

國立臺灣大學生物資源暨農學院農藝系

碩士論文

Department or Graduate Institute of Agronomy

College of Bio-Resources and Agriculture

National Taiwan University

Master's Thesis

依據出現型資料下的共同物種數估計

Estimate the Richness of Shared Species Based
on Incidence Data

郭昱嫻

KUO, YU-HSIEN

指導教授：邱春火 博士

Advisor: Chiu, Chun-Huo Ph.D

中華民國 113 年 6 月

June, 2024

國立臺灣大學碩(博)士學位論文
口試委員會審定書

MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

依據出現型資料下的共同物種數估計

Estimate the Richness of Shared Species Based on
Incidence Data

本論文係 郭昱嫻 R110621201 在國立臺灣大學農藝學系完成之碩士學位論文，於民國 113 年 6 月 321 日承下列考試委員審查通過及口試及格，特此證明。

口試委員 Oral examination committee:

(指導教授 Advisor)

系(所、學位學程)主管 Director: _____

摘要

兩群落之間之共同物種在群落中所佔的比例，可以做為一種表示 *Beta* 多樣性最簡單且直觀的指標之一。在過去的許多文獻中，以針對不同抽樣方式、調查方法或是資料型態，提出許多針對群落之間共同物種數的估計方法。且在估計方法的建構中，其中一項常見的估計方式便是使用有母數估計。而在過去也有許多研究針對估計單群落中的物種數，在假設出現型資料中，物種在出現的區塊數占群落中之比例服從 Beta 二項分佈 (Beta-binomial distribution) 的情況下使用動差法 (moment method) 建立估計式，且都得到良好的估計效果。因此，本文基於群落中物種出現區塊比例為 Beta 二項分佈的假設，使用動差法分別針對取後放回與取後不放回兩種抽樣方式，建立三個估計式。並透過多次的電腦模擬，評估估計式的優劣與穩定性。在結果方面，相較於原有的共同種估計方法，本文所提出的方式獲得較小的偏誤，且在 RMSE 與 95% 信賴區間涵蓋率 (95% confidence interval coverage) 兩項評估指標上，也相較現有的估計方法具有更好的表現。最後，將其應用至實例資料中，比較本篇所提出的估計式與現有估計式的結果差異，在計算群落共同種時，本文所提出的估計結果會高於現有估計式。

關鍵字：物種數、共同物種、Beta 二項分佈、動差法、有母數估計

Abstract

The proportion of shared species between two communities can serve as one of the simplest and most intuitive indicators of Beta diversity. Numerous studies in the past have proposed various methods for estimating the number of shared species between communities, taking into account different sampling methods, survey techniques, and data types. One common estimation approach among these methods is parametric estimation. Previous research has also developed estimation formulas for single-community species counts using the method of moments, assuming that the proportion of species occurrences in the sampling blocks follows a Beta-binomial distribution. These methods have yielded good estimation results. Therefore, based on the assumption that the proportion of species occurrences in a community follows a Beta-binomial distribution, this study employs the method of moments to establish three estimation formulas for two sampling methods: sampling with replacement and sampling without replacement. The study evaluates the accuracy and stability of these estimation formulas through extensive computer simulations. The results show that compared to existing methods for estimating shared species, the proposed methods exhibit smaller biases and perform better regarding RMSE and 95% confidence interval coverage. Finally, the study applies these methods to real-world data, comparing the proposed estimation formulas with existing ones. The proposed estimations yield higher values for the number of shared species between communities compared to existing methods.

Keywords: Species richness, shared species, Beta-binomial distribution, moment method, parametric method.

目次

摘要.....	i
Abstract.....	ii
目次.....	iii
圖次.....	v
表次.....	vii
第 1 章 緒論.....	1
第 2 章 模型符號介紹與相關文獻回顧	5
2.1 符號定義.....	5
2.2 相關文獻回顧.....	6
第 3 章 使用動差法估計共同物種數	17
3.1 取後放回之抽樣方法的估計方式.....	17
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	29
4.1 估計式整理.....	29
4.2 模型模擬設定.....	33
4.3 表格中名詞定義.....	34
4.4 取後放回的模擬結果.....	35
4.5 取後不放回與取後放回的模擬結果比較.....	41
4.6 真實資料模擬.....	46
4.7 討論.....	52
第 5 章 實例分析.....	55
5.1 紅杉國家公園內苔蘚資料.....	55
第 6 章 結論與後續研究	65
參考文獻.....	67
附錄.....	71

圖次

Figure 2.1 單一群落的物種分配示意圖。	7
Figure 2.2 兩群落的物種分配示意圖。	10
Figure 4.1 取後放回的抽樣方式在情況一下， 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。	37
Figure 4.2 取後放回的抽樣方式在情況一下， 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。	38
Figure 4.3 取後放回的抽樣方式在情況一下， 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。	39
Figure 4.4 取後放回的抽樣方式在情況一下， 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。	40
Figure 4.5 取後不放回的抽樣方式在情況一下， 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。	44
Figure 4.6 取後不放回的抽樣方式在情況一下， 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。	45
Figure 4.7 取後不放回的抽樣方式在情況一下， 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。	45
Figure 4.8 取後不放回的抽樣方式在情況一下， 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。.....	45
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。.....	47
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。.....	48
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。.....	49
Figure 4.12 取後不放回 BCI 資料集中的兩群落估計結果。.....	51
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。	
.....	58
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。.....	62
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。.....	62
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。.....	63

表次

Table 4.1 取後放回的抽樣方式在情況一下，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。	36
Table 4.2 取後放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的估計結果。	38
Table 4.3 取後放回的抽樣方式在情況一下，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。	39
Table 4.4 取後放回的抽樣方式在情況一下，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。	40
Table 4.5 取後不放回的抽樣方式在情況一下，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。	42
Table 4.6 取後不放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的估計結果。	43
Table 4.7 取後不放回的抽樣方式在情況一下，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。	43
Table 4.8 取後不放回的抽樣方式在情況一下，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。	44
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。	47
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。	48
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。	49
Table 4.12 取後不放回 BCI 資料集中的兩群落估計結果。	51

