

摘要

Abstract

目錄

摘要.....	i
Abstract.....	ii
目錄.....	iii
圖目錄.....	iv
表目錄.....	vi
第 1 章 緒論.....	1
第 2 章 模型符號介紹與相關文獻回顧.....	4
2.1 符號定義.....	4
2.2 相關文獻回顧.....	6
第 3 章 使用動差法估計共同物種數.....	18
3.1 取後放回之抽樣方法的估計方式.....	18
3.2 取後不放回之抽樣方法的估計方式.....	22
第 4 章 模擬研究與討論.....	30
4.1 估計式整理.....	30
4.2 模型模擬設定.....	34
4.3 表格中名詞定義.....	36
4.4 取後放回的模擬結果.....	37
4.5 取後不放回與取後放回的模擬結果比較.....	42
4.6 真實資料模擬.....	49
4.7 討論.....	55
第 5 章 實例分析.....	57
5.1 紅杉國家公園內苔蘚資料.....	57
第 6 章 結論與後續研究.....	67
參考文獻.....	69
附錄.....	72

圖目錄

Figure 4.1 群落一為同質模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	38
Figure 4.2 兩群落皆為均勻模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	39
Figure 4.3 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	40
Figure 4.4 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	41
Figure 4.5 群落一為同質模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	47
Figure 4.6 兩群落皆為均勻模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	47
Figure 4.7 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	48
Figure 4.8 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖)估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。	48
Figure 4.9 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Big Wet 估計結果。	50
Figure 4.10 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Post-Big Wet 估計結果。	51
Figure 4.11 取後放回澳洲三種極端氣候鳥類資料集中的 Big Wet 與 Post-Big Wet 估計結果。	52
Figure 4.12 取後不放回 BCI 資料集中的兩群落估計結果。	54

Figure 5.1 利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。.....	61
Figure 5.2 假設抽樣比例為 0.3 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。.....	65
Figure 5.3 假設抽樣比例為 0.5 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。.....	65
Figure 5.4 假設抽樣比例為 0.7 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。.....	66

表目錄

Table 4.1 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	38
Table 4.2 兩群落皆為均勻模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	39
Table 4.3 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	40
Table 4.4 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	41
Table 4.5 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	43
Table 4.6 兩群落皆為均勻模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	44
Table 4.7 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	45
Table 4.8 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S1 = S2 = 400$ ， $S12 = 300$ 。	46
Table 4.9 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Big Wet 估計結果。	50
Table 4.10 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Post-Big Wet 估計結果。	51
Table 4.11 取後放回澳洲三種極端氣候鳥類資料集中的 Big Wet 與 Post-Big Wet 估計結果。	52
Table 4.12 取後不放回 BCI 資料集中的兩群落估計結果。	54

Table 5.1 取後放回之紅杉國家公園內苔蘚的兩群落間共同種估計結果。.....	59
Table 5.2 取後放回之紅杉國家公園內苔蘚的兩群落間物種數與 Jaccard 估計結果。.....	60
Table 5.3 不同抽樣比例之單群落物種數估計 (mean \pm SD) 。.....	62
Table 5.4 假設抽樣比例為 0.3 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。.....	63
Table 5.5 假設抽樣比例為 0.5 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。.....	63
Table 5.6 假設抽樣比例為 0.7 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。.....	64
Table 5.7 在不同抽樣比例的假設下，取後不放回之紅杉國家公園內苔蘚的兩群落間物種數與 Jaccard 估計結果。.....	64

第1章緒論

物種豐富度，即在某一特定群落中所存在的物種數，在生態多樣性的研究中，物種豐富度是一種最為簡單且直觀的指標。特別是在維持多種生態功能，以及多樣性的保護與管理上。此外，也有研究表明，物種豐富度與生態系統的功能性呈正相關 (Shmida et.al, 1985, Maestre et.al, 2012)。物種豐富度的存在在生態研究中顯得格外重要。然而在生態調查中，往往因為人力、資金以及時間等成本因素，而無法準確調查到群落中存在的所有物種。因此，如何準確地估計物種豐富度便成為一大難題。

針對物種豐富度所進行的估計方式可依據所蒐集的資料型態大致分為兩種：豐富度數據與出現率數據。所謂豐富度數據通常指的是描述一個生態系統中物種數量和種類的數據。這些數據通常是基於樣本收集的，例如在一個特定的地點收集樣本，後對樣本進行物種鑑定和計數。豐富度數據通常包括物種數量、物種種類、物種相對豐度等資訊；而出現率數據通常指的是描述一個生態系統中物種出現與否的數據。這些數據通常是基於區塊、陷阱或是時間作為單位所收集，例如在特定目標地區中，抽取其中部分的區塊做為樣本使用；或是在一個特定的地點定期收集樣本。出現率數據往往僅記錄物種在樣本中出現與否，而非實際出現的次數或個體數。兩者皆可用來作為計算生物多樣性指標 (Chao & Chiu, 2016)。

除了抽樣數據的不同之外，抽樣方式則是可以分為取後放回 (sampling with replacement) 以及取後不放回 (sampling without replacement) 兩種常見的抽樣方式。並在物種豐富度的估計上，針對不同的數據型態、不同的抽樣方式，進行相對應的模型假設。

在大多數先前研究所提出估計式中，依據估計方法的不同可分為參數化方法以及非參數化方法。在非參數化方法所建立估計式的過程中，通常不必假設

一個特定的機率分佈。模型的形式通常是通過估計一個未知的密度函數或分佈函數來進行的。例如在物種豐富度估計中常見的下界估計方法 (Chao, 1984, 1987) 以及 jackknife 估計方法 (Burnham & Overton, 1978, 1979)。另一方面，參數化方法則需要假設一個特定的機率分佈，並且假設這個分佈的參數化是已知的或可以通過樣本估計的。在參數化方法中，模型的參數化通常可以通過最大概似估計 (MLE) 等方式進行估計。在物種豐富度的估計方面，針對物種出現機率的隨機變數進行分佈的假設，藉此減少參數數量並使用傳統統計方式估計 (Dorazio & Royle, 2003)。大部分情況下，非參數化方法通常比參數化方法更靈活，因為其無須假設一個特定的分佈形式，然而非參數化方法卻需要更多的數據和計算資源來估計未知的密度函數或分佈函數。反觀，參數化方法通常比非參數化方法更簡單，因為它們可以藉由假設一個特定的分佈簡化問題，且在樣本數夠大的情況下，利用參數化方法所建立的估計方式，通常具有更好的效果 (Wu et.al, 2006)。

而在生態上的應用以及環境決策中，除了單一群落的物種豐富度之外，往往也需要針對多個群落之間進行比較。這些群落可能是一保護區隨時間的變化，或是不同海拔或緯度的生態差異等 (Chao, et.al, 2000)。在這其中，共同物種豐富度，即共同物種數，便是建構群落之間相似性以及 *Beta* 多樣性重要的基礎之一。因此，估計多個群落間存在的共同物種豐富度也是一樣重要的議題。在過去，已有許多研究針對不同的抽樣方式、參數化或非參數化的方法來建立兩群落的共同物種數進行估計 (Chao, et.al, 2000, Pan et.al, 2009, Chao & Lin, 2012)。此外，也有許多研究提出使用參數化方式，假設物種出現的機率為 *Beta* 分佈之隨機變數 (Shen and He, 2008)。因此，本文將針對不同取樣方式下，出現率數據所估計之共同物種數，結合 *Beta* - 二項式模型 (beta-binomial model) 修正原有的估計方法。

以下為本篇文章的章節安排：在第二章中，首先將針對文章中的數學符

號、出現率數據、不同抽樣方式下的單群落與兩群落的物種數估計，以及一些估計指標進行回顧。接著，在第三章中將針對本篇文章所提的修正估計方法進行完整的推導與描述。隨後在第四章中，利用所模擬出的群落數據，以及兩筆真實資料：澳洲三種極端氣候鳥類資料與 BCI 資料，作為母體 (Wilson et.al, 2022, Condit et.al, 2019)，以電腦模擬的方式呈現估計結果，並評估估計式的穩定性。並在第五章中使用紅杉國家公園內苔蘚資料進行實例分析 (Coleman et.al, 2023)，將估計式實際應用於資料分析中。最後針對本篇文章給予一個總結，同時針對研究的未來發展提出討論。

第2章 模型符號介紹與相關文獻回顧

2.1 符號定義

在常見的生態資料的蒐集上，常見的抽樣方式一般分為兩種：其一為以個體為抽樣單位，即豐富度數據 (abundance data)；其次為依照區塊為抽樣單位，即出現率數據 (incidence data)。本文主要使用以區塊抽樣為主的發生率資料。在估計式中，常見符號由以下清單表示：

- S ：混合群落中的總相異物種數。
- S_i ：第 i 群落的物種數， $i = 1, 2$ 。
- S_{12} ：兩群落共同物種數。
- π_i ：群落中物種 i 出現的機率，即相對豐富度。 $i = 1, 2, \dots, S$ 。
- π_{ji} ：第 j 群落中物種 i 出現的機率，即相對豐富度。 $i = 1, 2, \dots, S, j = 1, 2$ 。
- X_i ：第一群落樣本，第 i 物種出現的區塊數量。
- Y_i ：第二群落樣本，第 i 物種出現的區塊數量。
- D_{12} ：兩樣本中出現的共同物種數。 $D_{12} = \sum_{i=1}^S I(X_i > 0 \text{ and } Y_i > 0)$ 。
- T_i ：第 i 群集的總區塊數量， $i = 1, 2$ 。
- t_i ：第 i 群集的抽樣區塊數量， $i = 1, 2$ 。
- q ：抽樣比例， $q = \frac{t_1}{T_1} = \frac{t_2}{T_2}$ 。
- p_k ：群落樣本中物種的共出現在 k 個區塊的機率。 $k = 1, 2, \dots, T$ 。
- p_{kl} ：第一群落樣本中物種的共出現在 k 個區塊，且同時第二群落樣本中物種的共出現在 l 個區塊的機率。 $k = 1, 2, \dots, T_1, l = 1, 2, \dots, T_2$ 。

- Q_k ：樣本中剛好出現 k 個區塊的物種數。 $k = 1, 2, \dots, n_i$ ， $i = 1, 2$ 。

$$Q_k = \sum_{i=1}^{S_1} I(X_i = k) \text{ or } \sum_{i=1}^{S_2} I(Y_i = k)$$

- Q_{kl} ：樣本中在第一族群出現 k 個區塊，並在第二族群出現 l 個區塊的物種數。 $k = 1, 2, \dots, S_1$ ， $l = 1, 2, \dots, S_2$ 。

$$Q_{k_1, k_2} = \sum_{i=1}^S I(X_i = k_1, Y_i = k_2)$$

- Q_{k+} ：樣本中在第一族群出現 k 次，並在第二族群出現的物種數。 $k = 1, 2, \dots, S_1$ 。

$$Q_{k+} = \sum_{i=1}^{S_1} I(X_i = k, Y_i > 0)$$

- Q_{+l} ：樣本中在第一族群出現，並在第二族群出現 l 次的物種數。 $l = 1, 2, \dots, S_2$ 。

$$Q_{+l} = \sum_{i=1}^{S_2} I(X_i > 0, Y_i = l)$$

2.2 相關文獻回顧

2.2.1 出現率數據

在大多數的生態調查研究中，抽樣為隨機且獨立的。且抽樣單位通常為陷阱、區塊與定時調查。在出現率數據的抽樣中，大多數的方法是將其中的研究區域劃分為多個面積大致相同的 T 個區塊，並從中隨機選擇特定的 t 個區塊做為抽樣樣本進行調查。對於不同類型的物種，準確計算每個抽樣區塊中出現的個體數往往是一件相對困難的。因此在多數情況下，調查時僅記錄該物種在 t 個抽樣區塊中的出現率，即紀錄物種在該區塊出現與否。

出現率數據由一組為 t 個抽樣區塊的樣本所組成，並記錄每個區塊中每種物種的出現或是未出現，以形成一個具有 S 行與 t 列的矩陣 W_{ij} 。其中，若是在第 j 個區塊中發現 i 物種，則計為 1；反之若尚未觀測到該物種則計為 0。

又 X_i 被定義為該樣本中的出現頻率向量， $X_i = \sum_{j=1}^t W_{ij}$ ，表示在該樣本中第 i 物種出現的總區塊數量。若 $X_i = 0$ 則表示該抽樣樣本中並無觀測到該物種，且在樣本中觀測到的物種總數為 S_{obs} ，故 $0 \leq X_i < S_{obs}$ 。

並且，可令 Q_k 表示在出現頻率向量 X_i 中出現 k 次的物種數， $Q_k = \sum_{i=1}^S I(X_i = k)$ ，且 $k=1, 2, \dots, t$ 。故 Q_1 為在該樣本中僅出現在一個區塊的物種數， Q_2 為在該樣本中出現在兩個區塊的物種數，並依此類推。除此之外， Q_0 為在該樣本並未被觀測到的物種數。而真實的物種數 S ，應為被觀測到的物種數 S_{obs} 與未被觀測到的物種數 Q_0 之總和。

2.2.2 取後放回之抽樣方式

在生態調查的研究中，物種數或稱物種豐富度是最直接呈現多樣性的指標之一，往往需要消耗大量的人力、經費與時間等成本。這使得在抽樣的結果中，能看見所有物種皆出現之狀況的機率大幅降低。也就是說，在大部分的生態調查結果中，皆可能存在部分未被觀測到的物種。因此，需針對該部分未被

觀測到的物種進行估計，以獲取到更接近於真實物種豐富度的結果。

對於出現率數據所開發之物種豐富度估計的模型多數皆式依據捉放法 (capture-recapture) 的抽樣方式為基礎所建立。一般而言，傳統的捉放法是藉由單一物種「個體數」，針對該物種在群落中所佔比例進行估計。而在物種豐富度的估計中，可將捉放法「個體數」對應至「物種數」，已估計群落中的物種數作為物種豐富度的指標所使用。

在物種豐富度的調查結果中，又可將物種大致分為豐富物種與稀有物種。在大多數情況下，稀有物種對於未被觀測到的物種提供了更為豐富的資訊。這是由於，相較於皆為豐富物種的樣本，在某樣本中含有大量的稀有物種時，通常情況下，在該樣本所抽樣地區應存在更多尚未被觀測到的物種。因此在過去許多研究中，皆是藉由稀有物種對物種豐富度的估計進行修正。

依據 2.2.1 所述，出現率數據的樣本來自於依照物種抽樣區塊的出現率矩陣 W_{ij} 所組成。並可將該矩陣整理成出現頻率向量 X_i 。 W_{ij} 應為服從伯努力分佈 (Bernoulli distribution) 的隨機變數，且當 $W_{ij} = 1$ 時機率為 π_i ，而 $W_{ij} = 0$ 時機率為 $1 - \pi_i$ 。則發生率矩陣的機率分佈為：

$$P(W_{ij} = w_{ij} | \pi_i) = \prod_{i=1}^S \prod_{j=1}^t \pi_i^{w_{ij}} (1 - \pi_i)^{1-w_{ij}}$$

又，物種數出現頻率向量 X_i 服從二項分佈 (binomial distribution)：

$$P(X_i = x_i | \pi_i) = \sum_{i=1}^S \binom{t}{x_i} \pi_i^{x_i} (1 - \pi_i)^{t-x_i}$$

單群落物種數估計

Chao (1987) 針對出現率數據建立物種豐富度的無母數估計模型 *Chao2*。所謂無母數估計意旨在該估計方法中，不對物種豐富度或者物種出現機率的分布進行假設。無母數的物種豐富度估計是一個基本且直接的觀念，利用稀有物種中所含的資訊以估計真實的物種豐富度。並根據樣本中物種計數的邊際機率分佈， $E(Q_k)$ 可以表示為：

$$E(Q_k) = \sum_{i=1}^S P(X_i = k) = \sum_{i=1}^S \binom{t}{k} \pi_i^k (1 - \pi_i)^{t-k}$$

因此求得在樣本中未出現以及分別出現一次與兩次的物種數之期望值為：

$$E(Q_0) = \sum_{i=1}^S (1 - \pi_i)^t$$

$$E(Q_1) = \sum_{i=1}^S \binom{t}{1} \pi_i (1 - \pi_i)^{t-1}$$

$$E(Q_2) = \sum_{i=1}^S \binom{t}{2} \pi_i^2 (1 - \pi_i)^{t-2}$$

又根據柯西-施瓦茨不等式 (Cauchy-Schwarz inequality) 之概念可以推導出：

$$\left[\sum_{i=1}^S (1 - \pi_i)^t \right] \left[\sum_{i=1}^S \pi_i^2 (1 - \pi_i)^{t-2} \right] \geq \left[\sum_{i=1}^S \pi_i (1 - \pi_i)^{t-1} \right]^2$$

固可求得：

$$E(Q_0) \geq \frac{t-1}{t} \frac{[E(Q_1)]^2}{2E(Q_2)}$$

最終將結果帶入 $S = S_{obs} + Q_0$ ，可知估計式 *Chao2* 為：

$$\hat{S}_{Chao2} = \begin{cases} S_{obs} + \frac{t-1}{t} \left(\frac{Q_1^2}{2Q_2} \right), & \text{if } Q_2 > 0 \\ S_{obs} + \frac{t-1}{t} \left(\frac{Q_1(Q_1-1)}{2} \right), & \text{if } Q_2 = 0 \end{cases}$$

針對 *Chao2* 的方法，Chiu 等人 (2014) 應用 Good-Turing 頻率公式，並加入 Q_3 與 Q_4 的資訊對其進行修正，開發更為準確的下界估計式 *iChao2*：

$$\hat{S}_{iChao2} = \hat{S}_{Chao2} + \frac{t-3}{4t} \frac{Q_3}{Q_4} \times \max\left(Q_1 - \frac{t-3}{2(t-1)} \frac{Q_2 Q_3}{Q_4}, 0\right)$$

