilde{N}_{12}$	68.66	11.86	11.00	-11.34	15.78
$\widetilde{N}_{12}^*$	72.64	13.10	11.65	-7.36	13.76

II vs. II			$N_{12}=40$		
$CV_1^*=0.463$		$CV_2^*=0.463$		$CV_{12}^*=0.297$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	34.84		1.90	-5.16	5.50
$\widetilde{N}_{12}^o$	37.24	1.91	2.41	-2.76	3.66
$\widetilde{N}_{12}$	39.78	3.42	3.10	-0.22	3.10
$\widetilde{N}_{12}^*$	38.27	2.46	2.63	-1.73	3.15

II vs. II			$N_{12}=80$		
$CV_1^*=0.971$		$CV_2^*=0.971$		$CV_{12}^*=1.301$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	54.66		3.47	-25.34	25.57
$\widetilde{N}_{12}^o$	61.54	3.63	5.07	-18.46	19.14
$\widetilde{N}_{12}$	77.70	14.62	12.01	-2.30	12.20
$\widetilde{N}_{12}^*$	66.26	4.83	6.32	-13.74	15.12

II vs. III			$N_{12}=40$		
$CV_1^*=0.161$		$CV_2^*=-0.038$		$CV_{12}^*=-0.013$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	22.89		3.17	-17.11	17.39
$\widetilde{N}_{12}^o$	46.15	14.98	19.30	6.15	20.21
$\widetilde{N}_{12}$	31.75	6.06	6.78	-8.25	10.67
$\widetilde{N}_{12}^*$	46.92	15.02	19.73	6.92	20.86

II vs. III			$N_{12}=80$		
$CV_1^*=0.215$		$CV_2^*=-0.058$		$CV_{12}^*=-0.065$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	45.80		4.25	-34.20	34.46
$\widetilde{N}_{12}^o$	85.33	20.01	19.26	5.33	19.94
$\widetilde{N}_{12}$	69.91	11.32	12.17	-10.09	15.78
$\widetilde{N}_{12}^*$	88.70	21.91	20.66	8.70	22.37

II vs. IV			$N_{12}=40$		
$CV_1^*=0.839$		$CV_2^*=3.658$		$CV_{12}^*=4.038$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	31.59		2.17	-8.41	8.69
$\widetilde{N}_{12}^o$	34.91	2.66	2.90	-5.09	5.86
$\widetilde{N}_{12}$	40.01	6.25	5.50	0.01	5.49
$\widetilde{N}_{12}^*$	37.10	3.57	3.63	-2.90	4.64

II vs. IV			$N_{12}=80$		
$CV_1^*=1.603$		$CV_2^*=6.743$		$CV_{12}^*=14.118$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	45.24		3.38	-34.76	34.92
$\widetilde{N}_{12}^o$	52.74	4.58	5.19	-27.26	27.75
$\widetilde{N}_{12}$	71.75	22.56	13.99	-8.25	16.21
$\widetilde{N}_{12}^*$	58.73	5.91	6.80	-21.27	22.33

III vs. III			$N_{12}=40$		
$CV_1^*=0.034$		$CV_2^*=0.034$		$CV_{12}^*=0.002$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	14.98		3.09	-25.02	25.21
$\widetilde{N}_{12}^o$	45.71	6.63	31.66	5.71	32.09
$\widetilde{N}_{12}$	23.47	4.56	9.76	-16.53	19.18
$\widetilde{N}_{12}^*$	45.78	6.55	31.70	5.78	32.15

III vs. III			$N_{12}=80$		
$CV_1^*=0.300$		$CV_2^*=0.300$		$CV_{12}^*=0.125$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	41.58		4.11	-38.42	38.64
$\widetilde{N}_{12}^o$	65.37	13.19	11.89	-14.63	18.84
$\widetilde{N}_{12}$	58.04	10.39	10.59	-21.96	24.37
$\widetilde{N}_{12}^*$	65.49	13.29	12.01	-14.51	18.81