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。.....	57
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。	58
Table 5.3 不同抽樣比例之單群落物種數估計 (mean \pm SD) 。	59
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。	60
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。	60
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。	61
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。	61
Table S. 1 取後放回的抽樣方式在情況二下，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。	71
Table S. 2 取後放回的抽樣方式在情況二下，兩群落皆為均勻模型之情況下的估 計結果。	72
Table S. 3 取後放回的抽樣方式在情況二下，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。	72
Table S. 4 取後放回的抽樣方式在情況二下，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。	73
Table S. 5 取後放回的抽樣方式在情況三下，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。	73
Table S. 6 取後放回的抽樣方式在情況三下，兩群落皆為均勻模型之情況下的估 計結果。	73
Table S. 7 取後放回的抽樣方式在情況三下，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。	74
Table S. 8 取後放回的抽樣方式在情況三下，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。	74
Table S. 9 取後放回的抽樣方式在情況四下，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。	74
Table S. 10 取後放回的抽樣方式在情況四下，兩群落皆為均勻模型之情況下的估計結果。	75
Table S. 11 取後放回的抽樣方式在情況四下，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。	75
Table S. 12 取後放回的抽樣方式在情況四下，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。	75
Table S. 13 取後不放回的抽樣方式在情況二下，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。	76
Table S. 14 取後不放回的抽樣方式在情況二下，兩群落皆為均勻模型之情況下的估計結果。	76
Table S. 15 取後不放回的抽樣方式在情況二下，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。	77
Table S. 16 取後放回的抽樣方式在情況二下，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。	77
Table S. 17 取後不放回的抽樣方式在情況三下，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。	78
Table S. 18 取後不放回的抽樣方式在情況三下，兩群落皆為均勻模型之情況下的估計結果。	78
Table S. 19 取後不放回的抽樣方式在情況三下，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。	79

Table S. 20 取後放回的抽樣方式在情況三下，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。	79
Table S. 21 取後不放回的抽樣方式在情況四下，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。	80
Table S. 22 取後不放回的抽樣方式在情況四下，兩群落皆為均勻模型之情況下的估計結果。	80
Table S. 23 取後不放回的抽樣方式在情況四下，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。	81
Table S. 24 取後放回的抽樣方式在情況四下，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。	81

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，物種數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，以及在多樣性的保護與管理上。此外，也有研究表明，物種數與生態系統的功能性呈正向關係 (Shmida et al., 1985; Maestre et al., 2012)。故在群落物種數的監測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，如何透過樣本資料準確地估計物種數便成為一大難題。

在過去文獻中針對物種數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：個體豐富度資料與物種出現與否的出現型資料。所謂個體豐富度資料指具備物種和其個體數的資料。這些資料通常是基於個體為抽樣單位的隨機樣本，例如在特定的地點進行個體抽樣一段時間，並同時對抽取個進行物種鑑定和計數。因此，個體豐富度資料通常包括物種個體數量，物種種類數、物種的相對豐度等資訊；而物種出現型資料是指只紀錄物種出現與否的資料型態。這些資料通常是以區塊、陷阱或是特定的一段時間作為抽樣單位的隨機樣本，例如在特定目標地區中，抽取其中部分的區塊；或是在不同時間點進行調查，同時針對抽取的區塊，只紀錄物種出現與否而非物種實際出現的次數或個體數。文獻中，此兩種類型資料皆可用來作為量化生物多樣性指標 (Chao & Chiu, 2016)。

除了樣本資料型態的不同之外，抽樣方式則可分為取後放回 (sampling with replacement) 以及取後不放回 (sampling without replacement) 兩種常見的抽樣方式。因此在物種數估計方法的推導上，需針對不同的數據型態、不同的抽樣方式，進行相對應的抽樣機率模型和母體物種組成模型假設。

在大多數先前研究所提出的物種數的估計式中，依據估計方法的不同，大

致可分為有母數方法 (parametric method) 以及無母數方法 (nonparametric method)。其中，在無母數方法所建立估計式的過程中，無需對物種組成做任何機率模型假設。例如在物種數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，有母數方法則需要假設物種相對組成來自一個特定的機率分佈。在有母數方法中，模型的有母數方法通常可以通過最大概似估計 (MLE) 或者是動差法 (method of moments) 等方式估計。並且大部分情況下，無母數方法因為其無須假設一個特定的分佈形式，因此會比有母數方法更廣為被引用。反觀，有母數方法通常比無母數方法更簡單，因為它們可以藉由假設一個特定的分佈簡化問題。

而在生態上的應用以及環境決策中，除了單一群落的物種數之外，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。在這其中，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數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，依據無母數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，Shen 與 He (2008) 根據有母數方法，假設物種出現的機率為 Beta 分佈 (Beta distribution) 之隨機變數進行共同種類數估計。因此，本文將針對不同取樣方式下，根據物種出現型資料，結合假設物種在群落中出現的機率為 Beta 二項分佈 (Beta-binomial distribution)，用以改善兩群落之間共同物種數估計。

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、出現型數據、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，利用電腦模擬出的群落數據，以及兩筆真實資料：澳洲三種極端氣候鳥類資料與 BCI 資料，作為群落母體 (Connell et al.,

2022; Condit et al., 2019)。再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 (Wilson & Coleman, 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，依抽樣單位分為兩種資料型態：其一為以個體為抽樣單位的豐富度數據 (abundance data)；其次為依照區塊為抽樣單位的出現型數據 (incidence data)。本文主要以區塊為抽樣單位的出現型數據來進行共同物種數估計。本文使用的符號如下表示：

母體群落

- S_j ：第 j 群落的物種數， $j = 1, 2$ 。
- S_{12} ：兩群落共同物種數。
- S ：混合群落中的總相異物種數， $S = S_1 + S_2 - S_{12}$ 。
- T_j ：第 j 群集的總區塊數量， $j = 1, 2$ 。
- π_{ij} ：物種 i 在第 j 群落區塊中所出現的比例， $i = 1, 2, \dots, S$ ， $j = 1, 2$ 。