兩群落的共同物種數估計

在生態統計中，群落之間의 共同物種可以表示兩群落間의 物種多樣性，同時也能表現兩群落間의 相似性 (Chao, et al., 2000)。在兩群落的抽樣樣本中，除了共同種之外，也會分別存在只出現於其中單一群落的特有物種。為此，當在比較兩群落之間의 物種豐富度時，並非僅考慮單一群落的物種豐富度，而是必須針對群落間의 共同物種數進行估計。與單群落的物種數估計相似，在大多數情況下，抽樣樣本無法觀測到所有存在的共同種。因此需針對未被抽樣觀測到的共同種進行估計，並加上已存在於樣本中的共同物種數，作為修正的共同物種數所使用。Pan et al. (2009) 將 *Chao2* 的方法推廣至兩群落，建立一估計兩群落間存在的共同物種數之估計式。

假設在第一群落的樣本 (X_i) 與第二群落的樣本 (Y_i) 中，分別有 t_1 、 t_2 個抽樣區塊。且兩群落的第 i 物種出現機率分別表示為 π_{1i} 、 π_{2i} 。則 Q_{kl} 表示在出現頻率向量 X_i 中出現 k 個區塊，同時在 Y_i 中出現 l 個區塊的物種數。 $E(Q_{kl})$ 可以表示為：

$$E(Q_{kl}) = \sum_{i=1}^{S_{12}} P(X_i = k)P(Y_i = l) = \sum_{i=1}^{S_{12}} \binom{t_1}{k} \pi_{1i}^k (1 - \pi_{1i})^{t_1-k} \binom{t_2}{l} \pi_{2i}^l (1 - \pi_{2i})^{t_2-l}$$

因此求得在樣本中分別未出現於兩群落的期望值為：

$$\begin{aligned} E(Q_{00}) &= \sum_{i=1}^{S_{12}} (1 - \pi_{1i})^{t_1} (1 - \pi_{2i})^{t_2} \\ E(Q_{0+}) &= \sum_{i=1}^{S_{12}} (1 - \pi_{1i})^{t_1} [1 - (1 - \pi_{2i})^{t_2}] \\ E(Q_{+0}) &= \sum_{i=1}^{S_{12}} [1 - (1 - \pi_{1i})^{t_1}] (1 - \pi_{2i})^{t_2} \end{aligned}$$

為求得 $E(Q_{0+})$ 下界估計式，因此需要 $E(Q_{1+})$ 與 $E(Q_{2+})$ 的資訊：

$$\begin{aligned} E(Q_{1+}) &= \sum_{i=1}^{S_{12}} \binom{t_1}{1} \pi_{1i} (1 - \pi_{1i})^{t_1-1} [1 - (1 - \pi_{2i})^{t_2}] \\ E(Q_{2+}) &= \sum_{i=1}^{S_{12}} \binom{t_1}{2} \pi_{1i}^2 (1 - \pi_{1i})^{t_1-2} [1 - (1 - \pi_{2i})^{t_2}] \end{aligned}$$

又根據柯西-施瓦茨不等式之概念可以推導出：

$$\begin{aligned} & \left[\sum_{i=1}^{S_{12}} (1 - \pi_{1i})^{t_1} [1 - (1 - \pi_{2i})^{t_2}] \right] \left[\sum_{i=1}^{S_{12}} \pi_{1i}^2 (1 - \pi_{1i})^{t_1-2} [1 - (1 - \pi_{2i})^{t_2}] \right] \\ & \geq \left[\sum_{i=1}^{S_{12}} \pi_{1i} (1 - \pi_{1i})^{t_1-1} [1 - (1 - \pi_{2i})^{t_2}] \right]^2 \end{aligned}$$

固可求得：

$$E(Q_{0+}) \geq \frac{t_1 - 1}{t_1} \frac{[E(Q_{1+})]^2}{2E(Q_{2+})}$$

同理可以求得：

$$E(Q_{+0}) \geq \frac{t_2 - 1}{t_2} \frac{[E(Q_{+1})]^2}{2E(Q_{+2})}$$

而若是想要求得 $E(Q_{00})$ 下界估計式，則需要 $E(Q_{11})$ 與 $E(Q_{22})$ 的資訊：

$$\begin{aligned} E(Q_{11}) &= \sum_{i=1}^{S_{12}} \binom{t_1}{1} \pi_{1i} (1 - \pi_{1i})^{t_1-1} \binom{t_2}{1} \pi_{2i} (1 - \pi_{2i})^{t_2-1} \\ E(Q_{22}) &= \sum_{i=1}^{S_{12}} \binom{t_1}{2} \pi_{1i}^2 (1 - \pi_{1i})^{t_1-2} \binom{t_2}{2} \pi_{2i}^2 (1 - \pi_{2i})^{t_2-2} \\ &= \sum_{i=1}^{S_{12}} \left[\frac{t_1(t_1-1)}{2} \right] \pi_{1i}^2 (1 - \pi_{1i})^{t_1-2} \left[\frac{t_2(t_2-1)}{2} \right] \pi_{2i}^2 (1 - \pi_{2i})^{t_2-2} \end{aligned}$$

又根據柯西-施瓦茨不等式之概念可以推導出：

$$\begin{aligned} & \left[\sum_{i=1}^{S_{12}} (1 - \pi_{1i})^{t_1} (1 - \pi_{2i})^{t_2} \right] \left[\sum_{i=1}^{S_{12}} \pi_{1i}^2 (1 - \pi_{1i})^{t_1-2} \pi_{2i}^2 (1 - \pi_{2i})^{t_2-2} \right] \\ & \geq \left[\sum_{i=1}^{S_{12}} \pi_{1i} (1 - \pi_{1i})^{t_1-1} \pi_{2i} (1 - \pi_{2i})^{t_2-1} \right]^2 \end{aligned}$$

最終得：

$$E(Q_{00}) \geq \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{[E(Q_{11})]^2}{4E(Q_{22})}$$

將結果帶入 $S_{12} = D_{12} + Q_{0+} + Q_{+0} + Q_{00}$ ，可知估計式 Pan 為：

$$\begin{aligned} \hat{S}_{Pan} &= D_{12} + E(Q_{0+}) + E(Q_{+0}) + E(Q_{00}) \\ E(Q_{0+}) &= \begin{cases} \frac{t_1 - 1}{t_1} \frac{Q_{1+}^2}{Q_{2+}}, & \text{if } Q_{2+} > 0 \\ \frac{t_1 - 1}{t_1} \frac{Q_{1+}(Q_{1+} - 1)}{2}, & \text{if } Q_{2+} = 0 \end{cases} \\ E(Q_{+0}) &= \begin{cases} \frac{t_2 - 1}{t_2} \frac{Q_{+1}^2}{Q_{+2}}, & \text{if } Q_{+2} > 0 \\ \frac{t_2 - 1}{t_2} \frac{Q_{+1}(Q_{+1} - 1)}{2}, & \text{if } Q_{+2} = 0 \end{cases} \\ E(Q_{00}) &= \begin{cases} \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{Q_{11}^2}{Q_{22}}, & \text{if } Q_{22} > 0 \\ \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{Q_{11}(Q_{11} - 1)}{4}, & \text{if } Q_{22} = 0 \end{cases} \end{aligned}$$

2.2.3 取後不放回之抽樣方式

相對於取後放回的抽樣方式，另一種在生態資料中常見的抽樣方法為取後

不放回。在取後不放回的抽樣方法中，廣泛使用在林業調查中，依照所選區塊對樹木進行不重複取樣，或是用於陷阱或誘捕器的抽樣方式中，需要殺死個體的抽樣方法。

在這種類型抽樣方法的單群落情況中，假設將欲調查地區大致分為 T 個相等的區塊。又每個區塊物種存在的機率為 π_i ， $0 \leq \pi_i \leq 1$ ， π_i 為未知參數。假設在 T 區塊中，物種 i 僅能在 U_i 的目標區塊中被檢驗到， U_i 亦為未知參數，且 $1 \leq U_i \leq T$ 。則在 $U_i = u_i$ 給定 π_i 的條件下遵循參數 T 和 π_i 的零截尾二項分佈 (zero-truncated binomial distribution)：

$$P(U_i = u_i | \pi_i) = \binom{T}{u_i} \frac{\pi_i^{u_i} (1 - \pi_i)^{T-u_i}}{1 - (1 - \pi_i)^T}, i = 1, 2, \dots, S$$

當從 T 個區塊中，針對群落進行取後不放回之隨機抽樣，分別抽取 t 的區塊數，若在取樣區塊中發現該物種，則被紀錄為存在，反之則為不存在，僅記錄每個採樣樣本中物種的發生率，以形成出現頻率向量 X_i 。在給定 $U_i = u_i$ 的情況下， X_i 應服從超幾何分佈 (hypergeometric distribution)：

$$P(X_i = x_i | u_i) = \prod_{i=1}^S \left\{ \frac{\binom{u_i}{x_i} \binom{T-u_i}{t-x_i}}{\binom{T}{t}} \right\}, i = 1, 2, \dots, S$$

單群落物種數估計

Chao 與 Lin (2012) 針對 *Chao2* 進行修正，針對取後不放回的樣本資料開發新的估計方法。在該估計方法中 $E(Q_k)$ 可以表示為：

$$E(Q_k) = \sum_{i=1}^S P(X_i = k) = \sum_{i=1}^S \frac{\binom{u_i}{k} \binom{T-u_i}{t-k}}{\binom{T}{t}}$$

因此求得在樣本中未出現以及分別出現一次與兩次的物種數之期望值為：

$$E(Q_0) = \sum_{i=1}^S \frac{\binom{T-u_i}{t}}{\binom{T}{t}}$$

$$E(Q_1) = \sum_{i=1}^S \frac{\binom{u_i}{1} \binom{T-u_i}{t-1}}{\binom{T}{t}} = \sum_{i=1}^S \frac{tu_i}{(T-u_i-t+1)} \frac{\binom{T-u_i}{t}}{\binom{T}{t}}$$

$$E(Q_2) = \sum_{i=1}^S \frac{\binom{u_i}{2} \binom{T-u_i}{t-2}}{\binom{T}{t}} = \sum_{i=1}^S \frac{t(t-1)u_i(u_i-1)}{2(T-u_i-t+1)(T-u_i-t+2)} \frac{\binom{T-u_i}{t}}{\binom{T}{t}}$$

根據柯西-施瓦茨不等式之概念可以得：

$$\left[\sum_{i=1}^S \frac{\binom{T-u_i}{t}}{\binom{T}{t}} \right] \left[\sum_{i=1}^S \left(\frac{tu_i}{T-u_i-t+1} \right)^2 \frac{\binom{T-u_i}{t}}{\binom{T}{t}} \right] \geq \left[\sum_{i=1}^S \frac{tu_i}{T-u_i-t+1} \frac{\binom{T-u_i}{t}}{\binom{T}{t}} \right]^2$$

將其中 $\left(\frac{tu_i}{T-u_i-t+1} \right)^2$ 整理以便後續計算：

$$\left(\frac{tu_i}{T-u_i-t+1} \right)^2 = \frac{t}{t-1} \left[\frac{t(t-1)u_i(u_i-1)}{(T-u_i-t+1)^2} \right] + \frac{t^2u_i}{(T-u_i-t+1)^2}$$

則不等式中的第二項可寫作：

$$\left[\sum_{i=1}^S \left(\frac{tu_i}{T-u_i-t+1} \right)^2 \frac{\binom{T-u_i}{t}}{\binom{T}{t}} \right] \approx \frac{2t}{t-1} E(f_2) + \sum_{i=1}^S \frac{t^2u_i}{(T-u_i-t+1)^2} \frac{\binom{T-u_i}{t}}{\binom{T}{t}}$$

由於那些擁有較大 u_i 值的物種對上述方程式中最後一項的貢獻幾乎可以忽略不計。對於那些 u_i 遠小於 T 的物種，我們有以下結果：

$$\frac{t}{T - u_i - t + 1} = \frac{t/T}{(T - u_i - t + 1)/T} \approx \frac{t/T}{(T - t)/T} = \frac{q}{1 - q}$$

故可將不等式整理為：

$$E(Q_0) \left\{ \frac{t}{t-1} 2E(Q_2) + \frac{q}{1-q} E(Q_1) \right\} \geq E(Q_1)^2$$

移項後得：

$$E(Q_0) \geq \frac{E(Q_1)^2}{\frac{t}{t-1} 2E(Q_2) + \frac{q}{1-q} E(Q_1)}$$

最終將結果帶入 $S = S_{obs} + Q_0$ ，可知估計式 $wChao2$ 為：

$$\hat{S}_{wChao2} = S_{obs} + \frac{E(Q_1)^2}{w2E(Q_2) + rE(Q_1)}$$

其中， $w = \frac{t}{t-1}$ ， $r = \frac{q}{1-q}$ 。

兩群落的共同種估計

與取後放回的估計方法相似，在取後不放回的估計中也存在兩群落間的共同種估計需求。在此假設在第一群落的樣本 (X_i) 與第二群落的樣本 (Y_i) 中，分別有 t_1 、 t_2 個抽樣區塊。且兩群落的第 i 物種出現機率分別表示為 π_{1i} 、 π_{2i} 。則 Q_{kl} 表示在出現頻率向量 X_i 中出現 k 個區塊，同時在 Y_i 中出現 l 個區塊的物種

數。在給定 $U_{1i} = u_{1i}$ 與 $U_{2i} = u_{2i}$ 的情況下， $E(Q_{kl})$ 可以表示為：

$$E(Q_{kl}) = \sum_{i=1}^{S_{12}} P(X_i = k)P(Y_i = l) = \sum_{i=1}^{S_{12}} \frac{\binom{u_{1i}}{k} \binom{T_1 - u_{1i}}{t_1 - k}}{\binom{T_1}{t_1}} \frac{\binom{u_{2i}}{l} \binom{T_2 - u_{2i}}{t_2 - l}}{\binom{T_2}{t_2}}$$

同理於取後不放回的單群落物種數估計方法，藉由樣本中分別未出現於兩群落的期望值計算兩群落共同種，可得最終估計式為：

$$\begin{aligned} \hat{S}_{wChao2.12} = & D_{12} + \frac{E(Q_{1+})^2}{w_1 2E(Q_{2+}) + r_1 E(Q_{1+})} + \frac{E(Q_{+1})^2}{w_2 2E(Q_{+2}) + r_2 E(Q_{+1})} \\ & + \frac{E(Q_{11})^2}{4w_1 w_2 E(Q_{22}) + 2w_1 r_2 E(Q_{21}) + 2w_2 r_1 E(Q_{12}) + r_1 r_2 E(Q_{11})} \end{aligned}$$

其中， $w_i = \frac{t_i}{t_i - 1}$ ， $r_i = \frac{q_i}{1 - q_i}$ ， $i = 1, 2$

2.2.4 樣本涵蓋率

樣本涵蓋率的概念起源於第二次世界大戰，艾倫圖靈 (Alan Turing) 與其同事 I.J 古德 (I. J. Good.) 在密碼分析中所開發之估計方法。於出現率樣本中，樣本涵蓋率被定義為樣本中已檢測物種的總出現的比例，即物種的相對出現率總和。在隨機抽樣的假設下，物種出現頻率向量為 (X_1, X_2, \dots, X_S) ，而對抽樣區塊中所觀測到的物種機率估計為 $\hat{\pi}_i = \frac{X_i}{\sum_{i=1}^S X_i}$ ，則樣本覆蓋率可表示為：

$$C = \sum_{i=1}^S \pi_i I(X_i > 0) = 1 - \sum_{i=1}^S \pi_i I(X_i = 0)$$

由上述式子可以得知，物種的相對出現率總和，其值應界於 0 至 1 之間。並且， $1 - C$ 可被解釋為再多觀測一個新的區塊時，該區塊出現新物種的機率。由於該區塊出現的為新物種，固可使用樣本中僅出現一個區塊 (Q_i) 的資訊估計該指標，樣本涵蓋率的估計式為：

$$\hat{C} = 1 - \frac{Q_1}{\sum_{i=1}^S iQ_i}$$

2.2.5 標準差估計

根據 (Q_0, Q_1, \dots, Q_t) 的漸近分布，其服從大小為 S 以及機率為 $\left(\frac{E(Q_0)}{S}, \frac{E(Q_2)}{S}, \dots, \frac{E(Q_t)}{S}\right)$ 的多項分布 (multinomial distribution)。所提出的物種豐富度估計量的變異數估計量可以使用 multivariate delta-method 導出，表示為

$$\widehat{var}(\hat{S}_{12}) = \sum_{i=1}^t \sum_{j=1}^t \frac{\partial \hat{S}_{12}}{\partial Q_i} \frac{\partial \hat{S}_{12}}{\partial Q_j} cov(Q_i, Q_j)$$

$$\text{其中 } cov(Q_i, Q_j) = \begin{cases} Q_i \left(1 - \frac{Q_i}{S}\right), & \text{if } i = j \\ -\frac{Q_i Q_j}{S}, & \text{if } i \neq j \end{cases}$$

2.2.6 95%信賴區間

物種豐富的信賴區間通過假 $\hat{S}_{12} - D_{12}$ 符合對數常態分佈 (log normal distribution) (Chiu et al., 2014)，為此確保了信賴區間之下限值大於觀察到的物種豐富度。故，物種豐富度之 95%信賴區間為：

$$\left[D_{12} + \frac{\widehat{S}_{12} - D_{12}}{R}, D_{12} + (\hat{S}_{12} - D_{12}) \times R \right]$$

第3章 使用動差法估計共同物種數

3.1 取後放回之抽樣方法的估計方式

在單群落的情況下，假設在目標區域實際存在 S 種物種，且實際存在之物種數 S 為一未知參數。且抽樣單位式從目標區域中針對其中的抽樣區塊進行隨機抽樣，並記錄每個區塊中的物種存在與否。若是該樣本總共包含 t 個抽樣區塊，並且 X_i 表示第 i 物種在樣本中出現的區塊數量。則 X_i 遵循參數為 t 且機率為 π_i 的二項分佈 (binomial distribution)。在此， π_i 除了取決於群落規模外，也與其他多種的生物因素相關。