表1. (續)

III vs. IV			$N_{12}=40$			III vs. IV			$N_{12}=80$		
$CV_1^*=-0.101$		$CV_2^*=2.288$		$CV_{12}^*=-0.360$		$CV_1^*=-0.204$		$CV_2^*=3.352$		$CV_{12}^*=-1.300$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	20.43		2.89	-19.57	19.78	M_{12}	35.91		4.13	-44.09	44.28
$\widetilde{N}_{12}^{o}$	43.92	13.32	23.45	3.92	23.72	$\widetilde{N}_{12}^{o}$	80.16	23.29	30.66	0.16	30.59
$\widetilde{N}_{12}$	31.64	7.95	10.13	-8.36	13.12	$\widetilde{N}_{12}$	64.49	14.66	18.61	-15.51	24.19
$\widetilde{N}_{12}^{*}$	46.78	13.52	25.99	6.78	26.79	$\widetilde{N}_{12}^{*}$	87.48	24.84	35.80	7.48	36.49

IV vs. IV			$N_{12}=40$			IV vs. IV			$N_{12}=80$		
$CV_1^*=5.934$		$CV_2^*=5.934$		$CV_{12}^*=42.976$		$CV_1^*=10.862$		$CV_2^*=10.862$		$CV_{12}^*=145.069$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	29.08		2.34	-10.92	11.17	M_{12}	38.34		3.16	-41.66	41.77
$\widetilde{N}_{12}^{o}$	32.05	2.75	3.12	-7.95	8.54	$\widetilde{N}_{12}^{o}$	44.00	4.34	4.64	-36.00	36.29
$\widetilde{N}_{12}$	39.93	10.37	7.36	-0.07	7.34	$\widetilde{N}_{12}$	68.19	38.13	16.14	-11.81	19.96
$\widetilde{N}_{12}^{*}$	34.74	3.75	4.17	-5.26	6.71	$\widetilde{N}_{12}^{*}$	50.35	5.99	6.63	-29.65	30.38

表2. 各估計量之比較, $(n_1, n_2)=(300,300)$

I vs. I			$N_{12}=40$			I vs. I			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=0.000$		$CV_{12}^*=0.000$		$CV_1^*=0.000$		$CV_2^*=0.000$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	34.52		1.95	-5.49	5.82	M_{12}	62.30		3.08	-17.70	17.96
$\widetilde{N}_{12}^{o}$	36.33	1.67	2.35	-3.67	4.35	$\widetilde{N}_{12}^{o}$	68.47	3.26	4.35	-11.53	12.32
$\widetilde{N}_{12}$	39.33	3.70	3.75	-0.67	3.80	$\widetilde{N}_{12}$	78.03	8.23	8.03	-1.97	8.25
$\widetilde{N}_{12}^{*}$	37.53	2.33	2.85	-2.47	3.76	$\widetilde{N}_{12}^{*}$	71.05	4.07	4.94	-8.95	10.22

I vs. II			$N_{12}=40$			I vs. II			$N_{12}=80$		
$CV_1^*=0.000$		$CV_2^*=0.223$		$CV_{12}^*=-0.000$		$CV_1^*=0.000$		$CV_2^*=0.451$		$CV_{12}^*=0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.98		2.13	-6.03	6.39	M_{12}	61.62		3.40	-18.39	18.70
$\widetilde{N}_{12}^{o}$	37.77	2.77	3.10	-2.23	3.81	$\widetilde{N}_{12}^{o}$	72.41	4.92	5.56	-7.59	9.40
$\widetilde{N}_{12}$	38.98	3.35	3.48	-1.02	3.61	$\widetilde{N}_{12}$	77.31	6.99	7.59	-2.69	8.04
$\widetilde{N}_{12}^{*}$	39.33	3.62	3.70	-0.67	3.75	$\widetilde{N}_{12}^{*}$	76.05	6.04	6.68	-3.95	7.75