樣本

- X_i ：第一群落樣本中第 i 物種出現的區塊數量。
- Y_i ：第二群落樣本中第 i 物種出現的區塊數量。
- $S_{obs(j)}$ ：第 j 群落樣本中出現的物種數， $j = 1, 2$ ， $S_{obs(j)} = \sum_{i=1}^j I(X_i > 0)$ 。
- 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_j ：第 j 個集的抽樣區塊數量， $j = 1, 2$ 。
- q ：抽樣比例， $q = \frac{t_j}{T_j}$ ， $j = 1, 2$ 。
- $Q_{k(j)}$ ：第 j 群落樣本中出現 k 個區塊的物種數，即物種的區塊計數。 $k = 1, 2, \dots, S_j$ ， $j = 1, 2$ ， $Q_{k(j)} = \sum_{i=1}^{S_j} I(X_i = k)$ 。
- Q_{kl} ：在第一群落樣本出現 k 個區塊，並在第二群落樣本出現 l 個區塊的物種數。 $k = 1, 2, \dots, S_1$ ， $l = 1, 2, \dots, S_2$ ， $Q_{kl} = \sum_{i=1}^S I(X_i = k, Y_i = l)$ 。

- Q_{k+} ：在第一群落樣本出現 k 個區塊，並在第二群落樣本出現至少一個區塊的物種數。 $k = 1, 2, \dots, S_1$ ， $Q_{k+} = \sum_{i=1}^{S_1} I(X_i = k, Y_i > 0)$ 。
- Q_{+l} ：在第一群落樣本出現至少一個區塊，並在第二群落樣本出現 l 個區塊的物種數。 $l = 1, 2, \dots, S_2$ ， $Q_{+l} = \sum_{i=1}^{S_2} I(X_i > 0, Y_i = l)$ 。

2.2 相關文獻回顧

2.2.1 出現型數據

文獻中，大多數的物種數估計研究，其理論依據皆假設抽樣單位為隨機且獨立，但在真實野外的情境下，由於物種會因不同的習性以及特性而聚集，導致在某些特定情況下，更容易觀測到相同物種。因此以個體為抽樣單位的方式進行調查較難符合該假設。而相較於個體抽樣，僅記錄物種在該區塊出現與否的區塊抽樣較能符合該理論假設。此時，區塊為抽樣單位，如陷阱、區塊、穿越線或固定時間的調查。在出現型數據的抽樣中，可先將研究區域劃分為多個面積大致相同的區塊，並從中隨機選擇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抽樣區塊中的物種出現與否。

出現型數據由 t_j 個抽樣區塊的樣本所組成，其中每個抽樣區塊只紀錄物種的出現或是未出現，以形成一個以物種為列區塊為行的矩陣 $[W_{ij}]$ 。其中，若是 i 物種出現在 j 區塊中，無論個體總數為何，皆計為 1 ($W_{ij} = 1$)；反之若尚未觀測到該物種在區塊中則計為 0 ($W_{ij} = 0$)。

以第一群落為例， X_i 為落樣本中第 i 物種出現的區塊數量 ($X_i = \sum_{j=1}^{t_1} W_{ij}$)，表示在該樣本中第 i 物種出現的區塊數量，故 $0 < X_i \leq t_1$ 。

因此，可得到第一樣本中剛好出現 k 個區塊的物種數， $Q_{k(1)} = \sum_{i=1}^{S_1} I(X_i =$

k), 且 $k = 1, 2, \dots, t_1$ 。其中, $Q_{1(1)}$ 為在該樣本中僅出現在一個區塊的物種數, $Q_{2(1)}$ 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, $Q_{0(1)}$ 將被定義為在該樣本並未被觀測到的物種數。而真實的物種數 S_1 , 應為被觀測到的物種數 $S_{obs(1)}$ 與未被觀測到的物種數 $Q_{0(1)}$ 之和 ($S_1 = S_{obs(1)} + Q_{0(1)}$), Figure 2.1)。

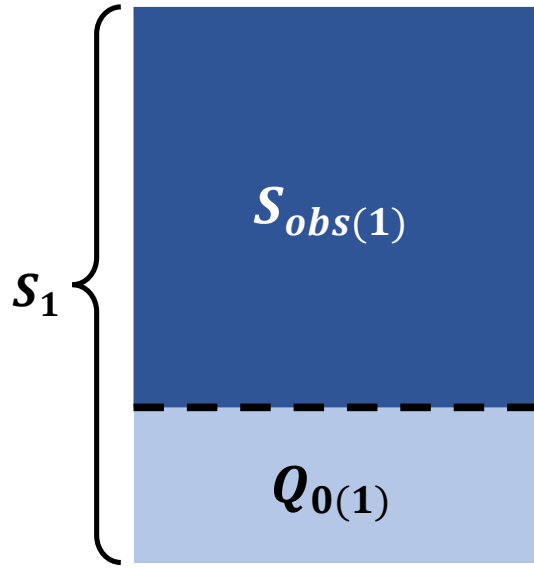


Figure 2.1 單一群落的物種分配示意圖。

2.2.2 取後放回之抽樣方式

在多數生態調查的研究中, 計算物種數在群落之中的占比, 是最直觀量化多樣性的指標之一。但普查物種數量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, 因此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。故需針對這些未被觀測到的物種進行估計, 以獲取到更接近於真實物種數的結果。

對於出現型數據所開發之物種數估計的模型多數皆是依據捉放法 (capture-recapture) 的理論架構為基礎所建立。而在物種數的估計中, 可將捉放法「個體」對應至「物種」, 以估計群落中的物種數。

在物種數的調查結果中，可依據物種在樣本中出現的區塊計數，將物種大致分為豐富物種與稀有物種。當樣本中皆為豐富物種時(出現頻率較高)，通常會被認為該群落之物種皆已被調查完全；相反，當樣本中的多數物種為稀有物種(出現頻率較低)時，在直覺上會認為群落中仍存在更多未被觀測到的物種，因此樣本中稀有物種對於未被觀測到的物種數提供較多的資訊。在過去許多研究中，最廣為應用估計方法是藉由稀有物種提供的資訊進行物種數估計。