Chiu (2022) 使用的混合二項式模型，建立一個新的針對單群落物種數的估計式。假設出現頻率向量 X_i 遵循二項分佈 $Binomial(t, \pi_i)$ ，其中 π_i ， $i = 1, 2, \dots, S$ ，為機率密度函數為獨立同分佈的隨機變數 $g(\pi)$ 。同時，假設 $g(\pi)$ 服從 $Beta(\alpha, \beta)$ ，獲得以下樣本之物種出現頻率的邊際分佈如下， p_k 為樣本中的物種豐富度正好為 k 的平均機率：

$$\begin{aligned} p_k &= P(X_i = k) = \int_0^1 P(X_i = k | \pi_i = \pi) \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1 - \pi)^{\beta-1} d\pi \\ &= \int_0^1 \binom{t}{k} \pi^k (1 - \pi)^{t-k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1 - \pi)^{\beta-1} d\pi \\ &= \binom{t}{k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + k)\Gamma(\beta + t - k)}{\Gamma(\alpha + \beta + t)}, \quad k = 0, 1, 2, \dots, t \end{aligned}$$

又令 Q_k 表示在 k 個區塊中準確觀測到的物種數，而 $Q_k = \sum_{i=1}^S I(X_i = k)$ 為在單群落樣本中出現 k 個區塊數。並根據柯西-施瓦茨不等式 (Cauchy-Schwarz inequality) 之概念與 Good-Turing 頻率公式 (Good, 1953, 2000) 得出近似式：

$$\frac{tE(Q_0)}{E(Q_1)} \geq \frac{(t-1)E(Q_1)}{2E(Q_2)} \geq \frac{(t-2)E(Q_2)}{3E(Q_3)} \dots。由該近似式可以得知，出現於較少區塊的稀有$$

物種可以為未被觀測到的物種豐富度提供更多的估計資訊。

並根據樣本中物種出現頻率的邊際機率分佈，可知 $E(Q_k)$ 可以表示為：

$$E(Q_k) = S \times P(X = k) = S \binom{t}{k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + k)\Gamma(\beta + t - k)}{\Gamma(\alpha + \beta + t)}$$

依據上述式子，可獲得未觀測以及出現一次至三次的區塊數的期望值：

$$\begin{aligned} E(Q_0) &= S \times P(X = 0) = S \binom{t}{0} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha)\Gamma(\beta + t)}{\Gamma(\alpha + \beta + t)} \\ E(Q_1) &= S \times P(X = 1) = S \binom{t}{1} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 1)\Gamma(\beta + t - 1)}{\Gamma(\alpha + \beta + t)} \\ E(Q_2) &= S \times P(X = 2) = S \binom{t}{2} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 2)\Gamma(\beta + t - 2)}{\Gamma(\alpha + \beta + t)} \\ E(Q_3) &= S \times P(X = 3) = S \binom{t}{3} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + 3)\Gamma(\beta + t - 3)}{\Gamma(\alpha + \beta + t)} \end{aligned}$$

並依據上述概念，將其推廣至兩群落。 p_{kl} 為兩樣本的物種豐富度正好分別為 k 和 l 的平均機率。則：

$$\begin{aligned} p_{kl} &= P(X_i = k)P(Y_i = l) \\ &= \binom{t_1}{k} \binom{t_2}{l} \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1 + k)\Gamma(\beta_1 + t_1 - k)}{\Gamma(\alpha_1 + \beta_1 + t_1)} \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \frac{\Gamma(\alpha_2 + l)\Gamma(\beta_2 + t_2 - l)}{\Gamma(\alpha_2 + \beta_2 + t_2)} \end{aligned}$$

而 k 與 l 的範圍分別為 $k = 0, 1, 2, \dots, t_1$ 與 $l = 0, 1, 2, \dots, t_2$ 。隨後，又令 $Q_{kl} = \sum_{i=1}^S I(X_i = k \text{ and } Y_i = l)$ 表示在第一群落 k 個區塊且在第二群落 l 個區塊數中準確觀測到的物種數。則樣本中觀測到的共同物種數為 $D_{12} = \sum_{i=1}^{T_1} \sum_{i=1}^{T_2} Q_{kl}$ 。藉此，可獲得在第一群落中分別出現未觀測到以及一至二個區塊，且同時在第二群落中出現過至少一次的期望值：

$$\begin{aligned}
E(Q_{0+}) &= S_{12} \times P(X_i = 0) \times (1 - P(Y_i = 0)) \\
&= S_{12} \times \left[\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1)\Gamma(\beta_1 + t_1)}{\Gamma(\alpha_1 + \beta_1 + t_1)} \right] \\
&\quad \times \left\{ 1 - \left[\frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + t_2)}{\Gamma(\alpha_2 + \beta_2 + t_2)} \right] \right\}
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
E(Q_{1+}) &= S_{12} \times P(X_i = 1) \times (1 - P(Y_i = 0)) \\
&= S_{12} \times \binom{t}{1} \left[\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1 + 1)\Gamma(\beta_1 + t_1 - 1)}{\Gamma(\alpha_1 + \beta_1 + t_1)} \right] \\
&\quad \times \left\{ 1 - \left[\frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + t_2)}{\Gamma(\alpha_2 + \beta_2 + t_2)} \right] \right\}
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
E(Q_{2+}) &= S_{12} \times P(X_i = 2) \times (1 - P(Y_i = 0)) \\
&= S_{12} \times G\left(\frac{t}{2}\right) \left[\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1 + 2)\Gamma(\beta_1 + t_1 - 2)}{\Gamma(\alpha_1 + \beta_1 + t_1)} \right] \\
&\quad \times \left\{ 1 - \left[\frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + t_2)}{\Gamma(\alpha_2 + \beta_2 + t_2)} \right] \right\}
\end{aligned} \tag{3.3}$$

將 α 設定為 1，且 $T \gg \beta$ 。可藉由式 (3.1)、式 (3.2) 與式 (3.3) 成立以下近似值：

$$\begin{aligned}
\frac{E(Q_{0+})}{E(Q_{1+})} &= \frac{(\beta_1 + t_1 - 1)}{t_1 \alpha_1} \approx \frac{1}{\alpha_1} \left(\frac{\beta_1}{t_1} + 1 \right) \\
\frac{E(Q_{1+})}{E(Q_{2+})} &= \frac{2}{\alpha_1 + 1} \approx \left(\frac{\beta_1}{t_1} + 1 \right) \\
\frac{E(Q_{2+})}{E(Q_{3+})} &= \frac{3}{\alpha_1 + 1} \approx \left(\frac{\beta_1}{t_1} + 1 \right)
\end{aligned}$$

並經由化簡 $\frac{E(Q_{1+})/E(Q_{2+})}{E(Q_{2+})/E(Q_{3+})}$ 的結果，求得 $\alpha_1 \approx \frac{4E(Q_{2+})^2 - E(Q_{1+})E(Q_{3+})}{3E(Q_{1+})E(Q_{3+}) - 2E(Q_{2+})^2}$ ；又透過化簡

$\frac{E(Q_{0+})/E(Q_{1+})}{E(Q_{1+})/E(Q_{2+})}$ 之後，可求得 $E(Q_{0+}) = \frac{E(Q_{1+})^2}{2E(Q_{2+})} \left(1 + \frac{1}{\alpha_1} \right)$ ；隨後，依據不等式

$\frac{E(Q_{1+})^2}{E(Q_{2+})} \left(\frac{1}{\alpha_1} \right) = \frac{E(Q_{1+})^2}{E(Q_{2+})} \frac{3E(Q_{1+})E(Q_{3+}) - 2E(Q_{2+})^2}{4E(Q_{2+})^2 - E(Q_{1+})E(Q_{3+})} \geq \frac{E(Q_{1+})^2}{E(Q_{2+})} \left(1 - \frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)$ ，故將 $\left(1 - \frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)$

$\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})})$ 作為 $\frac{1}{\alpha_1}$ 帶入 $E(Q_{0+}) = \frac{E(Q_{1+})^2}{2E(Q_{2+})} (1 + \frac{1}{\alpha_1})$ ，最終可獲得：

$$E(Q_{0+}) = \frac{E(Q_{1+})^2}{2E(Q_{2+})} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right)$$

$(A)^-$ 表示：若 $A \leq 1$ 時，則 $(A)^- = \max(0.5, A)$ ；若 $A > 1$ 時，則 $(A)^- = A$ 。同理，

可經由上述相同方式推導出 $E(Q_{+0})$ 與 $E(Q_{00})$ ：

$$E(Q_{+0}) = \frac{E(Q_{+1})^2}{2E(Q_{+2})} \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right)$$

$$E(Q_{00}) = \frac{E(Q_{11})^2}{4E(Q_{22})} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right) \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right)$$

並加入 $\frac{t_i-1}{t_i}$ 對估計式進行修正，最終得估計式：

$$S_{New} = D_{12} + E(Q_{0+}) + E(Q_{+0}) + E(Q_{00})$$

其中，

$$E(Q_{0+}) = \frac{t_1-1}{t_1} \frac{E(Q_{1+})^2}{2E(Q_{2+})} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right)$$

$$E(Q_{+0}) = \frac{t_2-1}{t_2} \frac{E(Q_{+1})^2}{2E(Q_{+2})} \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right)$$

$$E(Q_{00}) = \frac{E(Q_{11})^2}{4E(Q_{22})} \frac{t_1-1}{t_1} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right) \frac{t_2-1}{t_2} \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right)$$

且 $(A)^-$ 表示：若 $A \leq 1$ 時，則 $(A)^- = \max(0.5, A)$ ；若 $A > 1$ 時，則 $(A)^- = A$ 。

3.2 取後不放回之抽樣方法的估計方式

依據 2.2.3 節所示，在目標區塊中檢視到物種 i 所存在的區塊數 U_i ，其分布應服從一個二項分佈；且第 i 物種出現的區塊數所組成的出現頻率向量 X_i ，在給定 $U_i = u_i$ 的情況下，應服從超幾何分佈。又 $P(X_i = k | u_i)$ 與 u_i 有關， X_i 來自於 π_i ，因此可推導出：

$$\begin{aligned} P(X_i = k | \pi_i) &= \sum_{\psi_i} P(X_i = k | u_i) P(u_i | \pi_i) \\ &= \binom{t}{k} \frac{\pi_i^k (1 - \pi_i)^{t-k}}{1 - (1 - \pi_i)^T} \frac{(1 - \pi_i)^T I(k = 0)}{1 - (1 - \pi_i)^T} \end{aligned}$$

其中， $I(A)$ 為指標函數，表示若出現 A 情況時，則該式為 1，反之則即為 0。

Shen 和 He (2008) 針對取後不放回的抽樣方式，開發的 *Beta* 二項式模型：假設 π_i 為一來自 *Beta* 分佈的隨機樣本，故可將式子表示為：

$$f(\pi_i) = G(\alpha, \beta, T) (1 - (1 - \pi_i)^T) \pi_i^{\alpha-1} (1 - \pi_i)^{\beta-1}, 0 < \pi_i < 1$$

其中 $G(\alpha, \beta, T) = \left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)} \right]^{-1}$ 。 X_i 的邊際分佈可通過 $P(X_i = k | \psi_i)$

與 $f(\pi_i)$ 獲得，故樣本之物種出現頻率的邊際分佈如下， p_k 為樣本中的物種豐富度正好為 k 的平均機率：

$$\begin{aligned} p_k = P(X_i = k) &= \int_0^1 P(X_i = k | \pi_i) f(\pi_i) d\pi_i \\ &= \begin{cases} G(\alpha, \beta, T) \left[\frac{\Gamma(\alpha)\Gamma(t+\beta)}{\Gamma(t+\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)} \right], & \text{if } k = 0 \\ G(\alpha, \beta, T) \binom{t}{k} \frac{\Gamma(\alpha+k)\Gamma(t+\beta-k)}{\Gamma(t+\alpha+\beta)}, & \text{if } k > 0 \end{cases} \end{aligned}$$

並令 Q_k 表示在 k 個區塊中準確觀測到的物種數，而 $Q_k = \sum_{i=1}^S I(X_i = k)$ 為在單群落樣本中出現 k 個區塊數。並且 Chiu (2023) 又基於 Good-Turing 頻率公式與柯西不等式之概念，針對單一群落的估計得出近似式： $\frac{TE(Q_0)}{E(Q_1)} \geq \frac{(T-1)E(Q_1)}{2E(Q_2)} \geq \frac{(T-2)E(Q_2)}{3E(Q_3)}$ 。從中可以得知，在物種估計時，採取出現較少次的物種，可以更多提供未出現物種的資訊，有助於縮小物種豐富度的估計結果。根據 $P(X_i = k)$ ，可定義出 $E(Q_k)$ 為：

$$E(Q_k) = P(X_i = k) = \begin{cases} S_{12} \times G(\alpha, \beta, T) \left[\frac{\Gamma(\alpha)\Gamma(t+\beta)}{\Gamma(t+\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)} \right], & \text{if } k = 0 \\ S_{12} \times G(\alpha, \beta, T) \binom{t}{k} \frac{\Gamma(\alpha+k)\Gamma(t+\beta-k)}{\Gamma(t+\alpha+\beta)}, & \text{if } k > 0 \end{cases}$$

依據上述式子可知未觀測到、僅出現在一個與兩個區塊中的豐富度的期望值為：

$$\begin{aligned} E(Q_0) &= P(X_i = 0) = S_{12} \times G(\alpha, \beta, T) \left[\frac{\Gamma(\alpha)\Gamma(t+\beta)}{\Gamma(t+\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T)}{\Gamma(\alpha+\beta+T)} \right] \\ E(Q_1) &= P(X_i = 1) = S_{12} \times G(\alpha, \beta, T) \binom{t}{1} \frac{\Gamma(\alpha+1)\Gamma(t+\beta-1)}{\Gamma(t+\alpha+\beta)} \\ E(Q_2) &= P(X_i = 2) = S_{12} \times G(\alpha, \beta, T) \binom{t}{2} \frac{\Gamma(\alpha+2)\Gamma(t+\beta-2)}{\Gamma(t+\alpha+\beta)} \end{aligned}$$

將隨後利用上述概念，將估計推廣至兩群落。令 p_{kl} 為兩樣本的物種豐富度正好分別為 k 和 l 的平均機率：

$$p_{kl} = P(X_i = k)P(Y_i = l) =$$

$$\begin{cases} G_1 \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] G_2 \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right], \text{if } k = 0 \text{ and } l = 0 \\ G_1 \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] G_2 \binom{t_2}{l} \frac{\Gamma(\alpha_2+l)\Gamma(t_2+\beta_2-l)}{\Gamma(t_2+\alpha_2+\beta_2)}, \text{if } k = 0 \text{ and } l > 0 \\ G_1 \binom{t_1}{k} \frac{\Gamma(\alpha_1+k)\Gamma(t_1+\beta_1-k)}{\Gamma(t_1+\alpha_1+\beta_1)} G_2 \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right], \text{if } k > 0 \text{ and } l = 0 \\ G_1 \binom{t_1}{k} \frac{\Gamma(\alpha_1+k)\Gamma(t_1+\beta_1-k)}{\Gamma(t_1+\alpha_1+\beta_1)} G_2 \binom{t_2}{l} \frac{\Gamma(\alpha_2+l)\Gamma(t_2+\beta_2-l)}{\Gamma(t_2+\alpha_2+\beta_2)}, \text{if } k > 0 \text{ and } l > 0 \end{cases}$$

隨後又令 $Q_{kl} = \sum_{i=1}^S I(X_i = k \& Y_i = l)$ 為在樣本中第一群落出現 k 次且第二群即出現 l 次的區塊數，則 D_{12} 為樣本中觀測到的共同物種數量， $D_{12} = \sum_{k=1}^{T_1} \sum_{l=1}^{T_2} Q_{kl}$ 。藉此，可獲得在第一群落中分別出現未觀測到以及一至二個區塊，且同時在第二群落中出現過至少一次的期望值：

$$\begin{aligned} E(Q_{0+}) &= S_{12} \times P(X_i = 0) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \\ &\quad \times \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned} \quad (3.4)$$

$$\begin{aligned} E(Q_{1+}) &= S_{12} \times P(X_i = 1) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \binom{t_1}{1} \left(\frac{\Gamma(\alpha_1+1)\Gamma(t_1+\beta_1-1)}{\Gamma(t_1+\alpha+\beta_1)} \right) \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned} \quad (3.5)$$

$$\begin{aligned} E(Q_{2+}) &= S_{12} \times P(X_i = 2) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \times \binom{t_1}{2} \left(\frac{\Gamma(\alpha_1+2)\Gamma(t_1+\beta_1-2)}{\Gamma(t_1+\alpha+\beta_1)} \right) \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned} \quad (3.6)$$

依據式 (3.4) 與式 (3.5)。成立以下近似值。將 α 設定為 1，且 $T \gg \beta$ ：

$$\frac{E(Q_{0+})}{E(Q_{1+})} = \left(\frac{\beta_1 + t_1 - 1}{t_1} \frac{T_1 - t_1}{T_1 + \beta_1} \right) \approx \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right]$$

$$\frac{E(Q_{1+})}{E(Q_{2+})} = \frac{t_1 + \beta_1 - 2}{t_1 - 1} \approx \left(\frac{\beta_1}{t_1} + 1 \right)$$

得 $\beta_1 = \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})} \right) t_1$ ，代入 $E(Q_{0+}) = \frac{E(Q_{1+})^2}{E(Q_{2+})} \left(\frac{(T_1 - t_1)}{T_1 + \beta_1} \right)$ 。得：

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))^+}, & \text{if } E(Q_{2+}) > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q} (E(Q_{1+}) - 1), & \text{if } E(Q_{2+}) = 0 \end{cases}$$