表2. (續)

I vs. III			$N_{12} = 40$		
$CV_1^* = 0.000$		$CV_2^* = 0.017$		$CV_{12}^* = 0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.69		2.07	-6.31	6.64
$\widetilde{N}_{12}^{o}$	36.43	2.24	2.69	-3.57	4.46
$\widetilde{N}_{12}$	38.99	3.85	3.71	-1.01	3.83
$\widetilde{N}_{12}^{*}$	37.81	2.92	3.16	-2.19	3.84

I vs. III			$N_{12} = 80$		
$CV_1^* = 0.000$		$CV_2^* = 0.146$		$CV_{12}^* = -0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	61.03		3.14	-18.96	19.22
$\widetilde{N}_{12}^{o}$	69.72	4.24	4.92	-10.28	11.39
$\widetilde{N}_{12}$	77.13	7.81	7.85	-2.87	8.34
$\widetilde{N}_{12}^{*}$	72.62	5.03	5.86	-7.38	9.42

I vs. IV			$N_{12} = 40$		
$CV_1^* = 0.000$		$CV_2^* = 2.540$		$CV_{12}^* = -0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.85		2.08	-6.14	6.49
$\widetilde{N}_{12}^{o}$	36.57	2.22	2.83	-3.43	4.44
$\widetilde{N}_{12}$	38.62	3.46	3.53	-1.38	3.78
$\widetilde{N}_{12}^{*}$	37.92	2.96	3.42	-2.08	3.99

I vs. IV			$N_{12} = 80$		
$CV_1^* = 0.000$		$CV_2^* = 4.297$		$CV_{12}^* = -0.000$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	61.34		3.41	-18.66	18.97
$\widetilde{N}_{12}^{o}$	70.25	4.27	5.28	-9.75	11.08
$\widetilde{N}_{12}$	78.01	7.76	8.25	-1.99	8.47
$\widetilde{N}_{12}^{*}$	73.35	5.20	6.18	-6.65	9.07

II vs. II			$N_{12} = 40$		
$CV_1^* = 0.463$		$CV_2^* = 0.463$		$CV_{12}^* = 0.297$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	34.28		1.83	-5.71	6.00
$\widetilde{N}_{12}^{o}$	36.09	1.67	2.27	-3.91	4.52
$\widetilde{N}_{12}$	39.35	3.81	4.10	-0.65	4.15
$\widetilde{N}_{12}^{*}$	37.33	2.34	2.91	-2.67	3.95

II vs. II			$N_{12} = 80$		
$CV_1^* = 0.971$		$CV_2^* = 0.971$		$CV_{12}^* = 1.301$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	63.31		2.87	-16.68	16.93
$\widetilde{N}_{12}^{o}$	69.17	3.16	3.90	-10.83	11.50
$\widetilde{N}_{12}$	79.11	8.32	7.35	-0.89	7.39
$\widetilde{N}_{12}^{*}$	71.88	3.99	4.61	-8.12	9.33

II vs. III			$N_{12} = 40$		
$CV_1^* = 0.161$		$CV_2^* = -0.038$		$CV_{12}^* = -0.013$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.67		2.13	-6.33	6.68
$\widetilde{N}_{12}^{o}$	37.40	2.73	2.95	-2.60	3.93
$\widetilde{N}_{12}$	38.40	3.17	3.42	-1.60	3.77
$\widetilde{N}_{12}^{*}$	38.81	3.55	3.46	-1.19	3.65

II vs. III			$N_{12} = 80$		
$CV_1^* = 0.215$		$CV_2^* = -0.058$		$CV_{12}^* = -0.065$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	62.20		3.04	-17.80	18.06
$\widetilde{N}_{12}^{o}$	73.08	4.87	5.16	-6.92	8.63
$\widetilde{N}_{12}$	77.48	6.68	6.49	-2.52	6.94
$\widetilde{N}_{12}^{*}$	76.64	6.03	6.01	-3.36	6.87