依據 2.2.1 所述，並以第一群落為例，出現型矩陣 W_{ij} 可整理為 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 $\pi_{i(1)}$ ，而 $W_{ij} = 0$ 時機率為 $1 - \pi_{i(1)}$ 。則發生率矩陣的機率分佈為：

$$P(W_{ij} = w_{ij} | \pi_{i(1)}) = \pi_{i(1)}^{w_{ij}} (1 - \pi_{i(1)})^{1-w_{ij}}, i = 1, 2, \dots, S_1; j = 1, 2$$

而 X_i 服從二項分佈 (Binomial distribution)，則其機率分佈為：

$$P(X_i = x_i | \pi_{i(1)}) = \binom{t_1}{x_i} \pi_{i(1)}^{x_i} (1 - \pi_{i(1)})^{t_1 - x_i}, i = 1, 2, \dots, S_1; j = 1, 2$$

單群落物種數估計

在所有估計方法中，最被廣為利用的估計方法是 Chao (1987) 針對出現型數據建立物種數的無母數估計模型 *Chao2*。無母數的物種數估計是一個基本且直接的觀念，所謂無母數估計意旨在該估計方法中，不對物種數或者物種出現機率的分布進行假設。以第一群落作為單群落的例子，根據樣本中物種出現區塊計數的機率分佈， $E(Q_{k(1)})$ 可以表示為：

$$E(Q_{k(1)}) = \sum_{i=1}^{S_1} P(X_i = k) = \sum_{i=1}^{S_1} \binom{t_1}{k} \pi_{i1}^k (1 - \pi_{i1})^{t_1 - k}$$

又可依據柯西-施瓦茨不等式 (Cauchy-Schwarz inequality)，利用稀有物種中所含的資訊以估計真實的物種數。最終求得 *Chao2* 估計式為：

$$\hat{S}_{Chao2} = \begin{cases} S_{obs(1)} + \frac{t_1-1}{t_1} \left(\frac{Q_{1(1)}^2}{2Q_{2(1)}} \right), & \text{if } Q_{2(1)} > 0 \\ S_{obs(1)} + \frac{t_1-1}{t_1} \left(\frac{Q_{1(1)}(Q_{1(1)}-1)}{2} \right), & \text{if } Q_{2(1)} = 0 \end{cases}$$

並針對 *Chao2* 估計式，Chiu (2022) 使用動差法 (moment method)，推導出新的有母數方法的估計式 *BB*：

$$\hat{S}_{BB} = S_{obs(1)} + \hat{Q}_{0(1).Chao2} \left(2 - \left(\frac{2Q_{2(1)}^2}{3Q_{1(1)}Q_{3(1)}} \right)^- \right)$$

在上述式子中， $\hat{Q}_{0(1).Chao2} = \begin{cases} \frac{t_1-1}{t_1} \left(\frac{Q_{1(1)}^2}{2Q_{2(1)}} \right), & \text{if } Q_{2(1)} > 0 \\ \frac{t_1-1}{t_1} \left(\frac{Q_{1(1)}(Q_{1(1)}-1)}{2} \right), & \text{if } Q_{2(1)} = 0 \end{cases}$ ；且 $(A)^-$ 皆表示為：

且在上述等式中 $(A)^-$ 皆表示為：若 $A \leq 0.5$ 時，則 $(A)^- = 0.5$ ；而若 $A > 0.5$ 時，則表示為 $(A)^- = A$ 。

兩群落的共同物種數估計

在生態統計中，群落之間共同物種在樣本中所佔的比例，可以表示兩群落間的物種多樣性，同時也能表現兩群落間的相似性 (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 (Figure 2.2)。

為此，當在比較兩群落之間的物種數時，並非僅考慮單一群落或是混合群落的物種數，也必須針對群落之間所存在的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，並加上於樣本中已被觀測到的共同物種數，作為修正的共同物種數所使用。

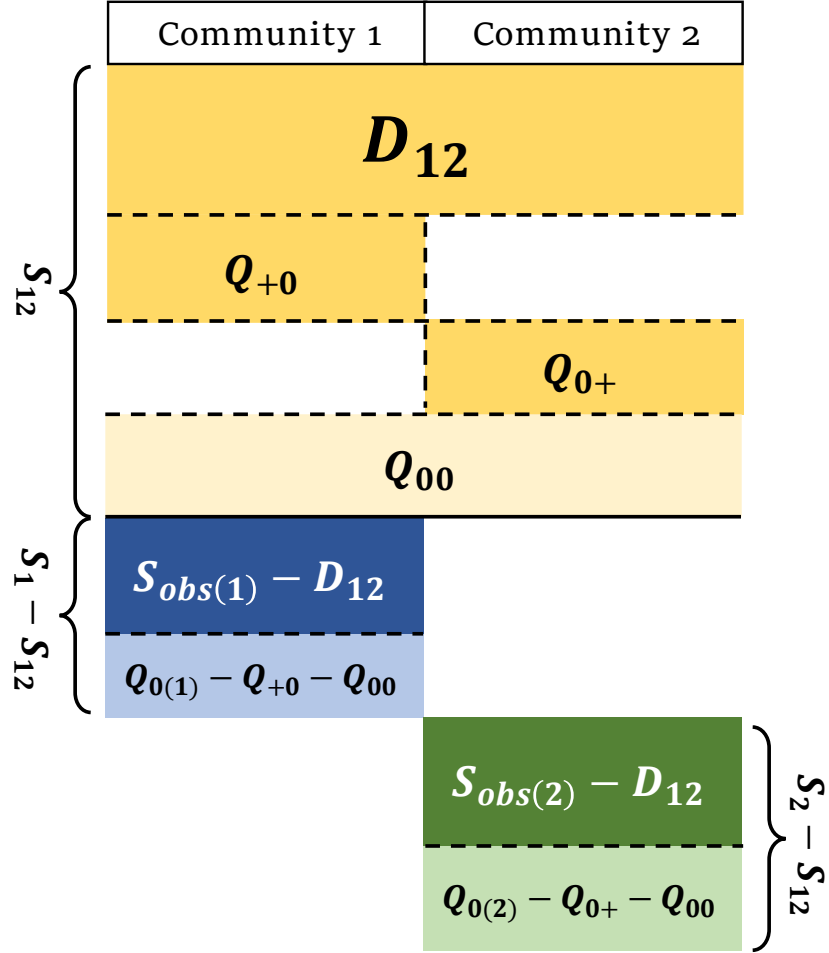


Figure 2.2 兩群落的物種分配示意圖。

Pan 等人(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。根據樣本中物種出現區塊計數的機率分佈， $E(Q_{kl})$ 可以表示為：

$$E(Q_{kl}) = \sum_{i=1}^{S_{12}} P(X_i = k)P(Y_i = l) = \sum_{i=1}^{S_{12}} \binom{t_1}{k} \pi_{i1}^k (1 - \pi_{i1})^{t_1-k} \binom{t_2}{l} \pi_{i2}^l (1 - \pi_{i2})^{t_2-l}$$