$(A)^+$ 表示：若 $A > 0$ 時，則等於 $(A)^+ = A$ ；若 $A \leq 0$ ，則等於 $(A)^+ = 1$ 。

同理 $E(Q_{+0})$ 也依此證明，得 $\beta_2 = \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})} \right) t_2$ ：

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + q(E(Q_{+1}) - E(Q_{+2}))^+}, & \text{if } E(Q_{+2}) > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q} (E(Q_{+1}) - 1), & \text{if } E(Q_{+2}) = 0 \end{cases}$$

$(A)^+$ 表示：若 $A > 0$ 時，則等於 $(A)^+ = A$ ；若 $A \leq 0$ ，則等於 $(A)^+ = 1$ 。

同理，又可得在兩群群落同時為未觀測到、觀測到一至二個區塊的期望值分別為：

$$\begin{aligned}
E(Q_{00}) &= S_{12} \times P(X_i = 0) \times P(Y_i = 0) \\
&= S_{12} G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\
&\quad \times \left[\frac{\Gamma(\alpha_1) \Gamma(t_1 + \beta_1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} - \frac{\Gamma(\alpha_1) \Gamma(\beta_1 + T_1)}{\Gamma(\alpha_1 + \beta_1 + T_1)} \right] \\
&\quad \times \left[\frac{\Gamma(\alpha_2) \Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2) \Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right]
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
E(Q_{11}) &= S_{12} \times P(X_i = 1) \times P(Y_i = 1) \\
&= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\
&\quad \times \binom{t_1}{1} \binom{t_2}{1} \left(\frac{\Gamma(\alpha_1 + 1) \Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 1) \Gamma(t_2 + \beta_2 - 1)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right)
\end{aligned} \tag{3.8}$$

$$\begin{aligned}
E(Q_{22}) &= S_{12} \times P(X_i = 2) \times P(Y_i = 2) \\
&= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\
&\quad \times \binom{t_1}{2} \binom{t_2}{2} \left(\frac{\Gamma(\alpha_1 + 2) \Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 2) \Gamma(t_2 + \beta_2 - 2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right)
\end{aligned} \tag{3.9}$$

並依據式 (3.7) 與式 (3.8) 成立以下近似值：

$$\begin{aligned}
\frac{E(Q_{00})}{E(Q_{11})} &= \left(\frac{\beta_1 + t_1 - 1}{t_1} \frac{T_1 - t_1}{T_1 + \beta_1} \right) \left(\frac{\beta_2 + t_2 - 1}{t_2} \frac{T_2 - t_2}{T_2 + \beta_2} \right) \\
&\approx \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right]
\end{aligned}$$

$$\text{將} \begin{cases} \beta_1 = \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})} \right) t_1 \\ \beta_2 = \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})} \right) t_2 \end{cases}, \text{代入 } E(Q_{00}) \approx \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right] E(Q_{11})$$

後，並加入 $\left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \right)$ 對估計式進行調整，最終得估計式：

$$S_{wNew1} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中：

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))^+}, & \text{if } E(Q_{2+}) > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q} (E(Q_{1+}) - 1), & \text{if } E(Q_{2+}) = 0 \end{cases} \quad (3.10)$$

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + w_2(E(Q_{+1}) - E(Q_{+2}))^+}, & \text{if } E(Q_{+2}) > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q} (E(Q_{+1}) - 1), & \text{if } E(Q_{+2}) = 0 \end{cases} \quad (3.11)$$

$$E(Q_{00}) = \left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \right) \times \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \times \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right] E(Q_{11}) \quad (3.12)$$

$(A)^+$ 表示：若 $A > 0$ 時，則等於 $(A)^+ = A$ ；若 $A \leq 0$ ，則等於 $(A)^+ = 1$ 。

並在 S_{wBB1} 的基礎上，加入 Q_{12} 、 Q_{21} 對 β_1 、 β_2 的估計進行修正，依據式(3.8)

與式(3.9)成立以下近似值：

$$\frac{E(Q_{11})}{E(Q_{22})} \approx \left(\frac{\beta_1}{t_1} + 1 \right) \left(\frac{\beta_2}{t_2} + 1 \right)$$

並經由該式可推得出：

$$\begin{cases} \beta_1 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_2}{\beta_2 + t_2} \right) t_1 - t_1 \\ \beta_2 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_1}{\beta_1 + t_1} \right) t_2 - t_2 \end{cases} \quad (3.13)$$

又經由 $E(Q_{kl}) = S_{12} \times P(X_i = k) \times P(Y_i = l)$ 可以得知， $E(Q_{12})$ 以及 $E(Q_{21})$ 分別為：

$$\begin{aligned} E(Q_{12}) &= S_{12} \times P(X_i = 1) \times P(Y_i = 2) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\ &\times \binom{t_1}{1} \binom{t_2}{2} \left(\frac{\Gamma(\alpha_1 + 1) \Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 2) \Gamma(t_2 + \beta_2 - 2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{aligned} \quad (3.14)$$

$$\begin{aligned} E(Q_{21}) &= S_{12} \times P(X_i = 2) \times P(Y_i = 1) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\ &\times \binom{t_1}{2} \binom{t_2}{1} \left(\frac{\Gamma(\alpha_1 + 2) \Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 1) \Gamma(t_2 + \beta_2 - 1)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{aligned} \quad (3.15)$$

並依式 (3.14) 與式 (3.15) 成立以下近似式：

$$\frac{E(Q_{12})}{E(Q_{21})} = \left(\frac{t_2 - 1}{t_1 - 1} \right) \left(\frac{t_1 + \beta_1 - 2}{t_2 + \beta_2 - 2} \right) \approx \left(\frac{t_2}{t_1} \right) \left(\frac{t_1 + \beta_1}{t_2 + \beta_2} \right)$$

由上述式子推得：

$$\begin{cases} \beta_1 = \frac{E(Q_{12})}{E(Q_{21})} \left(\frac{t_1}{t_2} \right) (t_2 + \beta_2) - t_1 \\ \beta_2 = \frac{E(Q_{21})}{E(Q_{12})} \left(\frac{t_2}{t_1} \right) (t_1 + \beta_1) - t_2 \end{cases} \quad (3.16)$$

又可從式 (3.13) = 式 (3.16) 得：

$$\begin{aligned} (t_1 + \beta_1)^2 &= \frac{E(Q_{11}) E(Q_{12})}{E(Q_{22}) E(Q_{21})} t_1^2 \Rightarrow \beta_2^2 + 2t_2\beta_2 + t_2^2 \left(1 - \frac{E(Q_{11}) E(Q_{12})}{E(Q_{22}) E(Q_{21})} \right) = 0 \\ (t_2 + \beta_2)^2 &= \frac{E(Q_{11}) E(Q_{21})}{E(Q_{22}) E(Q_{12})} t_2^2 \Rightarrow \beta_2^2 + 2t_2\beta_2 + t_2^2 \left(1 - \frac{E(Q_{11}) E(Q_{21})}{E(Q_{22}) E(Q_{12})} \right) = 0 \end{aligned}$$

並依公式解 $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ ，得：

$$\begin{cases} \beta_1 = -t_1 + t_1 \sqrt{\frac{E(Q_{11}) E(Q_{12})}{E(Q_{22}) E(Q_{21})}} = \left(-1 + \sqrt{\frac{E(Q_{11}) E(Q_{12})}{E(Q_{22}) E(Q_{21})}} \right) t_1 \\ \beta_2 = -t_2 + t_2 \sqrt{\frac{E(Q_{11}) E(Q_{21})}{E(Q_{22}) E(Q_{12})}} = \left(-1 + \sqrt{\frac{E(Q_{11}) E(Q_{21})}{E(Q_{22}) E(Q_{12})}} \right) t_2 \end{cases} \quad (3.17)$$

最終得：

$$S_{wNew2} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中，將式 (3.17) 的 β_1 與 β_2 的結果分別帶入 $E(Q_{0+}) = \frac{E(Q_{1+})^2}{E(Q_{2+})} \left(\frac{T_1 - t_1}{T_1 + \beta_1} \right)$ 與

$E(Q_{+0}) = \frac{E(Q_{+1})^2}{E(Q_{+2})} \left(\frac{T_2 - t_2}{T_2 + \beta_2} \right)$ 中可分別求得 $E(Q_{0+})$ 與 $E(Q_{+0})$ 的估計式；而 $E(Q_{00})$

則與 $wNew1$ 相同使用式 (3.12) 估計。

第4章 模擬研究與討論

4.1 估計式整理

整理上一章所推倒之估計式，以及過去文獻中所存在的估計式。藉由電腦模擬的方式，在不同模型的設定下，比較三種取後不放回以及兩種取後放回的估計方式之表現，並同時估計其標準差。

以下為不同估計方法之整理：

4.1.1 取後放回

1. S_{New}

$$S_{New} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中：

$$\begin{aligned} E(Q_{00}) &= \frac{E(Q_{11})^2}{4E(Q_{22})} \frac{t_1 - 1}{t_1} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right) \frac{t_2 - 1}{t_2} \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right) \\ E(Q_{0+}) &= \frac{t_1 - 1}{t_1} \frac{E(Q_{1+})^2}{2E(Q_{2+})} \left(2 - \left(\frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} \right)^- \right) \\ E(Q_{+0}) &= \frac{t_2 - 1}{t_2} \frac{E(Q_{+1})^2}{2E(Q_{+2})} \left(2 - \left(\frac{2E(Q_{+2})^2}{3E(Q_{+1})E(Q_{+3})} \right)^- \right) \end{aligned}$$

且 $(A)^-$ 表示：若 $A \leq 1$ 時，則 $(A)^- = \max(0.5, A)$ ；若 $A > 1$ 時，則 $(A)^- = A$ 。

2. S_{Pan}

$$S_{Pan} = D_{12} + \frac{t_1 - 1}{t_1} \frac{E(Q_{1+})^2}{2E(Q_{2+})} + \frac{t_2 - 1}{t_2} \frac{E(Q_{+1})^2}{2E(Q_{+2})} + \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{E(Q_{11})^2}{4E(Q_{22})}$$

4.1.2 取後不放回

3. S_{wNew1}

$$S_{wNew1} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

其中：

$$E(Q_{00}) = \left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \right) \times \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \times \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right] E(Q_{11})$$

並在 $E(Q_{00})$ 的估計式中代入 $\begin{cases} \beta_1 = \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})} \right) t_1 \\ \beta_2 = \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})} \right) t_2 \end{cases}$ 。

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))^+}, & \text{if } E(Q_{2+}) > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q} (E(Q_{1+}) - 1), & \text{if } E(Q_{2+}) = 0 \end{cases}$$

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + w_2(E(Q_{+1}) - E(Q_{+2}))^+}, & \text{if } E(Q_{+2}) > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q} (E(Q_{+1}) - 1), & \text{if } E(Q_{+2}) = 0 \end{cases}$$

$(A)^+$ 表示：若 $A > 0$ 時，則等於 $(A)^+ = A$ ；若 $A \leq 0$ ，則等於 $(A)^+ = 1$ 。

4. S_{wNew2}

$$S_{wNew2} = D_{12} + E(Q_{00}) + E(Q_{0+}) + E(Q_{+0})$$

$$E(Q_{00}) = \left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \right) \times \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \times \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right] E(Q_{11})$$

並在 $E(Q_{00})$ 的估計式中代入 $\begin{cases} \beta_1 = \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{12})}{E(Q_{22})E(Q_{21})}} \right) t_1 \\ \beta_2 = \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{21})}{E(Q_{22})E(Q_{12})}} \right) t_2 \end{cases}$ 。而 $E(Q_{+0})$ 與

$E(Q_{0+})$ 的估計與 S_{wNew1} 相同。

5. $S_{12, \text{wor2}}$

$$S_{12, \text{wor2}} = D_{12} + \frac{Q_{1+}^2}{2k_1 Q_{2+} + r Q_{1+}} + \frac{Q_{+1}^2}{2k_2 Q_{+2} + r Q_{+1}} + \frac{Q_{11}^2}{4k_1 k_2 Q_{22} + 2k_1 r_2 Q_{21} + 2k_2 r_1 Q_{12} + r_1 r_2 Q_{11}}$$

其中 $k_i = \frac{t_i}{t_i - 1}$ 與 $r_j = \frac{\frac{t_j}{T}}{(1 - \frac{t_j}{T})}$ 。

4.1.3 標準差、95%信賴區間、樣本涵蓋率與變異係數估計

6. 標準差估計

$$\widehat{var}(\widehat{S}_{12}) = \sum_{i=1}^t \sum_{j=1}^t \frac{\partial \widehat{S}_{12}}{\partial Q_i} \frac{\partial \widehat{S}_{12}}{\partial Q_j} \widehat{cov}(Q_i, Q_j)$$

其中 $\widehat{cov}(Q_i, Q_j) = \begin{cases} Q_i \left(1 - \frac{Q_i}{\hat{S}}\right), & \text{if } i = j \\ -\frac{Q_i Q_j}{\hat{S}}, & \text{if } i \neq j \end{cases}$

7. 95%信賴區間

(1) $wNew1$ and $wNew2$

$$\left[\hat{S}_{12} - 1.96 \left(\widehat{var}(\widehat{S}_{12}) \right)^{\frac{1}{2}}, \hat{S}_{12} + 1.96 \left(\widehat{var}(\widehat{S}_{12}) \right)^{\frac{1}{2}} \right]$$

(2) Other

$$\left[D_{12} + \frac{\widehat{S}_{12} - D_{12}}{R}, D_{12} + (\hat{S}_{12} - D_{12}) \times R \right]$$

其中 $R = \left\{ 1.96 \left[\log \left(1 + \frac{\widehat{var}(\widehat{S}_{12})}{(\widehat{S}_{12} - D_{12})^2} \right) \right]^{\frac{1}{2}} \right\}$ 以此計算 95%信賴區間的樣本涵蓋率

(95% confidence interval coverage rate, 95% CI Coverage)。

8. Sample Coverage

為樣本中檢測到的物種的總發生率的百分比。

$$\hat{C} = 1 - \frac{Q_1}{\sum_{i=1}^S iQ_i}$$

9. 樣本變異係數 (Sample CV)

$$\hat{\gamma}^2 = \max \left\{ \frac{S_{obs}}{\hat{C}} \frac{t}{t-1} \frac{\sum_{i=1}^t i(i-1)Q_i}{\sum_{i=1}^t iQ_i \sum_{i=1}^t iQ_i - 1} - 1, 0 \right\}$$

$\hat{\gamma}^2$ 為變異係數平方的估計式，即 $CV = \hat{\gamma}$ 。

4.2 模型模擬設定

4.2.1 模擬模型假設

經由電腦模擬，可以設定不同母體並產生不同物種豐富度的結構，透過重複抽樣資料以評估估計式之估計表現。在模擬的過程中，首先需決定兩族群的共同種與特有種的數量，再選擇兩種不同種模型作為母體使用，以下為四種模擬模型之設定：

- I. 同質性模型 (homogeneity model), $p_i = 0.1$, $i = 1, 2, \dots, S$ ($mean = 0.1, CV = 0$)
- II. 均勻模型 (uniform model), $p_i = ca_i$, $i = 1, 2, \dots, S$ ($mean = 0.12, CV = 0.66$), 其中 a_i 服從均勻分佈, c 為調整常數。
- III. Broken-stick 模型, $p_i = ca_i$, $i = 1, 2, \dots, S$ ($mean = 0.15, CV = 0.94$), 其中 a_i 服從指數函數分佈, c 為調整常數。
- IV. 對數常態模型 (log-normal model), $p_i = ca_i$, $i = 1, 2, \dots, S$ ($mean = 0.15, CV = 1.17$), 其中 a_i 服從對數常態分佈, c 為調整常數。

4.2.2 物種與區塊數的假設

在電腦模擬時，需針對真實的群落物種與區塊數進行假設，以評估估計結果之優劣程度。故假設以下四種組合為電腦模擬的真實物種與區塊狀況做使用。

1. 假設總物種數為 500 種物種 ($S = 500$)，其中群落一與第二群落皆存在 400 種物種 ($S_1 = S_2 = 400$)，兩群落間共有 300 種共同種 ($S_{12} = 300$)。且兩族群皆為 100 區塊 ($T_1 = T_2 = 100$)，並從中依比例抽取 t_1 與 t_2 個區塊作為樣本使用。
2. 假設總物種數皆為 700 種物種 ($S = 700$)，其中群落一與第二群落分別存在 400 與 600 種物種 ($S_1 = 400$, $S_2 = 600$)，兩群落間共有 300 種共同種 ($S_{12} = 300$)。且兩族群皆為 100 區塊 ($T_1 = T_2 = 100$)，並從中依比例抽取 t_1 與 t_2 個區塊作為樣本使用。

3. 假設總物種數為 500 種物種 ($S = 500$)，其中群落一與第二群落皆存在 400 種物種 ($S_1 = S_2 = 400$)，兩群落間共有 300 種共同種($S_{12} = 300$)；兩族群皆為 100 區塊 ($T_1 = T_2 = 100$)，並從中依比例抽取 t_1 與 t_2 個區塊作為樣本使用，而 $t_2 = 2t_1$ 。
4. 假設總物種數為 500 種物種 ($S = 500$)，其中群落一與第二群落皆存在 400 種物種 ($S_1 = S_2 = 400$)，兩群落間共有 300 種共同種($S_{12} = 300$)。且兩族群分別為 100 與 200 區塊 ($T_1 = 100, T_2 = 200$)，並從中依比例抽取 $t_1 = qT_1$ 與 $t_2 = qT_2$ 個區塊作為樣本使用。

而在每種群落假設下，樣本數的模擬次數為 $R = 1000$ 次。

4.3 表格中名詞定義

- q ：為抽樣比例， $q = \frac{t_i}{T_i}$ 。
- Sample size：抽樣區塊數， t 。
- Estimator：所採用之估計方法。
- Obs：兩樣本觀測到的共同物種。
- AVG Estimate： R 次模擬之平均估計值。
- Bias： R 次模擬之偏差 (bias)， $bias = \hat{S}_{12} - S_{12}$ 。
- Sample SE： R 次模擬之樣本標準差。
- Est. SD： R 次模擬之標準差估計值。
- RMSE： R 次模擬估計量之樣本均方根誤差。
- 95% CI Coverage： R 次模擬中理論值於信賴區間的比例。