II vs. IV			$N_{12} = 40$		
$CV_1^* = 0.839$		$CV_2^* = 3.658$		$CV_{12}^* = 4.038$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	34.02		2.07	-5.98	6.33
$\widetilde{N}_{12}^{o}$	36.81	2.27	2.70	-3.19	4.18
$\widetilde{N}_{12}$	39.43	3.82	3.80	-0.57	3.83
$\widetilde{N}_{12}^{*}$	38.19	2.97	3.16	-1.81	3.64

II vs. IV			$N_{12} = 80$		
$CV_1^* = 1.603$		$CV_2^* = 6.743$		$CV_{12}^* = 14.118$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	62.13		3.24	-17.87	18.16
$\widetilde{N}_{12}^{o}$	71.12	4.35	5.05	-8.88	10.21
$\widetilde{N}_{12}$	79.39	8.02	8.27	-0.61	8.27
$\widetilde{N}_{12}^{*}$	74.24	5.20	6.09	-5.76	8.38

表2. (續)

III vs. III			$N_{12}=40$			III vs. III			$N_{12}=80$		
$CV_1^*=0.034$		$CV_2^*=0.034$		$CV_{12}^*=0.002$		$CV_1^*=0.300$		$CV_2^*=0.300$		$CV_{12}^*=0.125$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.84		1.92	-6.16	6.45	M_{12}	62.74		3.09	-17.26	17.53
$\widetilde{N}_{12}^o$	35.60	1.61	2.38	-4.40	5.00	$\widetilde{N}_{12}^o$	68.87	3.27	4.02	-11.13	11.83
$\widetilde{N}_{12}$	38.71	3.79	3.64	-1.29	3.85	$\widetilde{N}_{12}$	78.99	8.46	7.81	-1.01	7.86
$\widetilde{N}_{12}^*$	36.76	2.26	2.78	-3.24	4.26	$\widetilde{N}_{12}^*$	71.62	4.14	4.71	-8.38	9.61

III vs. IV			$N_{12}=40$			III vs. IV			$N_{12}=80$		
$CV_1^*=-0.101$		$CV_2^*=2.288$		$CV_{12}^*=-0.360$		$CV_1^*=-0.204$		$CV_2^*=3.352$		$CV_{12}^*=-1.300$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	33.77		2.27	-6.22	6.62	M_{12}	62.04		3.19	-17.96	18.24
$\widetilde{N}_{12}^o$	37.68	2.83	3.57	-2.32	4.25	$\widetilde{N}_{12}^o$	72.85	4.88	5.12	-7.15	8.79
$\widetilde{N}_{12}$	38.85	3.38	3.95	-1.15	4.10	$\widetilde{N}_{12}$	76.72	6.44	6.85	-3.28	7.58
$\widetilde{N}_{12}^*$	39.16	3.62	4.21	-0.84	4.28	$\widetilde{N}_{12}^*$	76.18	6.02	6.29	-3.82	7.35

IV vs. IV			$N_{12}=40$			IV vs. IV			$N_{12}=80$		
$CV_1^*=5.934$		$CV_2^*=5.934$		$CV_{12}^*=42.976$		$CV_1^*=10.862$		$CV_2^*=10.862$		$CV_{12}^*=145.069$	
	est.	$\widehat{s.e.}$	s.e.	bias	rmse		est.	$\widehat{s.e.}$	s.e.	bias	rmse
M_{12}	34.24		1.80	-5.75	6.03	M_{12}	62.83		3.27	-17.17	17.48
$\widetilde{N}_{12}^o$	36.07	1.67	2.15	-3.93	4.48	$\widetilde{N}_{12}^o$	69.05	3.29	4.56	-10.95	11.86
$\widetilde{N}_{12}$	38.99	3.64	3.54	-1.01	3.67	$\widetilde{N}_{12}$	78.66	8.22	7.69	-1.34	7.79
$\widetilde{N}_{12}^*$	37.21	2.30	2.57	-2.79	3.79	$\widetilde{N}_{12}^*$	71.73	4.12	5.16	-8.27	9.74