同樣可根據柯西-施瓦茨不等式，最終可以推導出估計式 Pan ：

$$\hat{S}_{Pan} = D_{12} + E(Q_{0+}) + E(Q_{+0}) + E(Q_{00})$$

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

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

$$E(Q_{00}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{Q_{11}^2}{Q_{22}}, & \text{if } Q_{22} > 0 \\ \frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \frac{Q_{11}(Q_{11} - 1)}{4}, & \text{if } Q_{22} = 0 \end{cases}$$

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，該抽樣方法廣泛使用在林業調查中，例如依照所選區塊對樹木進行不重複取樣，或是使用陷阱或誘捕器的抽樣方式等需要殺死個體的抽樣方法中。

在這種類型抽樣方法的單群落情況下，以第一群落的為例，假設將欲調查地區大致分為 T_1 個相等的區塊，又每個區塊物種存在於該群落的比例為 π_{i1} ，而 $0 \leq \pi_{i1} \leq 1$ ，且為未知參數。並假設在 T_1 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 $1 \leq U_i \leq T_1$ 。則在 $U_i = u_i$ 給定 π_{i1} 的條件下遵循參數 T_1 和 π_{i1} 的零截尾二項分佈 (zero-truncated binomial distribution)：

$$P(U_i = u_i | \pi_{i1}) = \binom{T_1}{u_i} \frac{\pi_{i1}^{u_i} (1 - \pi_{i1})^{T_1 - u_i}}{1 - (1 - \pi_{i1})^{T_1}}, i = 1, 2, \dots, S_1$$

當從 T_1 個區塊中，針對群落進行取後不放回之隨機抽樣，抽取 t_1 的區塊數，若在取樣區塊中發現該物種出現與否的情況。最終將各物種出現區塊的總和整理整理成物種出現的區塊數量 X_i 。且在給定 $U_i = u_i$ 的情況下， X_i 應服從超幾何分佈 (hypergeometric distribution)：

$$P(X_i = x_i | u_i) = \frac{\binom{u_i}{x_i} \binom{T_1 - u_i}{t_1 - x_i}}{\binom{T_1}{t_1}}, i = 1, 2, \dots, S_1$$

單群落物種數估計

Chao and Lin (2012) 基於 *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。以第一群落作為單一群落的情況下，在該估計方法中 $E(Q_{k(1)})$ 可以表示為：

$$E(Q_{k(1)}) = \sum_{i=1}^{S_1} P(X_i = k) = \sum_{i=1}^{S_1} \frac{\binom{u_i}{k} \binom{T_1 - u_i}{t_1 - k}}{\binom{T_1}{t_1}}$$

隨後根據柯西-施瓦茨不等式的概念推導出估計式，建構出針對取後不放回的樣本資料新的估計方法 *wChao2* 為：

$$\hat{S}_{wChao2} = S_{obs} + \frac{E(Q_{1(1)})^2}{w_1 2E(Q_{2(1)}) + r_1 E(Q_{1(1)})}$$

其中， $w_1 = \frac{t_1}{t_1 - 1}$ ， $r_1 = \frac{q_1}{1 - q_1}$ 。

而在 Chiu (2023) 使用動差法，推導出新的有母數方法的估計式 *MoRE*：

$$\hat{S}_{MoRE} = \begin{cases} S_{obs(1)} + \frac{t_1 - 1}{t_1} \frac{(1 - w_1) Q_{1(1)}^2}{Q_{2(1)} + w_1 (Q_{1(1)} - Q_{2(1)})^+}, & \text{if } Q_2 > 0 \\ S_{obs(1)} + \frac{t_1 - 1}{t_1} \frac{1 - w_1}{w_1} (Q_{1(1)} - 1), & \text{if } Q_2 \leq 0 \end{cases}$$

在上述式子中 $w_1 = \frac{t_1}{t_1 - 1}$ ；此外 $(A)^+$ 表示為若 $A > 0$ 時，則等於 $(A)^+ = A$ ；若

$A \leq 0$ ，該式則表示為 $(A)^+ = 1$ 。

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。Chao 與 Lin (2012) 將估計由單群落的物種數估計延伸至兩群落共同物種估計，建構一個針對取後不放回的抽樣方式，估計兩群落共同種的估計方法。在給定 $U_{i1} = u_{i1}$ 與 $U_{i2} = u_{i2}$ 的情況下， $E(Q_{kl})$ 可以表示為：

$$E(Q_{kl}) = \sum_{i=1}^{S_{12}} P(X_i = k)P(Y_i = l) = \sum_{i=1}^{S_{12}} \frac{\binom{u_{i1}}{k} \binom{T_1 - u_{i1}}{t_1 - k}}{\binom{T_1}{t_1}} \frac{\binom{u_{i2}}{l} \binom{T_2 - u_{i2}}{t_2 - l}}{\binom{T_2}{t_2}}$$

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， $w_j = \frac{t_j}{t_j - 1}$ ， $r_j = \frac{q_j}{1 - q_j}$ ， $j = 1, 2$