4.4 取後放回的模擬結果

首先針對樣本涵蓋率型計算，在多種組合模擬下，除了在樣本大小為 10 的情況下，其餘樣本大小所抽樣本，無論是在何種模型或模擬組合下，所估計之樣本涵蓋率皆大於 0.95。

且依據模擬結果可以得知，在第一種物種分配的假設下，在偏差的表現上，在四種模擬情況下，兩估計方法下所估計的群落共同種，皆有低估的情況發生。尤其在小樣本時，低估情況更加明顯。平均而言，在兩種估計方法的偏差結果中皆顯示，*New* 所估計之群落共同種相較於原有的 *Pan* 方法，在偏差的估計表現上更為優秀。在樣本標準差 (Sample SE) 以及平均標準差估計 (Est. SD) 的結果中，可以發現無論在何種模擬情形下兩種方法皆有低估的情況發生。然而，單憑偏差與標準差的估計的結果無法從中判定估計式的好壞，在評估一個估計方法的優劣之時，還需要考慮該估計量的準確度以及精確度。因此以 RMSE 作為衡量估計式好壞之準則。在 RMSE 的結果方面，在兩群落皆為小樣本時，*New* 的表現較優秀。但在大樣本地抽樣下，*Pan* 的表現會略優於 *New*。另一方面，95%信賴區間涵蓋率也是能作為評估估計式優劣的標準之一。在 95%信賴區間涵蓋率的結果中 *New* 的結果皆略優於 *Pan*。

在第二種與第三種物種分配的假設下，所進行的模擬結果，無論是在何種估計值的結果，皆與第一種物種分配的假設並無太大的差異。此外，針對第四種物種的假設下，無論是在觀測物種或是平均的估計結果上，相較於第一種假設在 *New* 的估計下，雖然也具有較小的偏差，但會有高估的情況發生；在 RMSE 的表現上，*New* 的估計也獲得較差的結果；同時，隨著樣本數的增加，在 95%信賴區間涵蓋率的結果反而下降。

綜上所述，在評估各項模擬與所有指標之後，本文所提出 *New* 估計方式的表現普遍優於 *Pan* 對於兩群落共同種的估計結果。且在目標區塊數以及共同物種數的真值皆為相同的情況下，群落中的特有種數量並不會影響共同物種數

的估計結果。但當目標區塊與抽樣區塊不相同時，則會影響到估計式的表現結果。

Table 4.1 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	125.12	270.79	-29.21	61.1	46.64	67.69	0.86
	Pan		246.81	-53.19	40.24	27.93	66.68	0.82
30	New	240.32	287.18	-12.82	19.64	18.34	23.44	0.91
	Pan		277.62	-22.38	13.64	11.92	26.2	0.89
50	New	267.3	294.68	-5.32	16.04	14.36	16.89	0.87
	Pan		288.29	-11.71	11.15	9.45	16.16	0.86
70	New	277.28	297.93	-2.07	14.03	12.42	14.18	0.83
	Pan		292.66	-7.34	9.78	8.19	12.22	0.82

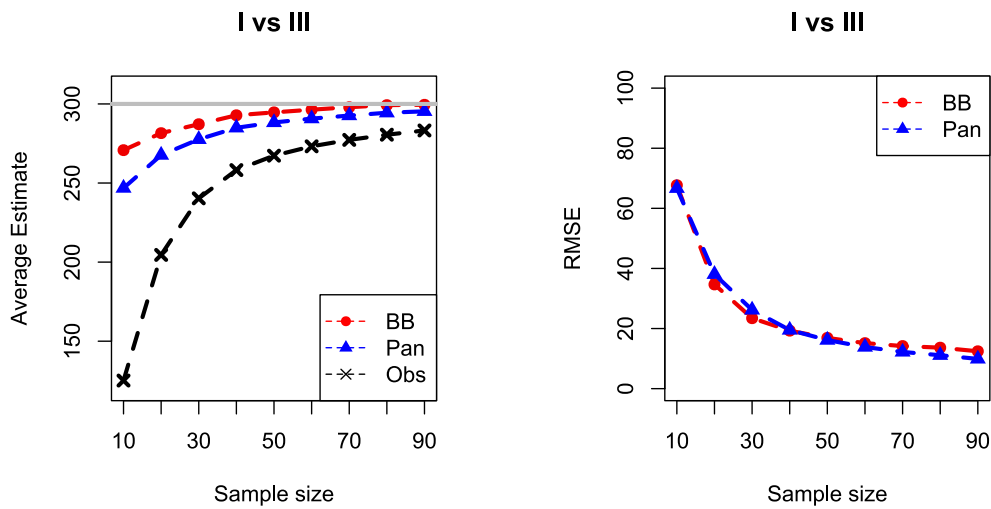


Figure 4.1 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Table 4.2 兩群落皆為均勻模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	123.34	244.33	-55.67	52.13	40.13	76.25	0.86
	Pan		220.02	-79.98	32.53	22.51	86.34	0.83
30	New	225.85	277.64	-22.36	22.77	19.67	31.9	0.89
	Pan		265.65	-34.35	15.31	12.33	37.6	0.88
50	New	255.45	289.5	-10.5	16.01	15.42	19.14	0.9
	Pan		281.66	-18.34	11.37	10.15	21.58	0.89
70	New	268.74	295.39	-4.61	13.68	13.72	14.43	0.91
	Pan		289.14	-10.86	9.76	9.21	14.6	0.9

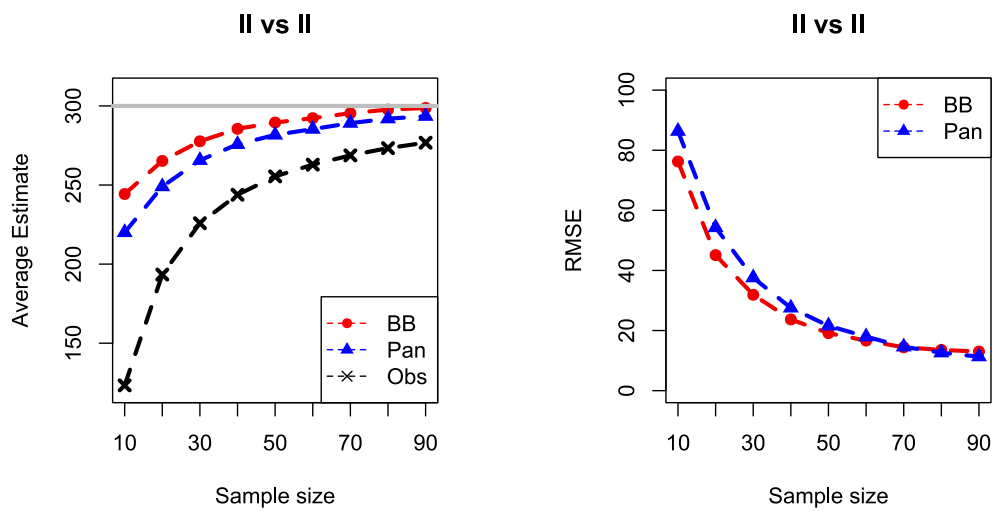


Figure 4.2 兩群落皆為均勻模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Table 4.3 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	123.27	232.31	-67.69	47.01	37.17	82.4	0.85
	Pan		207.78	-92.22	29.49	20.99	96.82	0.83
30	New	218.54	276.15	-23.85	24.65	21.91	34.29	0.89
	Pan		262.17	-37.83	16.27	13.71	41.18	0.88
50	New	248.57	290.11	-9.89	19.9	17.98	22.21	0.88
	Pan		280.02	-19.98	13.76	11.77	24.26	0.88
70	New	263.41	295.6	-4.4	17.15	15.73	17.69	0.88
	Pan		287.96	-12.04	12.17	10.46	17.11	0.86

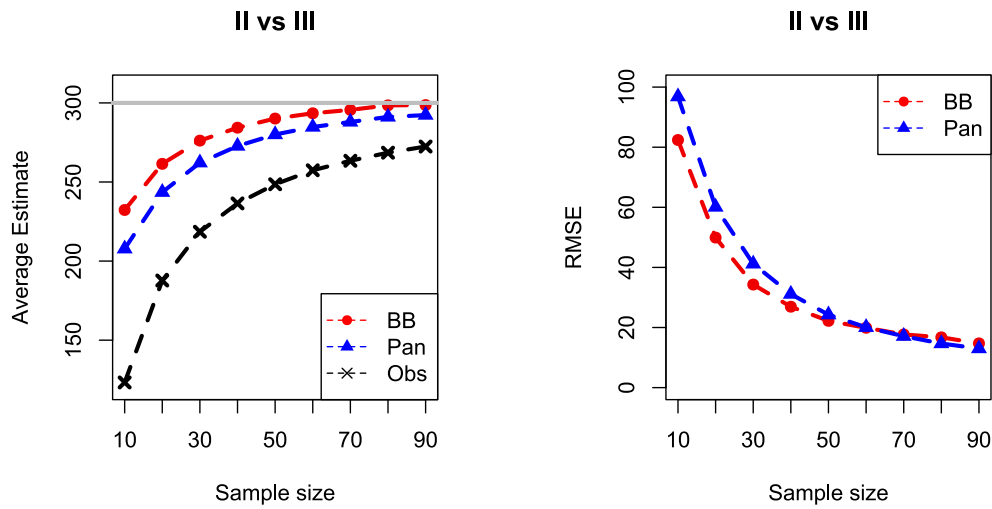


Figure 4.3 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Table 4.4 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	115.62	228.28	-71.72	52.12	40.34	88.64	0.85
	Pan		199.49	-100.51	31.57	22.73	105.34	0.83
30	New	208.96	277.73	-22.27	29.47	25.47	36.93	0.9
	Pan		261.16	-38.84	18.92	15.57	43.2	0.88
50	New	242.64	292.56	-7.44	22.86	20.58	24.03	0.88
	Pan		280.66	-19.34	15.44	13.18	24.74	0.87
70	New	260.08	298.13	-1.87	19.15	17.44	19.23	0.88
	Pan		289.36	-10.64	13.26	11.47	16.99	0.88

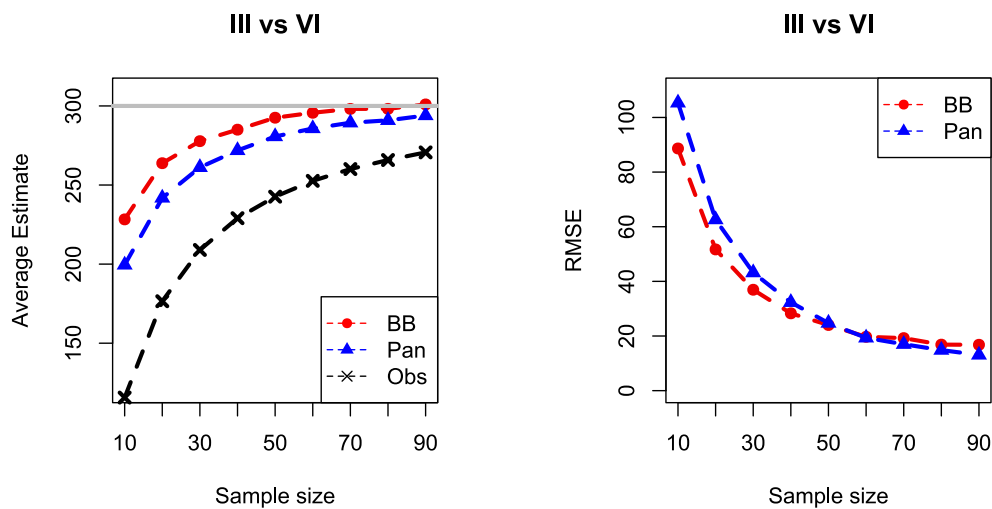


Figure 4.4 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

4.5 取後不放回與取後放回的模擬結果比較

在取後不放回的模擬針對樣本涵蓋率型計算，同樣於在多種組合模擬下，除了在抽樣比例為 0.1 的情況底下，其餘樣本大小所抽樣本，無論是在何種模型或模擬組合下，所估計之樣本涵蓋率皆大於 0.95。且模擬同質群落與均勻群落在抽樣比例為 0.1 的情況下，樣本變異係數通常小於 0.4。而隨抽樣比例的增加將越接近母體變異數。

依據第一種物種假設的模擬結果可以得知，在偏差的表現上，在 *wNew1* 與 *wNew2* 小樣本的估計，在同質與 Broken-stick 以及兩群落皆為均勻的兩種模擬組合下有高估的表現；而在均勻與 Broken-stick 以及 Broken-stick 與常對數的兩種模擬組合下有低估的表現。但在大樣本的情況下，則是三種估計方法皆呈現低估的表現。

在樣本標準差 (Sample SE) 以及平均標準差估計 (Est. SD) 的結果中，可以發現無論在何種模擬情形下三種方法皆有低估的情況發生。除此之外，在 RMSE 的結果中 *wNew1* 與 *wNew2* 兩者無明顯差異，且都優於 *wChao2*。在 95%信賴區間涵蓋率方面，由於 *wChao2* 已知為下界估計式，因此在 95%信賴區間涵蓋率的估計上使用 log 修正的方式進行估計，而在 *wNew1* 與 *wNew2*，則使用未修正的估計方式。在該項結果下 *wNew1* 的表現優於其他兩者，且隨著抽樣比例的增加，*wNew1* 與 *wNew2* 無明顯差異，且略優於 *wChao2*。

在第二種與第三種物種分配的假設底下，所進行的模擬結果，無論是在何種估計值的結果，皆與第一種物種分配的假設並無太大的差異。並且，不同於第四種物種的假設在取後放回的抽樣模擬中與前三種的估計結果上偶差異，在取後不放回的模擬下，四種假設的結果皆無明顯差異。

綜合以上論述可以得知，普遍而言在小樣本時，取後不放回的兩估計結果較不穩定，會因群落之變異大小所致，使得高估或低估的現象皆有可能發生。而依據多次模擬結果可以得知，通常當其中一個群落的估計之樣本變異係數小

於 0.4 時 ($\hat{p} < 0.4$)， $wNew1$ 與 $wNew2$ 更容易發生高估的情況。反之在大樣本的條件下，採用取後放回的估計方法 New 則是會有明顯高估的情況發生。

因此在小樣本且當其中一個群落樣本所估計的變異係數小於 0.4 時的估計中，建議採用取後放回的抽樣方式 (New) 估計兩群落的共同種；而在抽樣比例達到 0.4 時，取後不放回的估計式才趨近穩定，並相較於取後放回的方法有更良好的估計表現。故在抽樣比例大於 0.4 時，須以取後不放回的兩估計是對群落的共同種進行估計。

Table 4.5 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	130.68	362.99	62.99	57.31	51.9	85.14	0.92
	wNew2		367.63	67.63	67.25	56.69	95.36	0.92
	wChao2		246.53	-53.47	30.75	24.33	61.67	0.89
	New		282.03	-17.97	57.47	45.49	60.19	0.89
0.3	wNew1	251.5	301.72	1.72	12.87	9.29	12.97	0.86
	wNew2		303.33	3.33	17.58	14.6	17.88	0.9
	wChao2		281.55	-18.45	9.7	9.15	20.85	0.92
	New		300.26	0.26	19.47	19.11	19.46	0.91
0.5	wNew1	277.66	298.82	-1.18	6.77	7.46	6.87	0.98
	wNew2		297.5	-2.5	6.1	5.87	6.59	0.94
	wChao2		292.87	-7.13	6.15	5.81	9.41	0.89
	New		313.19	13.19	16.35	16.84	21	0.91
0.7	wNew1	289.19	299.95	-0.05	3.67	4.06	3.67	0.97
	wNew2		298.72	-1.28	3.32	3.51	3.56	0.97
	wChao2		298.19	-1.81	3.62	3.74	4.05	0.9
	New		324.51	24.51	14.65	17.23	28.55	0.94

Table 4.6 兩群落皆為均勻模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	127.72	306.03	6.03	44.38	44.69	44.77	0.95
	wNew2		310.52	10.52	52.3	43.21	53.32	0.91
	wChao2		215.98	-84.02	23.97	19.41	87.37	0.87
	New		246.21	-53.79	46.24	38.37	70.92	0.89
0.3	wNew1	234.14	293.65	-6.35	15.26	14.24	16.52	0.93
	wNew2		298.53	-1.47	21.44	17.07	21.48	0.9
	wChao2		269.61	-30.39	11.25	9.92	32.4	0.89
	New		294.95	-5.05	24.75	21.59	25.25	0.9
0.5	wNew1	266.2	296.79	-3.21	8.9	8.52	9.46	0.93
	wNew2		299.18	-0.82	10.62	8.93	10.65	0.9
	wChao2		287.64	-12.36	7.82	6.8	14.62	0.89
	New		315.69	15.69	20.48	19.29	25.79	0.9
0.7	wNew1	282.94	299.05	-0.95	5.2	4.97	5.28	0.94
	wNew2		299.36	-0.64	5.33	4.82	5.37	0.93
	wChao2		296.33	-3.67	5.08	4.55	6.26	0.89
	New		337.83	37.83	20.16	21.22	42.87	0.94