表3. 隨機模式下估計量之比較

$N_1=N_2=100 \quad \alpha_1=\alpha_2=0.5 \quad \lambda_1=\lambda_2=2$											
$N_{12}=40$						$N_{12}=80$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	12.18	2.79	-27.82	27.96		M_{12}	24.91	3.80	-55.09	55.22	
$\widetilde{N}_{12}^o$	16.93	8.36	-23.07	24.53	0.423	$\widetilde{N}_{12}^o$	32.30	6.84	-47.70	48.19	0.404
$\widetilde{N}_{12}$	17.83	7.96	-22.17	23.55	0.446	$\widetilde{N}_{12}$	35.84	7.71	-44.16	44.83	0.448
$\widetilde{N}_{12}^*$	19.78	13.13	-20.22	24.09	0.494	$\widetilde{N}_{12}^*$	37.15	9.61	-42.85	43.91	0.464
$\widetilde{N}_{12}^t$	46.31	31.94	6.31	32.48	1.158	$\widetilde{N}_{12}^t$	87.61	29.99	7.61	30.87	1.095

表3.（續）

$N_1 = N_2 = 200$ $\alpha_1 = \alpha_2 = 0.5$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 80$						$N_{12} = 160$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	24.50	4.39	-55.51	55.68		M_{12}	49.20	5.62	-110.80	110.94	
$\widetilde{N}_{12}^{o}$	31.20	6.57	-48.80	49.24	0.390	$\widetilde{N}_{12}^{o}$	61.90	8.95	-98.10	98.50	0.387
$\widetilde{N}_{12}$	35.26	8.41	-44.74	45.52	0.441	$\widetilde{N}_{12}$	69.99	11.02	-90.01	90.68	0.437
$\widetilde{N}_{12}^{*}$	36.01	8.78	-44.00	44.86	0.450	$\widetilde{N}_{12}^{*}$	71.29	12.02	-88.71	89.52	0.446
$\widetilde{N}_{12}^{t}$	84.63	27.86	4.63	28.17	1.058	$\widetilde{N}_{12}^{t}$	166.76	35.21	6.76	35.76	1.042

$N_1 = N_2 = 500$ $\alpha_1 = \alpha_2 = 0.5$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 200$						$N_{12} = 400$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	60.85	5.93	-139.15	139.28		M_{12}	122.83	8.67	-277.17	277.31	
$\widetilde{N}_{12}^{o}$	76.55	9.63	-123.45	123.82	0.383	$\widetilde{N}_{12}^{o}$	154.45	14.58	-245.55	245.98	0.386
$\widetilde{N}_{12}$	87.33	12.92	-112.67	113.41	0.437	$\widetilde{N}_{12}$	175.85	18.33	-224.15	224.89	0.440
$\widetilde{N}_{12}^{*}$	88.06	13.03	-111.94	112.69	0.440	$\widetilde{N}_{12}^{*}$	178.16	18.70	-221.84	222.62	0.445
$\widetilde{N}_{12}^{t}$	203.76	40.82	3.76	40.89	1.019	$\widetilde{N}_{12}^{t}$	414.15	56.92	14.15	58.51	1.035