2.2.4 標準差估計

根據 $(Q_{0(j)}, Q_{1(j)}, \dots, Q_{t(j)})$ 的漸近分布，其服從大小為 S_j 以及機率為 $\left(\frac{E(Q_{0(j)})}{S_j}, \frac{E(Q_{1(j)})}{S_j}, \dots, \frac{E(Q_{t(j)})}{S_j}\right)$ 的多項分布 (multinomial distribution)。所提出的物種數估計量的變異數估計量可以使用 multivariate delta-method 導出，表示為

$$\widehat{var}(\hat{S}_{12}) = \sum_{i=1}^{t_1} \sum_{j=1}^{t_2} \frac{\partial \hat{S}_{12}}{\partial Q_i} \frac{\partial \hat{S}_{12}}{\partial Q_j} \widehat{cov}(Q_i, Q_j)$$

$$\text{其中 } \widehat{cov}(Q_i, Q_j) = \begin{cases} Q_i \left(1 - \frac{Q_i}{\hat{S}_{12}}\right), & \text{if } i = j \\ -\frac{Q_i Q_j}{\hat{S}_{12}}, & \text{if } i \neq j \end{cases}$$

2.2.5 95%信賴區間

物種豐富的信賴區間通過假設 $\hat{S}_{12} - D_{12}$ 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數。故，物種數之 95%信賴區間為：

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

2.2.6 樣本涵蓋率

樣本涵蓋率 (sample coverage) 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，物種出現區塊數為 (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 $\hat{\pi}_{i1} = \frac{X_i}{\sum_{i=1}^{S_1} X_i}$ ，則樣本覆蓋率可表示為：

$$C = \sum_{i=1}^{S_1} \pi_{i1} I(X_i > 0) = 1 - \sum_{i=1}^{S_1} \pi_{i1} I(X_i = 0)$$

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 ($Q_{i(j)}$) 的資訊估計該指標，樣本涵蓋率的估計式為：

$$\hat{C} = 1 - \frac{Q_{1(1)}}{\sum_{i=1}^{S_1} i Q_{i(1)}}$$

2.2.7 Jaccard 指數

在量化兩群落之間的相似性時，其中一項常見的指標為 Jaccard 指數 (Jaccard index)。該指標為 Jaccard (1901) 提出，利用共同物種數在兩群落中所佔的比例，定義群落之間的相似程度。

$$J_{index} = \frac{S_{12}}{S} = \frac{S_{12}}{S_1 + S_2 - S_{12}}$$

Jaccard 指數範圍由 0 至 1，0 表示兩群落物種組成完全相異；反之，1 表示為 0 表示兩群落物種組成完全相同。

並可將其轉換為 Jaccard 距離 (Jaccard distance) 用來表示兩群落間的相異性：

$$J_d = 1 - \frac{S_{12}}{S} = 1 - \frac{S_{12}}{S_1 + S_2 - S_{12}}$$

Jaccard 距離為 0 至 1，0 表示兩群落物種組成完全相同；反之，1 表示為 0 表示兩群落物種組成完全相異。

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，以第一群落的樣本為例，假設在目標區域實際存在 S_1 種物種，其中， S_1 為一未知參數。且抽樣方法是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_1 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 遵循總數為 t_1 ，機率为 π_{i1} 的二項分佈 (binomial distribution)。在此， π_{i1} 除了取決於第 i 物種的族群規模外，也與其多種的生物特徵因素相關。

Chiu (2022) 使用 Beta 二項式模型，建立一個新的針對單群落物種數的有母數估計方法。假設 X_i 遵循二項分佈，且 π_{i1} ($i = 1, 2, \dots, S_1$) 為機率密度函數 $g(\pi_{i1})$ 服從 $Beta(\alpha, \beta)$ 的隨機變數。假設 p_k 為樣本中出現的區塊數正好為 k 的平均機率，最終獲得以下樣本之物種出現區塊數的機率分佈如下：

$$\begin{aligned} p_k &= P(X_i = k) = \int_0^1 P(X_i = k | \pi_{i1} = \pi) g(\pi_{i1}) d\pi \\ &= \int_0^1 P(X_i = k | \pi_{i1} = \pi) \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1 - \pi)^{\beta-1} d\pi \\ &= \int_0^1 \binom{t_1}{k} \pi^k (1 - \pi)^{t_1-k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1} (1 - \pi)^{\beta-1} d\pi \\ &= \binom{t_1}{k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + k)\Gamma(\beta + t_1 - k)}{\Gamma(\alpha + \beta + t_1)}, \quad k = 0, 1, 2, \dots, t_1 \end{aligned}$$

又令 $Q_{k(1)}$ 表示在樣本中 k 個區塊的物種數，而 $Q_{k(1)} = \sum_{i=1}^{S_1} I(X_i = k)$ 。並根據樣本中物種出現次數的機率分佈，可知 $E(Q_{k(1)})$ 可以表示為：

$$E(Q_{k(1)}) = S_1 \times P(X_i = k) = S_1 \binom{t_1}{k} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + k)\Gamma(\beta + t_1 - k)}{\Gamma(\alpha + \beta + t_1)}$$

依據上述式子，可獲得 $E(Q_{0(1)})$ 、 $E(Q_{1(1)})$ 以及 $E(Q_{2(1)})$ ：

$$E(Q_{0(1)}) = S_1 \times P(X_i = 0) = S_1 \binom{t_1}{0} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha) \Gamma(\beta + t_1)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta + t_1)}$$

$$E(Q_{1(1)}) = S_1 \times P(X_i = 1) = S_1 \binom{t_1}{1} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + 1) \Gamma(\beta + t_1 - 1)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta + t_1)}$$

$$E(Q_{2(1)}) = S_1 \times P(X_i = 2) = S_1 \binom{t_1}{2} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + 2) \Gamma(\beta + t_1 - 2)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta + t_1)}$$

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： $\frac{t_1 E(Q_{0(1)})}{E(Q_{1(1)})} \geq \frac{(t_1 - 1) E(Q_{1(1)})}{2 E(Q_{2(1)})} \geq \frac{(t_1 - 2) E(Q_{2(1)})}{3 E(Q_{3(1)})} \dots$ 。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數提供更多的估計資訊。