Table 4.7 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	126.47	290.58	-9.42	43.78	38.74	44.76	0.92
	wNew2		298.77	-1.23	55.94	44.91	55.92	0.92
	wChao2		208.8	-91.2	24.3	19.13	94.38	0.86
	New		243.7	-56.3	47.86	38.89	73.87	0.88
0.3	wNew1	226.97	292.39	-7.61	16.63	14.85	18.28	0.93
	wNew2		292.88	-7.12	22.97	17.59	24.04	0.89
	wChao2		266.41	-33.59	12.31	10.79	35.77	0.89
	New		296.35	-3.65	27.04	24.18	27.27	0.88
0.5	wNew1	261.52	296.89	-3.11	8.5	9.19	9.05	0.97
	wNew2		295.16	-4.84	10.14	9.17	11.23	0.94
	wChao2		286.59	-13.41	7.55	7.44	15.38	0.92
	New		320.96	20.96	20.55	21.94	29.35	0.94
0.7	wNew1	281.06	299.58	-0.42	5.35	5.33	5.36	0.94
	wNew2		298.73	-1.27	5.41	5.07	5.56	0.93
	wChao2		296.51	-3.49	5.24	4.9	6.3	0.9
	New		342.9	42.9	20.83	23.26	47.69	0.95

Table 4.8 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	121.47	282.17	-17.83	45.02	41.28	48.4	0.94
	wNew2		289.76	-10.24	56.53	47.66	57.42	0.92
	wChao2		202.82	-97.18	25.04	19.94	100.36	0.87
	New		243.48	-56.52	50.92	41.92	76.06	0.87
0.3	wNew1	223.77	300.21	0.21	17.93	19.62	17.93	0.97
	wNew2		301.38	1.38	24.33	19.37	24.36	0.91
	wChao2		269.92	-30.08	13.26	11.78	32.87	0.91
	New		304.82	4.82	31.38	27.43	31.74	0.91
0.5	wNew1	262.89	300.71	0.71	9.14	9.81	9.16	0.97
	wNew2		301.44	1.44	11.58	9.83	11.66	0.9
	wChao2		289.26	-10.74	8.02	7.55	13.4	0.91
	New		321.77	21.77	22.08	21.99	31	0.94
0.7	wNew1	283.16	300.2	0.2	4.86	5.12	4.86	0.96
	wNew2		300.39	0.39	5.08	5.03	5.09	0.94
	wChao2		296.82	-3.18	4.69	4.58	5.66	0.9
	New		329.86	29.86	17.54	19.58	34.62	0.96

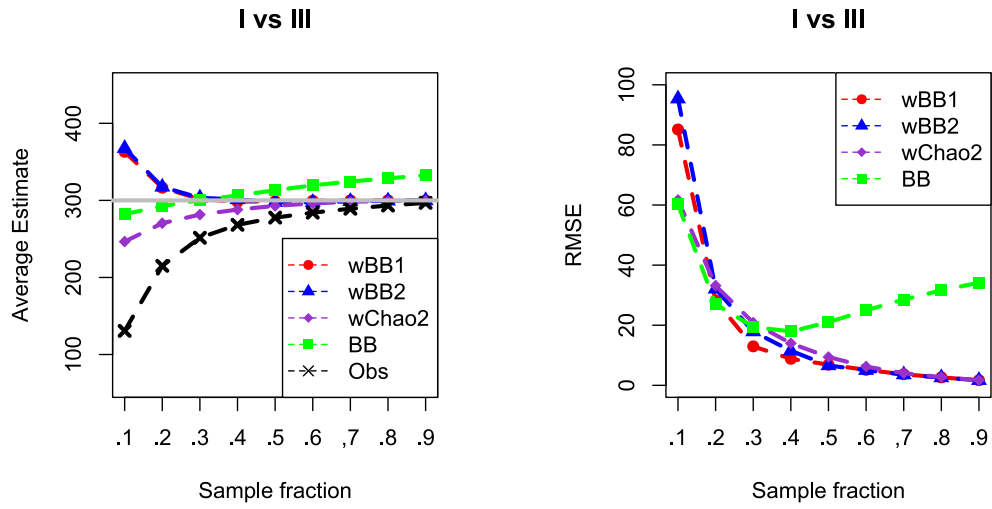


Figure 4.5 群落一為同質模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

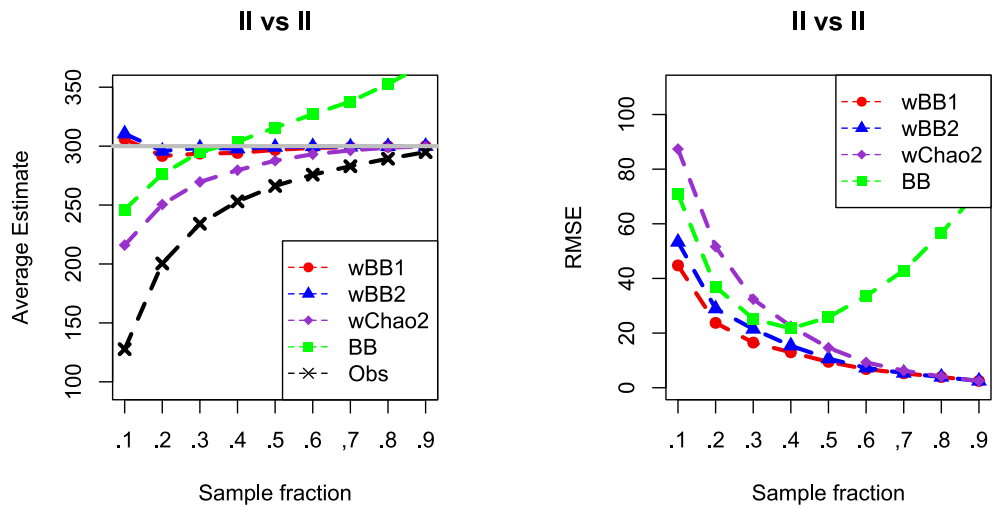


Figure 4.6 兩群落皆為均勻模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

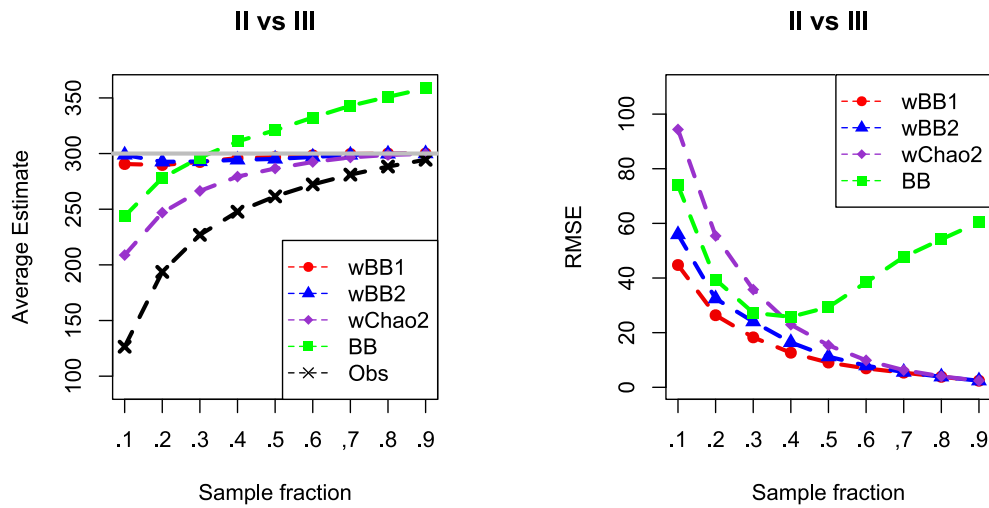


Figure 4.7 群落一為均勻模型，群落二為 Broken-stick 模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

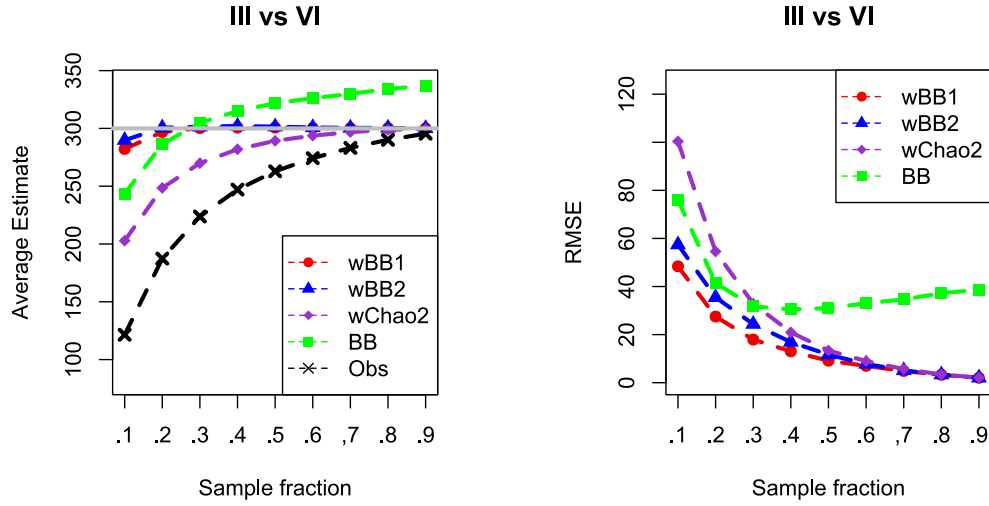


Figure 4.8 群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的估計結果。 $S_1 = S_2 = 400$ ， $S_{12} = 300$ 。

4.6 真實資料模擬

除了使用模擬母體進行重複抽樣的模擬之外，亦使用兩筆真實數據做為母體進行 $R = 1000$ 次的重複抽樣，以評斷估計式的結果優劣。

4.6.1 澳洲三種極端氣候鳥類資料

資料描述

針對澳洲東南部半乾旱地區的尤加利樹林地，量化三個氣候時期 178 個地點的極端氣候對鳥類物種出現 (species occurrence)、物種豐富度 (species richness) 和出現率 (incidence) 的影響——聖嬰現象相關的乾旱 (Big Dry)、反聖嬰現象破壞性降雨 (Big Wet)，以及反聖嬰現象事件三年後 (Post-Big Wet)。並使用 Big Dry 作為群落一使用 (Mean = 0.05, CV = 1.65)，其中共存在 55 種物種；Big Wet 作為群落二使用 (Mean = 0.13, CV = 1.53)，其中共存在 65 種物種。Post-Big Wet 作為群落三使用 (Mean = 0.12, CV = 1.61)，其中共存在 56 種物種。

模擬結果

在該筆資料中使用取後放回的估計方法下，假設該筆資料為母體，並抽取其中的區塊做為樣本所使用。使用 *New* 的估計結果與真值的差異，明顯表現較 *Pan* 佳。在估計式評估標準上，使用 RMSE 與 95% CI Coverage 進行評估，在兩種評估標準的結果表明，在小樣本中，RMSE 在 *New* 估計方式中的評估結果劣於 *Pan*；而在 95% CI Coverage 的模型評估上，兩者並無明顯差異，值得注意的是在本資料模擬中，隨著抽樣比例的增加，95% CI Coverage 的結果會隨之下降。

Table 4.9 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Big Wet 估計結果。

sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
20	New	25.54	40.84	-8.16	15.08	12.58	17.13	0.8
	Pan		37.54	-11.46	10.53	8.79	15.56	0.81
60	New	37.39	46.03	-2.97	9.69	8.18	10.13	0.75
	Pan		44.16	-4.84	6.97	5.89	8.48	0.77
100	New	41.48	47.89	-1.11	8.92	6.78	8.98	0.7
	Pan		46.44	-2.56	6.49	4.94	6.97	0.72
140	New	43.66	48.91	-0.09	7.21	6.13	7.21	0.69
	Pan		47.72	-1.28	5.35	4.46	5.49	0.71

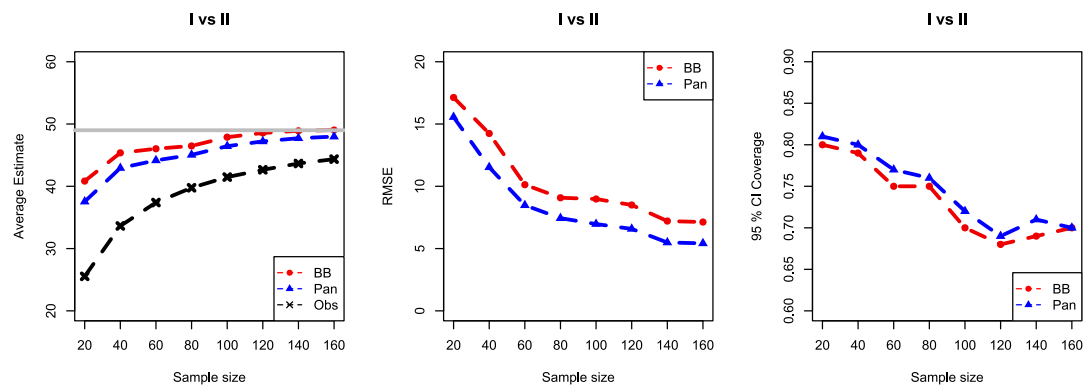


Figure 4.9 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Big Wet 估計結果。

Table 4.10 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Post-Big Wet 估計結果。

sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
20	New	24.27	40.19	-11.81	14.49	13.1	18.69	0.82
	Pan		36.69	-15.31	10.14	9.13	18.36	0.84
60	New	37.16	47.37	-4.63	10.54	9.4	11.51	0.8
	Pan		45.27	-6.73	7.85	6.67	10.33	0.79
100	New	41.93	49.97	-2.03	9.39	8	9.6	0.76
	Pan		48.09	-3.91	6.9	5.76	7.93	0.75
140	New	44.6	51.59	-0.41	8.53	7.41	8.53	0.77
	Pan		49.88	-2.12	6.22	5.31	6.56	0.76

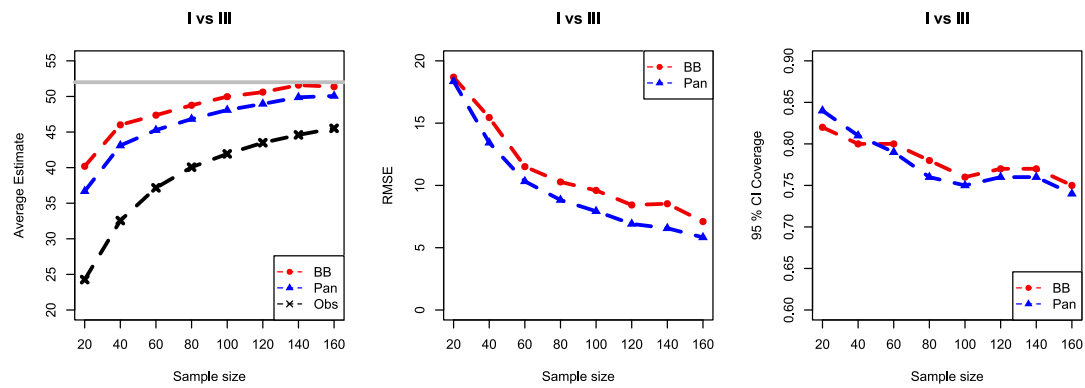


Figure 4.10 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Post-Big Wet 估計結果。

Table 4.11 取後放回澳洲三種極端氣候鳥類資料集中的 Big Wet 與 Post-Big Wet 估計結果。

sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
20	New	30.82	45.86	-10.14	14.51	12.74	17.7	0.8
	Pan		42.88	-13.12	10.5	8.99	16.8	0.81
60	New	42.91	51.52	-4.48	9.09	8.42	10.13	0.8
	Pan		49.88	-6.12	6.76	6.13	9.12	0.8
100	New	47.31	54.12	-1.88	8.64	7.23	8.84	0.76
	Pan		52.73	-3.27	6.3	5.29	7.1	0.77
140	New	49.53	55.47	-0.53	7.6	6.77	7.62	0.74
	Pan		54.15	-1.85	5.59	4.97	5.88	0.75

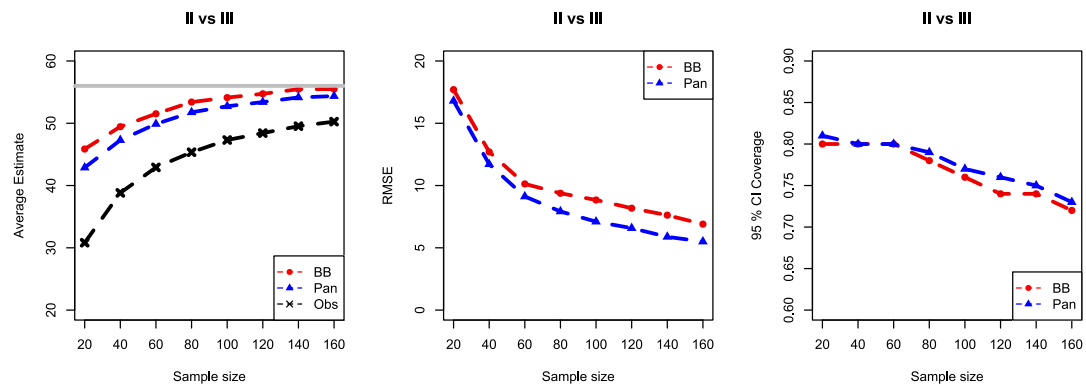


Figure 4.11 取後放回澳洲三種極端氣候鳥類資料集中的 Big Wet 與 Post-Big Wet 估計結果。

4.6.2 BCI

資料描述

本資料集中，針對 1981-2016 年間對 Barro Colorado Island 進行八次調查，紀錄該地區樹種與莖幹的生長狀況。將該地區非為 1250 個區塊進行調查，每塊區塊的大小為 400 平方公尺。以第一次與第八次的調查結果作為兩群落之母體資料，並使用母體的 10%、30%、50% 以及 70% 作為抽樣樣本大小。在兩母體中，第一次調查作為群落一 (Mean = 0.16, CV = 1.41)，第八次調查作為群落二 (Mean = 0.16, CV = 1.45)，在群落一中，包含了 307 種物種，而群落二中則擁有 299 種物種，兩群落的共同物種為 284 種。