$N_1 = N_2 = 1000$ $\alpha_1 = \alpha_2 = 0.5$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 400$						$N_{12} = 800$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	122.15	9.23	-277.86	278.01		M_{12}	244.01	11.82	-556.00	556.12	
$\widetilde{N}_{12}^{o}$	153.34	14.41	-246.66	247.08	0.383	$\widetilde{N}_{12}^{o}$	305.48	19.15	-494.52	494.89	0.382
$\widetilde{N}_{12}$	175.09	18.28	-224.91	225.65	0.438	$\widetilde{N}_{12}$	351.55	25.76	-448.45	449.18	0.439
$\widetilde{N}_{12}^{*}$	177.16	18.60	-222.84	223.61	0.443	$\widetilde{N}_{12}^{*}$	354.10	25.77	-445.90	446.64	0.443
$\widetilde{N}_{12}^{t}$	410.25	53.83	10.25	54.67	1.026	$\widetilde{N}_{12}^{t}$	821.99	80.45	21.99	83.20	1.027

$N_1 = N_2 = 100$ $\alpha_1 = \alpha_2 = 1$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 40$						$N_{12} = 80$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	17.77	2.76	-22.24	22.41		M_{12}	35.93	3.86	-44.08	44.24	
$\widetilde{N}_{12}^{o}$	23.18	4.57	-16.83	17.43	0.579	$\widetilde{N}_{12}^{o}$	47.71	7.50	-32.29	33.14	0.596
$\widetilde{N}_{12}$	25.55	6.50	-14.45	15.87	0.639	$\widetilde{N}_{12}$	51.90	8.48	-28.10	29.35	0.649
$\widetilde{N}_{12}^{*}$	25.77	5.69	-14.23	15.32	0.644	$\widetilde{N}_{12}^{*}$	53.08	9.37	-26.93	28.50	0.663
$\widetilde{N}_{12}^{t}$	41.29	10.54	1.29	10.59	1.032	$\widetilde{N}_{12}^{t}$	85.39	17.49	5.39	18.26	1.067

表3. (續)

$N_1 = N_2 = 200$ $\alpha_1 = \alpha_2 = 1$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 80$						$N_{12} = 160$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	35.46	4.87	-44.55	44.81		M_{12}	70.58	5.54	-89.43	89.60	
$\widetilde{N}^o_{12}$	46.37	8.43	-33.63	34.66	0.580	$\widetilde{N}^o_{12}$	90.71	10.07	-69.30	70.02	0.567
$\widetilde{N}_{12}$	49.76	9.20	-30.24	31.60	0.622	$\widetilde{N}_{12}$	101.01	13.37	-59.00	60.48	0.631
$\widetilde{N}^*_{12}$	51.13	10.18	-28.87	30.61	0.639	$\widetilde{N}^*_{12}$	101.23	12.83	-58.77	60.14	0.633
$\widetilde{N}^t_{12}$	81.46	19.06	1.46	19.07	1.018	$\widetilde{N}^t_{12}$	160.17	23.10	0.17	23.04	1.001

$N_1 = N_2 = 500$ $\alpha_1 = \alpha_2 = 1$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 200$						$N_{12} = 400$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	90.10	6.91	-109.90	110.12		M_{12}	178.04	8.99	-221.96	222.14	
$\widetilde{N}^o_{12}$	116.97	11.89	-83.03	83.87	0.585	$\widetilde{N}^o_{12}$	229.14	16.59	-170.86	171.66	0.573
$\widetilde{N}_{12}$	129.12	14.90	-70.88	72.42	0.646	$\widetilde{N}_{12}$	255.29	21.30	-144.71	146.26	0.638
$\widetilde{N}^*_{12}$	131.27	15.38	-68.73	70.42	0.656	$\widetilde{N}^*_{12}$	255.41	21.21	-144.59	146.13	0.639
$\widetilde{N}^t_{12}$	208.62	27.03	8.62	28.31	1.043	$\widetilde{N}^t_{12}$	405.07	37.83	5.07	38.07	1.013