將上述概念推廣至兩群落。假設兩群落皆為隨機且取後放回的抽樣，則 X_i 與 Y_i 皆分別遵循二項分佈 $Binomial(T_1, \pi_{i1})$ 以及 $Binomial(T_2, \pi_{i2})$ 。並假設 π_{i1} ， $i = 1, 2, \dots, S_1$ ； π_{i2} ， $i = 1, 2, \dots, S_2$ ，且機率密度函數分別為服從 $Bata(\alpha_1, \beta_1)$ 與 $Bata(\alpha_2, \beta_2)$ 的 $g(\pi_{i1})$ 與 $g(\pi_{i2})$ 。將 p_{kl} 定義為樣本中出現的區塊數正好分別為 k 和 l 的平均機率。則：

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

$$= \binom{t_1}{k} \binom{t_2}{l} \frac{\Gamma(\alpha_1 + \beta_1) \Gamma(\alpha_1 + k) \Gamma(\beta_1 + t_1 - k)}{\Gamma(\alpha_1) \Gamma(\beta_1) \Gamma(\alpha_1 + \beta_1 + t_1)} \frac{\Gamma(\alpha_2 + \beta_2) \Gamma(\alpha_2 + l) \Gamma(\beta_2 + t_2 - l)}{\Gamma(\alpha_2) \Gamma(\beta_2) \Gamma(\alpha_2 + \beta_2 + t_2)}$$

其中 $k = 0, 1, 2, \dots, t_1$ 與 $l = 0, 1, 2, \dots, t_2$ 。令 $Q_{kl} = \sum_{i=1}^{S_{12}} I(X_i = k \text{ and } Y_i = l)$ 表示在第一群落的樣本中存在 k 個區塊，且在第二群落的樣本中存在 l 個區塊的物種數。且樣本中觀測到的共同物種數為 $D_{12} = \sum_{k=1}^{t_1} \sum_{l=1}^{t_2} Q_{kl}$ 。藉此，可獲得 $E(Q_{0+})$ 、 $E(Q_{1+})$ 以及 $E(Q_{2+})$ ：

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

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

$$\begin{aligned}
E(Q_{2+}) &= S_{12} \times P(X_i = 2) \times (1 - P(Y_i = 0)) \\
&= S_{12} \times G \binom{t}{2} \left[\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} \frac{\Gamma(\alpha_1 + 2)\Gamma(\beta_1 + t_1 - 2)}{\Gamma(\alpha_1 + \beta_1 + t_1)} \right] \\
&\quad \times \left\{ 1 - \left[\frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} \frac{\Gamma(\alpha_2)\Gamma(\beta_2 + t_2)}{\Gamma(\alpha_2 + \beta_2 + t_2)} \right] \right\}
\end{aligned} \tag{3}$$

將 α 設定為 1，可藉由式 (1)、式 (2) 與式 (3) 成立以下近似值：

$$\frac{E(Q_{0+})}{E(Q_{1+})} = \frac{(\beta_1 + t_1 - 1)}{t_1 \alpha_1} \approx \frac{1}{\alpha_1} \left(\frac{\beta_1}{t_1} + 1 \right)$$

$$\frac{E(Q_{1+})}{E(Q_{2+})} = \frac{2}{t_1 - 1} \frac{\beta_1 - t_1 - 2}{\alpha_1 + 1} \approx \frac{2}{\alpha_1 + 1} \left(\frac{\beta_1}{t_1} + 1 \right)$$

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

並經由化簡 $\frac{E(Q_{0+})/E(Q_{1+})}{E(Q_{1+})/E(Q_{2+})}$ 可得：

$$E(Q_{0+}) = \frac{E(Q_{1+})^2}{2E(Q_{2+})} \left(1 + \frac{1}{\alpha_1} \right) \tag{4}$$

以及經由 $\frac{E(Q_{1+})/E(Q_{2+})}{E(Q_{2+})/E(Q_{3+})}$ 求得：

$$\alpha_1 \approx \frac{4E(Q_{2+})^2 - E(Q_{1+})E(Q_{3+})}{3E(Q_{1+})E(Q_{3+}) - 2E(Q_{2+})^2}$$

並由於 $\frac{E(Q_{1+})^2}{2E(Q_{2+})}$ 為適用於所有二項分佈的下界估計 (Pan et al., 2009; Chao 1987)，因此可依據 (4) 得知，在假設物種出現於群落的比例 (π_{ij}) 服從 Beta 二項分佈時，則估計量 $\frac{E(Q_{1+})^2}{2E(Q_{2+})}$ 的偏差近似於 $-\frac{E(Q_{1+})^2}{2E(Q_{2+})} \frac{1}{\alpha_1}$ 。

且由柯西-施瓦茨不等式可推得 $3E(Q_{1+})E(Q_{3+}) \geq 2E(Q_{2+})^2$ ，並結合式 (4) 可得知 $\frac{E(Q_{1+})^2}{2E(Q_{2+})}$ 的偏差絕對幅度的下界近似為不等式：

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

故將 $\left(1 - \frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})}\right)$ 帶入式 (4)，且為了確保估計是穩定的與 $\alpha > 0$ ，將滿足

$0.5 < \frac{2E(Q_{2+})^2}{3E(Q_{1+})E(Q_{3+})} < 1$ ，最終可獲得：

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

且在上述等式中 $(A)^-$ 皆表示為：若 $A \leq 0.5$ 時，則 $(A)^- = 0.5$ ；而若 $A > 0.5$ 時，則表示為 $(A)^- = A$ 。

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

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

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

最後加入 $\frac{t_j-1}{t_j}$ 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 $E(Q_{0+})$ 、 $E(Q_{+0})$ 以及 $E(Q_{00})$ 分別為：

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

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

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

且在上述等式中 $(A)^-$ 皆表示為：若 $A \leq 0.5$ 時，則 $(A)^- = 0.5$ ；而若 $A > 0.5$ 時，

則表示為 $(A)^- = A$ 。

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，並以第一群落樣本為例的情況下，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次數 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_{i1} ，因此可推導出：

$$\begin{aligned} P(X_i = k | \pi_{i1}) &= P(X_i = k | u_i) P(u_i | \pi_{i1}) \\ &= \binom{t_1}{k} \frac{\pi_{i1}^k (1 - \pi_{i1})^{t_1 - k}}{1 - (1 - \pi_{i1})^{T_1}} - \frac{(1 - \pi_{i1})^{T_1} I(k = 0)}{1 - (1 - \pi_{i1})^{T_1}} \end{aligned}$$