模擬結果

在該筆資料中使用取後不放回的估計方法下，假設該筆資料為母體，並抽取其中的區塊做為樣本所使用。在模擬結果中，使用 *wNew2* 的估計方法在小樣本中與真值差異最小。在抽樣比例為 0.1 時平均的樣本標準差估計 (Average estimated SE)，與樣本標準差 (Sample SE) 存在較大的差異，隨著抽樣比例增加差異逐漸減小。在估計式評估標準上，使用 RMSE 與 95% CI Coverage 進行評估，在兩種評估標準的結果表明，在小樣本中，*wNew1* 與 *wNew2* 的估計明顯較 *wChao2* 的結果較好，但在隨著樣本變大，RMSE 在三種估計方式中的評估結果沒有明顯差異；而在 95% CI Coverage 的模型評估上，*wNew1* 在小樣本中的結果明顯較其他兩者好。

Table 4.12 取後不放回 BCI 資料集中的兩群落估計結果。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	225.24	272.6	-11.4	24.29	26.06	26.82	0.97
	wNew2		297.63	13.63	47.35	34.37	49.25	0.92
	wChao2		253.31	-30.69	19.56	13.09	36.38	0.81
	New		263.89	-20.11	26.59	21.9	33.33	0.85
0.3	wNew1	245.65	281.82	-2.18	17.38	15.93	17.51	0.95
	wNew2		291.12	7.12	24.41	18.74	25.41	0.9
	wChao2		266.94	-17.06	12.53	9.38	21.17	0.83
	New		279.86	-4.14	23.86	19.53	24.2	0.84
0.5	wNew1	256.57	282.64	-1.36	11.78	10.51	11.85	0.93
	wNew2		287.81	3.81	14.91	11.69	15.38	0.88
	wChao2		272.93	-11.07	9.05	7.26	14.3	0.84
	New		285.87	1.87	20.04	17.11	20.12	0.85
0.7	wNew1	263.97	284.41	0.41	8.81	7.86	8.82	0.92
	wNew2		286.98	2.98	10.12	8.19	10.55	0.89
	wChao2		277.76	-6.24	7.26	6.06	9.57	0.85
	New		293.78	9.78	20.78	17.52	22.95	0.85

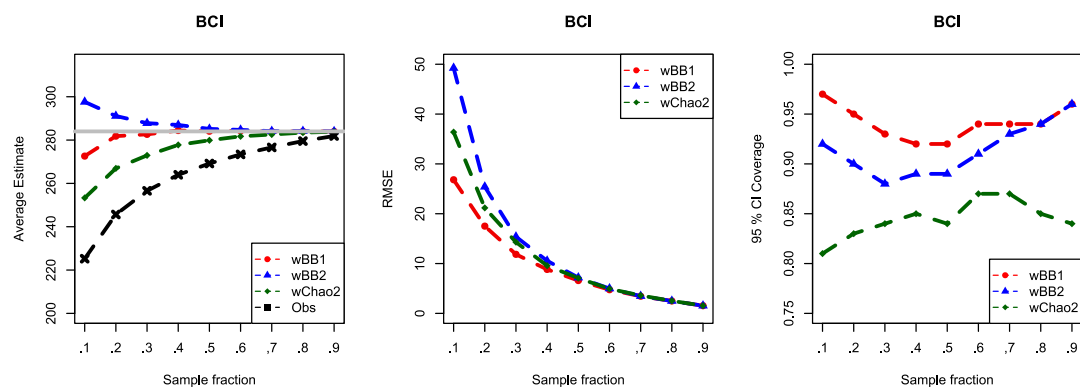


Figure 4.12 取後不放回 BCI 資料集中的兩群落估計結果。

4.7 討論

總結上述結果，在錯誤! 找不到參照來源。中使用了電腦模擬建立不同模型假設下的四種群落搭配四種不同的物種分配方式，以及分別使用兩筆真實資料作為群落母體。並從中隨機以取後放回以及取後不放回兩種不同的抽樣方式，重複抽取 1000 次，以計算平均的估計值、偏差、樣本標準差、族群標準差、RMSE 以及 95%信賴區間涵蓋率的估計結果。

在取後放回的估計方法中，其模擬結果顯示，不同樣本大小和物種分配假設對群落共同種估計的影響。且在第一種物種分配假設下，兩種估計方法都低估了共同種數，尤其在小樣本時更為明顯。*New* 方法相較於 *Pan* 方法在偏差和 RMSE 方面表現較好。95%信賴區間的涵蓋率方面，*New* 方法略優於 *Pan* 方法。而在第二和第三種假設下，估計結果與第一種假設相似。但是，在第四種假設下，*New* 方法偏差較小但容易高估，RMSE 較差，且隨著樣本增加，95%信賴區間涵蓋率下降。真實資料方面，澳洲三種極端氣候鳥類資料的模擬結果中，*New* 估計方式比 *Pan* 具有更好的表現，誤差較小。在評估標準上，*New* 在小樣本中的 RMSE 稍遜色於 *Pan*，但在 95% CI Coverage 上表現相近。

另一方面，在取後不放回的估計中，大多數情況下樣本涵蓋率都高於 0.95。隨著抽樣比例增加，樣本變異係數趨於接近母體變異數。在第一種物種分配假設下，不同模型和組合下，*wNew1* 和 *wNew2* 方法在小樣本中有稍微高估現象發生，在大樣本中則普遍為低估。在 RMSE 方面，*wNew1* 和 *wNew2* 表現相近且皆優於 *wChao2*。在 95%信賴區間涵蓋率方面，*wNew1* 表現為所有中最佳者。並且不同於取後放回的估計結果，在任何物種分配假設下結果皆無明顯差異。而在真實資料的部分，BCI 資料的模擬結果顯示，使用 *wNew2* 估計方法在小樣本中與真值的差異最小。抽樣比例為 0.1 時，平均估計的樣本標準差與實際樣本標準差存在較大差異，但隨著抽樣比例增加，差異逐漸減小。在評估中，RMSE 和 95% CI Coverage 顯示，在小樣本中，*wNew1* 和 *wNew2* 比

wChao2 表現更好。隨著樣本增加，三種估計方法的 RMSE 沒有明顯差異，但在 95% CI Coverage 方面，*wNew1* 在小樣本中表現明顯優於其他兩者。

總結上述，本文比較了取樣後放回和不放回的估計方法對群落共同種估計的影響。兩者在樣本涵蓋率、變異性和偏差方面表現類似。兩種抽樣方法中所獲得估計結果，除了取後不放回的小樣本之外，其餘皆低估了共同種數。在大多數情況下，樣本涵蓋率在大多數情況下都高於 0.85。總體而言，綜觀模擬出的各項指標與估計式評估結果，本文所提出的建議方法 (*New*, *wNew1*, *wNew2*) 在多數情況下表現優於原有方法 (*Pan*, *wChao2*)。但在取後不放回的估計中某些特定情況，特別是在小樣本和低估計方面，可能存在高估的風險，且對於不同假設條件下的估計結果表現也有所不同。

第5章 實例分析

5.1 紅杉國家公園內苔蘚資料

在紅杉國家公園內苔蘚 (Wilson and Coleman, 2022) 資料集中，蒐集加州內華達山脈 (California's Sierra Nevada) 的西部坡地，對苔蘚植物進行調查，範圍從海拔 380 公尺到 3,578 公尺之間，選擇 25 平方公尺的區塊作為抽樣地區。其中包含夏季乾燥且冬季溫暖的氣候地區的丘陵地區，以及夏季短、具有積雪的高山地區。在本資料集中依照海拔高低大致分為以下四個群落：

1. 山麓 (F, foothill) 海拔 1200 公尺以下，該地區以地中海型氣候為主要特徵：夏季炎熱乾燥；冬季則為涼爽，通常都在冰點以上。降水形式以降雨為主。其中包含 67 個區塊以及 132 種物種 (Mean = 0.08, CV = 1.17)，且僅出現在單一區塊的物種具有 48 種。該地區的樣本覆蓋率為 0.91。
2. 下針葉林帶 (LC, lower conifer) 海拔 1200 至 2440 公尺，在該地區主要的植被有巨型紅杉、雪松與白冷杉等。在下針葉林帶，主要的降水量以降雨和降雪為主。該地區氣候在夏季時非常乾燥，但在冬天時的積雪會滲透到土壤中變得相對涼爽。其中包含 100 個區塊以及 162 種物種 (Mean = 0.05, CV = 1.17)，且僅出現在單一區塊的物種具有 59 種。該地區的樣本覆蓋率為 0.93。
3. 上針葉林 (UC, upper conifer) 海拔 2440 至 2750 公尺，該地區氣溫較低，生長季較短。該地區被河流切割，與其他地區相比土壤相對較少，裸岩較多，且腐木會迅速乾枯，較不利於苔蘚生長。其中包含 17 個區塊以及 53 種物種 (Mean = 0.11, CV = 0.66)，且僅出現在單一區塊的物種具有 28 種。該地區的樣本覆蓋率為 0.74。

4. 高地地區 (H, high country) 海拔 2750 公尺以上，該地區冬季寒冷且漫長，夏季短而乾燥，大部分的水分來自於夏季融雪。植被相較其他區域更矮小，土壤為裸岩與碎石居多。但因融雪灌溉，因此該地潮濕的棲息地不斷增加，包括溪流、季節性滲漏 (seepages)、湖泊、草地和沼澤等地形。其中包含 69 個區塊以及 92 種物種 (Mean = 0.07, CV = 1.28)，且僅出現在單一區塊的物種具有 39 種。該地區的樣本覆蓋率為 0.91。

5.1.1 以取後放回的抽樣方法估計

在假設該筆資料為取後放回的抽樣方式，並依照相對應的估計方法分別針對各群落進行估計的結果，在山麓、下針葉林、上針葉林以及高地地區，各自分別擁有 273.9 ± 59.1 、 261.4 ± 33.9 、 96.9 ± 31.8 以及 159.9 ± 46 種物種，該估計結果為平均數±標準差。

隨後將該資料及做為樣本使用，估計兩群落之間共同物種數，在多數估計，無論是在共同物種數的估計值或是標準差的估計結果中，*BB* 所估計之結果高於 *Pan* 所估計之，此情況與上一章所呈現之電腦模擬的結果相符。

並經由 Jaccard 指數量化四群落間的 *Beta* 多樣性結果可以得知，山麓地區與下針葉林地地區具有最相似的群落組成，而與其不相鄰的兩區域則具有最高的群落相異性；在上針葉林地地區方面，與其相鄰的兩群落比起不相鄰的山麓地區具有更高的相似性。值得注意的是，在下針葉林的部分，在比較其與上針葉林地與高地地區的群落相異程度之後，發現在與之不相鄰的高地地區反而相較於相鄰的上針葉林地地區具有更高的群落相似性。

繪製在未使用估計式修正僅使用觀測值作為群落物種數，與使用估計式修正群落物種數後之估計結果的樹狀結構圖 (dendrogram)。在分類樹結果中，可以看到，僅由觀測值作為依據所繪製之分類樹結果，將山麓地區與下針葉林分為一類群，而上針葉林與高地地區做為第二類群；但在估計出的結果中，則是

在山麓地區與下針葉林具有最近的距離，隨後與該群相近的為高地地區，而在上針葉林則與其餘三群落具有更遠的距離。在使用估計式修正群落內物種數所量化之 *Beta* 多樣性結果，相較於直接使用觀測物種數所估計之結果之間會有所不同。

綜上所述，可知在分析群落間的 *Beta* 多樣性時，使用估計式修正群落物種數可能比直接使用觀測值計算的結果更能反映群落間的真實相似性關係。

Table 5.1 取後放回之紅杉國家公園內苔蘚的兩群落間共同種估計結果。

	Estimator	Obs	Estimate	Est. SD	95% CI bound	Jaccard
F & LC	BB	80	158.39	33.81	(114.89, 256.13)	0.53
	Pan		129.11	19.11	(103.53, 182.52)	0.47
F & UC	BB	21	53.13	17.61	(32.77, 108.72)	0.81
	Pan		41.03	11.42	(28.08, 77.68)	0.78
F & H	BB	24	101.43	41.27	(53.06, 230.33)	0.67
	Pan		73.19	27.2	(41.88, 159.37)	0.63
LC & UC	BB	43	62.42	16.19	(47.68, 123.66)	0.8
	Pan		58.36	8.48	(48.59, 85.21)	0.77
LC & H	BB	53	91.03	28.46	(63.29, 193.54)	0.71
	Pan		85.36	14.62	(66.91, 128.3)	0.67
UC & H	BB	30	41.25	12.27	(31.99, 93.77)	0.82
	Pan		41.7	7.11	(33.9, 65.12)	0.79

Table 5.2 取後放回之紅杉國家公園內苔蘚的兩群落間物種數與 Jaccard 估計結果。

	Estimator	Estimate	Jaccard
F & LC	Obs	214	0.63
	BB	339.99	0.53
	Chao2	297.99	0.47
F & UC	Obs	164	0.87
	BB	277.32	0.81
	Chao2	239.54	0.78
F & H	Obs	200	0.88
	BB	311.4	0.67
	Chao2	274.27	0.63
LC & UC	Obs	172	0.75
	BB	314.92	0.8
	Chao2	267.28	0.77
LC & H	Obs	201	0.74
	BB	316.99	0.71
	Chao2	278.33	0.67
UC & H	Obs	115	0.74
	BB	233.65	0.82
	Chao2	194.1	0.79

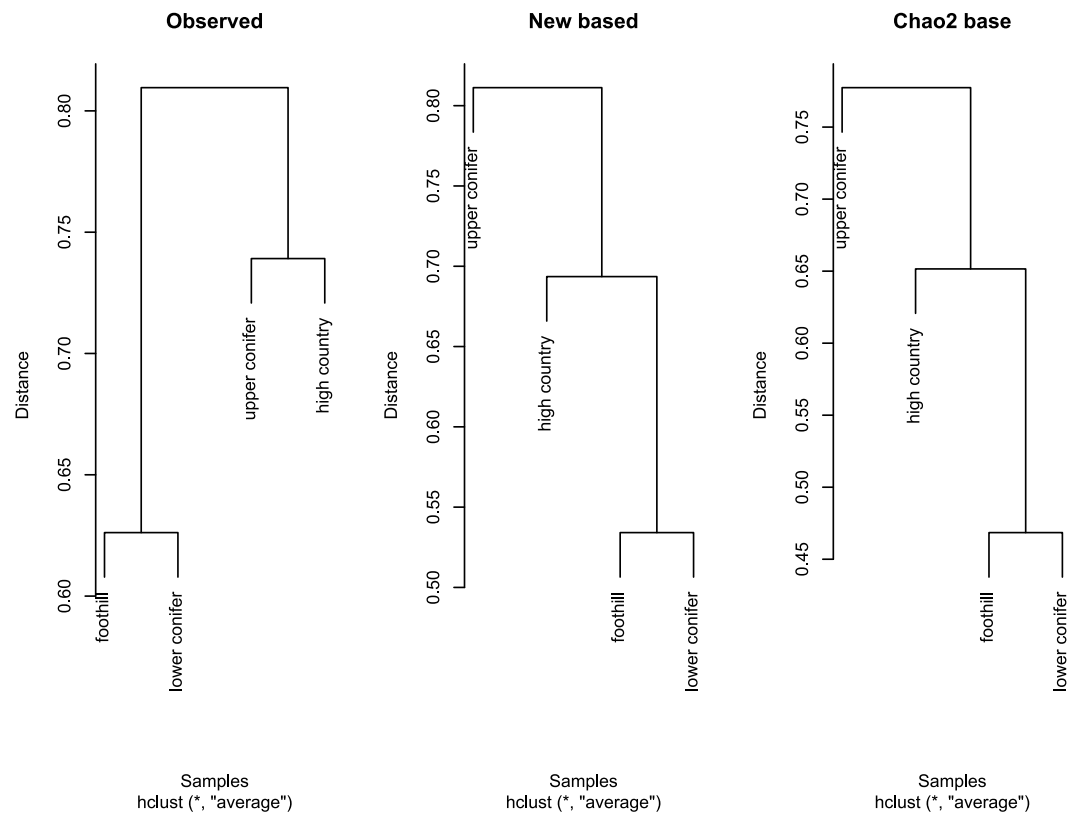


Figure 5.1 利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

5.1.2 以取後不放回的抽樣方法估計

在假設該筆資料為取後不放回的抽樣方式，並依照相對應的估計方法分別針對各群落進行估計的結果。且由於該資料集中尚未提及抽樣比例，故假設在抽樣比例為 0.3、0.5 與 0.7 的結果，估計四個群落的物種數與兩群落之間的共同物種數。

Table 5.3 不同抽樣比例之單群落物種數估計 (mean \pm SD)。

Community	0.3	0.5	0.7
foothill	196.7 \pm 7.1	168.4 \pm 5.6	150.1 \pm 4.1
lower conifer	241.5 \pm 8.7	221.9 \pm 7.4	205.5 \pm 6.2
upper conifer	47.5 \pm 1.4	35.7 \pm 5.1	28.9 \pm 6.8
high country	137.2 \pm 6.5	119.9 \pm 5.2	107.1 \pm 3.8

將該資料及做為樣本使用，估計兩群落之間的共同物種數。隨著抽樣比例的增加，所估計共同物種數隨之減少，這是由於在取後不放回的估計中，抽樣比例為重要參數，將影響估計結果。在量化四群落間的 *Beta* 多樣性結果中，同樣也獲得大多數情況下兩相鄰的群落會較不相鄰的群落之間，具有更高的群落相似性的結果，並在下針葉林以及高地地區的結果中有所例外。而在分類樹的建構中，當假設的抽樣比例較小時，與取後放回的估計結果相似——山麓地區與下針葉林地區為一類群，而上針葉林地區與高地地區則是分別作為一類群。其次，與山麓地區與下針葉林地區的類群更為相似的群落為高地地區。而隨著假設的抽樣比例的增加，所繪製之估計的分類樹結果，在結構方面與觀測值更為相似，且群落之間的距離差異增加。

綜上所述，結合上取後放回的估計結果，可以推斷在分類樹建構中，使用不同的估計方式所估計之 *Beta* 多樣性結果有所差異，隨之會導致生成的分類樹也有所不同。這表明在選用估計方式時，需針對不同取樣方式的樣本，配合相對應之估計方法，才得以確保估計時的準確性。