$N_1 = N_2 = 1000$ $\alpha_1 = \alpha_2 = 1$ $\lambda_1 = \lambda_2 = 2$											
$N_{12} = 400$						$N_{12} = 800$					
	est.	s.e.	bias	rmse	ratio		est.	s.e.	bias	rmse	ratio
M_{12}	176.50	9.13	-223.51	223.69		M_{12}	356.29	13.30	-443.72	443.91	
$\widetilde{N}^o_{12}$	226.72	15.30	-173.28	173.95	0.567	$\widetilde{N}^o_{12}$	458.12	23.07	-341.88	342.65	0.573
$\widetilde{N}_{12}$	253.26	21.13	-146.74	148.25	0.633	$\widetilde{N}_{12}$	510.08	29.29	-289.92	291.39	0.638
$\widetilde{N}^*_{12}$	253.47	20.38	-146.53	147.94	0.634	$\widetilde{N}^*_{12}$	512.08	28.56	-287.92	289.32	0.640
$\widetilde{N}^t_{12}$	401.68	35.38	1.68	35.34	1.004	$\widetilde{N}^t_{12}$	813.46	53.30	13.46	54.85	1.017

參考文獻

Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: a review. *J. Amer. Statist. Assoc.* **88**, 364-373.

Chao, A. and Lee,S. -M. (1992). Estimating the number of classes via sample coverage. *J. Amer. Statist. Assoc.* **87**, 210-217.

Chao, A., Ma, M. -C. and Yang, C. K. M. (1993). Stopping rules and estimation for recapture debugging with unequal failure eates. *Biometrika* **80**, 193-201.

Chen, W. -C. (1980). On the weak form of zipf's law. *J. Appl. Probab.* **18**, 611-622.

Coddington, J. A., Griswold, C. E., Silva Dávila, D., Peñaranda, E. and Larcher, S. F. (1991). Designing and Testing Sampling Protocols to Estimate Biodiversity in Tropical Ecosystems. In *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematic and Evolutionary Biology* (ed. E. C. Dudley), 44-60. Portland, Oregon: Dioscorides Press.

Colwell, R. K. (1973). Competition and coexistence in a simple tropical community. *Am. Nat.* **107**, 737-760.

Colwell, R. K. and Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond.* B **345**, 101-118.

Feinsinger, P. (1976). Organization of a tropical guild of nectarivorous birds. *Ecol. Monogr.* **46**, 257-291.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237-264.

Grassle, J. F. and Smith, W. (1976). A similarity measure sensitive to the contribution of rare species and its use in investigation of variation in marine benthic communities. *Oecologia* **25**, 13-22.

Karr, J. R., Robinson, S. K., Blake, J. G. and Bierregaard, R. O. (1990). Birds of Four Neotropical Forests. In *Four Neotropical Rainforests* (ed. A. H. Gentry), 237-269. Yale University Press.

Pielou, E. C. (1975). *Ecological Diversity*. Wiley Interscience, New York.

Pielou, E. C. (1977). *Mathematical Ecology.* Wiley, New York.

[民國 84 年 3 月 24 日收稿，民國 84 年 5 月 24 日修正]

Estimating the Number of Common Species–Analysis of the Number of Common Bird Species in Ke-yar Stream and Chung-kung Stream

Y.-C.Chen W.-H.Hwang Anne Chao

Institute of Statistics
National Tsing Hua University
Hsinchu, Taiwan, 30043, R.O.C.

C.-Y. Kuo

The Wild Bird Society of Hsin Chu

ABSTRACT

Assume that a multinomial sample is chosen from two populations respectively. Each observation is classified to its species identity. This paper uses the concept of sample coverage to estimate the number of common species of the two populations. Computer simulation investigates the performance of the proposed estimators. The result generalizes Chao and Lee (1992) to two populations problem. The bird data collected from April 1994 to March 1995 in Ke-yar and Chung-kung streams are used to illustrate the estimation procedure.

Key words and phrases: Multinomial, heterogeneity, sample coverage, coefficient of variation.

AMS 1991 subject classifications: Primary 62P10; secondary 62F10.