Table 5.4 假設抽樣比例為 0.3 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。

	Estimator	Obs	Est.	Est. SD	95% CI bound	Jaccard
F & LC	wNEW	80	135.45	19.2	(97.81, 173.09)	0.54
	wChao2		116.69	12.11	(99.53, 148.92)	0.49
F & UC	wNEW	21	50.41	16	(19.04, 81.77)	0.78
	wChao2		46.94	12.97	(31.28, 86.45)	0.76
F & H	wNEW	24	78.78	20.79	(38.03, 119.53)	0.71
	wChao2		80.71	26.26	(47.91, 158.54)	0.69
LC & UC	wNEW	43	64.86	9.53	(46.19, 83.53)	0.74
	wChao2		56.64	6.61	(48.55, 76.57)	0.72
LC & H	wNEW	53	90.06	13.4	(63.79, 116.33)	0.67
	wChao2		75.78	8.92	(63.87, 100.75)	0.64
UC & H	wNEW	30	45.65	7.26	(31.42, 59.88)	0.75
	wChao2		39.19	5.19	(33.28, 55.76)	0.72

Table 5.5 假設抽樣比例為 0.5 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。

	Estimator	Obs	Est.	Est. SD	95% CI bound	Jaccard
F & LC	wNEW	80	111.47	9.14	(93.56, 129.38)	0.57
	wChao2		103.86	7.52	(93.05, 123.63)	0.56
F & UC	wNEW	21	36.07	6.3	(23.73, 48.41)	0.82
	wChao2		34.25	5.72	(26.89, 50.81)	0.82
F & H	wNEW	24	45.93	7.16	(31.9, 59.96)	0.81
	wChao2		45.63	7.03	(35.62, 64.25)	0.8
LC & UC	wNEW	43	57.17	5.62	(46.16, 68.18)	0.74
	wChao2		53.18	4.77	(47.26, 67.36)	0.73
LC & H	wNEW	53	75.25	7.49	(60.57, 89.92)	0.69
	wChao2		69.22	6.07	(60.97, 85.99)	0.68
UC & H	wNEW	30	40.28	4.6	(31.26, 49.29)	0.73
	wChao2		37.12	3.91	(32.6, 49.47)	0.72

Table 5.6 假設抽樣比例為 0.7 下，紅杉國家公園內苔蘚各群落之間共同物種數之估計結果。

	Estimator	Obs	Est.	Est. SD	95% CI bound	Jaccard
F & LC	wNEW	80	95.76	4.91	(86.14, 105.38)	0.6
	wChao2		93.58	4.58	(87.13, 105.85)	0.59
F & UC	wNEW	21	28.39	3.35	(21.82, 34.96)	0.85
	wChao2		27.87	3.27	(23.83, 37.68)	0.84
F & H	wNEW	24	33.03	3.61	(25.95, 40.11)	0.85
	wChao2		32.91	3.61	(28.16, 43.13)	0.85
LC & UC	wNEW	43	51.35	3.55	(44.39, 58.3)	0.74
	wChao2		49.85	3.31	(45.79, 59.82)	0.73
LC & H	wNEW	53	64.6	4.23	(56.31, 72.88)	0.71
	wChao2		62.73	3.87	(57.59, 73.62)	0.71
UC & H	wNEW	30	36.04	2.98	(30.19, 41.88)	0.73
	wChao2		34.87	2.77	(31.72, 43.76)	0.73

Table 5.7 在不同抽樣比例的假設下，取後不放回之紅杉國家公園內苔蘚的兩群落間物種數與 Jaccard 估計結果。

	Estimator	Estimate			Jaccard		
		0.3	0.5	0.7	0.3	0.5	0.7
F & LC	Obs		214			0.63	
	BB	291.31	259.99	237.65	0.54	0.57	0.6
	Chao2	268.06	250.64	234.92	0.49	0.56	0.59
F & UC	Obs		164			0.87	
	BB	231.39	205.05	184.81	0.78	0.82	0.85
	Chao2	212.48	196.81	182.7	0.76	0.82	0.84
F & H	Obs		200			0.88	
	BB	274.34	245.6	223.28	0.71	0.81	0.85
	Chao2	250.35	235.22	220.71	0.69	0.8	0.85
LC & UC	Obs		172			0.75	
	BB	250.59	217.49	194.98	0.74	0.74	0.74
	Chao2	229.44	209.56	192.78	0.72	0.73	0.73
LC & H	Obs		201			0.74	
	BB	276.55	246.42	224.34	0.67	0.69	0.71
	Chao2	252.48	236.61	221.71	0.64	0.68	0.71
UC & H	Obs		115			0.74	
	BB	180.07	151.74	133.09	0.75	0.73	0.73
	Chao2	161.75	145.26	131.59	0.72	0.72	0.73

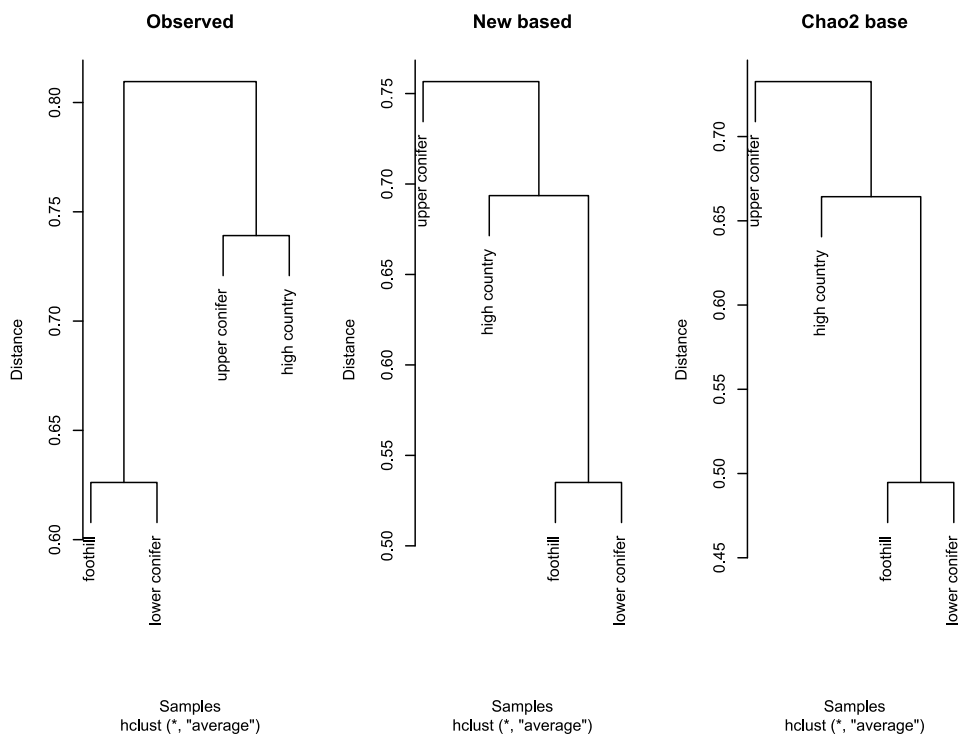


Figure 5.2 假設抽樣比例為 0.3 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

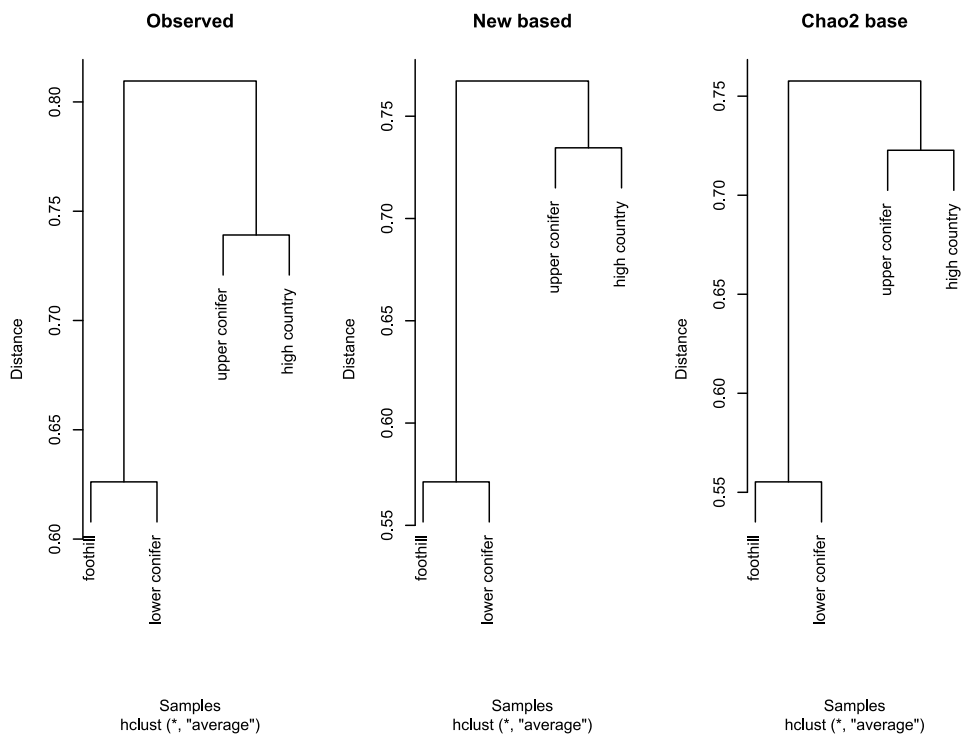


Figure 5.3 假設抽樣比例為 0.5 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

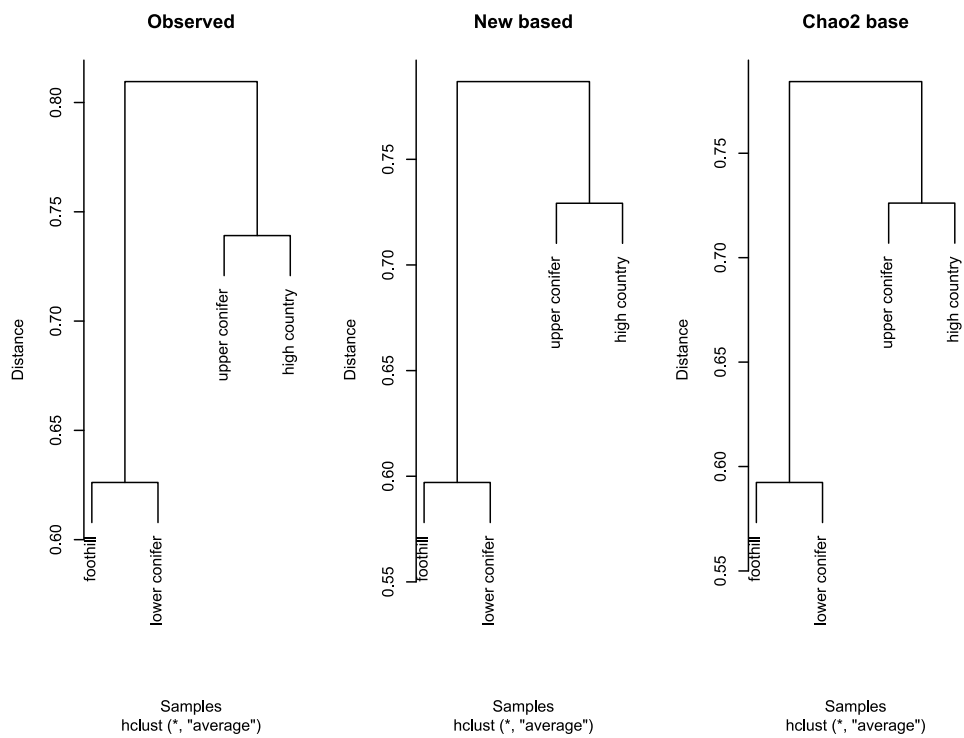


Figure 5.4 假設抽樣比例為 0.7 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

第6章 結論與後續研究

本文使用動差法，針對取後放回與取後不放回兩種不同的抽樣方式，分別提出 *New*、*wNew1* 以及 *wNew2* 三種估計式，用以修正兩群落間共同物種豐富度的估計。並依據電腦模擬，在不同模型假設以及物種分配的模擬群落，以及使用真實數據做為模擬群落的情況下，重複進行 1000 次的隨機抽樣。針對本文所提出的共同物種數估計方法與現有的方法，取後放回的估計方式 *Pan* 與針對取後不放回的估計方式 *wChao2*，進行比較，評估新提出的估計式。最後以紅杉國家公園的苔蘚資料，作為分析依據，對該地區不同海拔之間群落中的共同物種數進行估計，並以此計算 Jaccard 指標作為兩群落之間 *Beta* 多樣性的量化指標。

透過前半段使用電腦模擬群落進行估計中，綜合各項指標的結果，在本文所提出之新的估計方法中，比起原有的估計方法，皆取得偏誤更小的結果，具有更好的估計表現。值得注意的是，在取後不放回的模擬結果中，於小樣本的情況下，當其中一個群落的變異數小於 0.4 時，容易在估計共同種時發生高估的現象，特別是在某群落為同質群落的情況下更容易發生此現象。因此，當該情況發生時，可能採用取後放回的估計方法 *New*，會比使用 *wNew1* 或是 *wNew2* 獲得更優秀的估計表現。除此之外，在標準差的估計方面，無論是在取後放回或是取後不放回的模擬估計中，所得到的結果顯示在新提出的估計方法所估計的結果皆比現有結果還要大，尤其在小樣本方面的情況更甚。

因此針對以上現象，提出兩個議題：首先，針對取後不放回的小樣本中，共同種結果發生高估的情況，在本文模擬中，為其中一群落變異係數小於 0.4 時較為容易發生。而當該結果發生時，或許使用取後放回修正估計式 *New* 來進行估計，可能會得到較好的估計表現。然而，針對該現象所提出的評估標準「變異係數小於 0.4」是藉由多次電腦模擬所推算出的平均結果，相對而言較為

主觀。故針對該現象的發生，出了修正估計式以確保其穩定性之外，可能須另尋其他更明確且具有理論基礎的標準，來評估估計式的使用時機。此外，對於標準差估計的部分，由於所假設 Q_i 分佈為多項分佈，然而可能存在不同的分佈假設。是否能使用其他假設分佈，以更精準地估計標準差，也將成為未來在物種估計方面的一大挑戰。

最後，雖然共同物種數在作為評估兩群落間的物種相似性以及 Beta 多樣性的基礎之一，但在物種相似性以及 Beta 多樣性的量化指標上，往往是針對兩群落間的共同物種以及兩群落間物種數的比值作為標準。然而，並非在物種數估計準確的同時，多樣性的量化指標也會隨之準確。因此，若是想使用修正物種數去計算多樣性量化指標時，應注意修正後的估計結果合適的估計方式。

參考文獻

- Burnham, K. P., & Overton, W. S. (1978). Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika*, 65, 625-633.
- Burnham, K. P., & Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, 60, 927-936.
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*, 265-270.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 783-791.
- Chao, A., & Chiu, C. H. (2016). Species richness: estimation and comparison. *Wiley StatsRef: statistics reference online*, 1, 26.
- Chao, A., & Lin, C. W. (2012). Nonparametric lower bounds for species richness and shared species richness under sampling without replacement. *Biometrics*, 68(3), 912-921.
- Chao, A., Hwang, W. H., Chen, Y. C., & Kuo, C. Y. (2000). Estimating the number of shared species in two communities. *Statistica sinica*, 227-246.
- Chiu, C. H. (2022). Incidence-data-based species richness estimation via a Beta-Binomial model. *Methods in Ecology and Evolution*, 13(11), 2546-2558.
- Chiu, C. H. (2023). A species richness estimator for sample-based incidence data sampled without replacement. *Methods in Ecology and Evolution*.
- Chiu, C. H., Wang, Y. T., Walther, B. A., & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. *Biometrics*, 70(3), 671–682.

- Condit, R., Pérez, R., Aguilar, S., Lao, S., Foster, R., & Hubbell, S. (2019). Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years.
URL <https://doi.org/10.15146/5xcp-0d46>, 2, 13.
- Connell, J., Hall, M. A., Nimmo, D. G., Watson, S. J., & Clarke, M. F. (2022). Fire, drought and flooding rains: The effect of climatic extremes on bird species' responses to time since fire. *Diversity and Distributions*, 28(3), 417-438.
- Dorazio, R. M., & Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2), 351-364.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3-4), 237-264.
- Good, I. J. (2000). Turing's anticipation of empirical Bayes in connection with the cryptanalysis of the naval Enigma. *Journal of Statistical Computation and Simulation*, 66(2), 101-111.
- Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., ... & Watt, A. D. (1998). Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, 391(6662), 72-76.
- Maestre, F. T., Quero, J. L., Gotelli, N. J., Escudero, A., Ochoa, V., Delgado Baquerizo, M., ... & Zaady, E. (2012). Plant species richness and ecosystem multifunctionality in global drylands. *Science*, 335(6065), 214-218.
- Pan, H. Y., Chao, A., & Foissner, W. (2009). A nonparametric lower bound for the number of species shared by multiple communities. *Journal of agricultural, biological, and environmental statistics*, 14, 452-468.
- Shen, T. J., & He, F. (2008). An incidence-based richness estimator for quadrats sampled without replacement. *Ecology*, 89(7), 2052-2060.

- Shmida, A., & Wilson, M. (1985). Biological determinants of species diversity. *Journal of Biogeography*, 12, 1-20. <https://doi.org/10.2307/2845026>.
- Wilson, P., & Coleman, L. A. (2022). Niches and guilds of bryophytes along a 3000-meter elevational gradient. *The Bryologist*, 125(1), 115-134.
- Wu, B., Guan, Z., & Zhao, H. (2006). Parametric and Nonparametric FDR Estimation Revisited. *Biometrics*, 62. <https://doi.org/10.1111/j.1541-0420.2006.00531.x>.

附錄