其中， $I(A)$ 為指標函數，表示若 A 發生時，則該式結果計為 1，反之則該式結果計為 0。

Shen 和 He (2008) 針對取後不放回的抽樣方式，開發有母數的估計方法，假設 X_i 遵循超幾何分佈，且 π_{i1} ($i = 1, 2, \dots, S_1$) 為服從 $Bata(\alpha, \beta)$ 的隨機變數，故機率密函數 $f(\pi_{i1})$ 可表示為：

$$f(\pi_{i1}) = G_1(\alpha, \beta, T_1) (1 - (1 - \pi_{i1})^{T_1}) \pi_{i1}^{\alpha-1} (1 - \pi_{i1})^{\beta-1}, 0 < \pi_{i1} < 1$$

其中 $\alpha > 0$ 與 $\beta > 0$ ，且 $G_1(\alpha, \beta, T_1) = \left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T_1)}{\Gamma(\alpha+\beta+T_1)} \right]^{-1}$ 。

並且 X_i 的機率分布可依據 $P(X_i = k | \pi_{i1})$ 與 $f(\pi_{i1})$ 獲得 (Shen & He, 2008; Chiu, 2023)。設 p_k 為樣本中出現的區塊數正好為 k 的平均機率，則：

$$\begin{aligned} p_k &= P(X_i = k) = \int_0^1 P(X_i = k | \pi_{i1}) f(\pi_{i1}) d\pi_{i1} \\ &= \begin{cases} G_1(\alpha, \beta, T_1) \left[\frac{\Gamma(\alpha)\Gamma(t+\beta)}{\Gamma(t+\alpha+\beta)} - \frac{\Gamma(\alpha)\Gamma(\beta+T_1)}{\Gamma(\alpha+\beta+T_1)} \right], & \text{if } k = 0 \\ G_1(\alpha, \beta, T_1) \binom{t_1}{k} \frac{\Gamma(\alpha+k)\Gamma(t_1+\beta-k)}{\Gamma(t_1+\alpha+\beta)}, & \text{if } k > 0 \end{cases} \end{aligned}$$

令 $Q_{k(1)}$ 表示在 k 個區塊中準確觀測到的物種數，而 $Q_{k(1)} = \sum_{i=1}^{S_1} I(X_i = k)$ 為在單群落樣本中出現 k 個區塊數。故根據 $P(X_i = k)$ ，可定義出 $E(Q_{k(1)})$ 為：

$$E(Q_{k(1)}) = S_1 \times P(X_i = k)$$

$$= \begin{cases} S_1 \times G_1(\alpha, \beta, T_1) \left[\frac{\Gamma(\alpha)\Gamma(t_1 + \beta)}{\Gamma(t_1 + \alpha + \beta)} - \frac{\Gamma(\alpha)\Gamma(\beta + T_1)}{\Gamma(\alpha + \beta + T_1)} \right], & \text{if } k = 0 \\ S_1 \times G_1(\alpha, \beta, T_1) \binom{t_1}{k} \frac{\Gamma(\alpha + k)\Gamma(t_1 + \beta - k)}{\Gamma(t_1 + \alpha + \beta)}, & \text{if } k > 0 \end{cases}$$

依據上述式子可知 $E(Q_{0(1)})$ 、 $E(Q_{1(1)})$ 以及 $E(Q_{2(1)})$ 分別為：

$$E(Q_{0(1)}) = S_1 \times P(X_i = 0) = S_1 \times G(\alpha, \beta, T_1) \left[\frac{\Gamma(\alpha)\Gamma(t_1 + \beta)}{\Gamma(t_1 + \alpha + \beta)} - \frac{\Gamma(\alpha)\Gamma(\beta + T_1)}{\Gamma(\alpha + \beta + T_1)} \right]$$

$$E(Q_{1(1)}) = S_1 \times P(X_i = 1) = S_1 \times G(\alpha, \beta, T_1) \binom{t_1}{1} \frac{\Gamma(\alpha + 1)\Gamma(t_1 + \beta - 1)}{\Gamma(t_1 + \alpha + \beta)}$$

$$E(Q_{2(1)}) = S_1 \times P(X_i = 2) = S_1 \times G(\alpha, \beta, T_1) \binom{t_1}{2} \frac{\Gamma(\alpha + 2)\Gamma(t_1 + \beta - 2)}{\Gamma(t_1 + \alpha + \beta)}$$

並基於 Good-Turing 頻率公式與柯西不等式之概念，針對單一群落樣本的估計得出近似式： $\frac{T_1 E(Q_{0(1)})}{E(Q_{1(1)})} \geq \frac{(T_1 - 1)E(Q_{1(1)})}{2E(Q_{2(1)})} \geq \frac{(T_1 - 2)E(Q_{2(1)})}{3E(Q_{3(1)})}$ 。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，有助於減少物種數估計結果的偏差。

隨後依據單群落估計式的概念，將估計推廣至兩群落。則 X_i 與 Y_i 皆分別遵循超幾何分佈，並假設 π_{i1} ， $i = 1, 2, \dots, S_1$ ； π_{i2} ， $i = 1, 2, \dots, S_2$ ，且機率密度函數分別為服從 $Bata(\alpha_1, \beta_1)$ 與 $Bata(\alpha_2, \beta_2)$ 的 $f(\pi_{i1})$ 與 $f(\pi_{i2})$ 。將 p_{kl} 定義為樣本中出現的區塊數正好分別為 k 和 l 的平均機率，則 p_{kl} ：

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

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

其中， $G_i = G_i(\alpha_i, \beta_i, T_i) = \left[\frac{\Gamma(\alpha_i)\Gamma(\beta_i)}{\Gamma(\alpha_i+\beta_i)} - \frac{\Gamma(\alpha_i)\Gamma(\beta_i+T_i)}{\Gamma(\alpha_i+\beta_i+T_i)} \right]^{-1}$ 。

並又令 $Q_{kl} = \sum_{i=1}^{S_{12}} I(X_i = k \& Y_i = l)$ 為在樣本中第一群落樣本出現 k 次且第二群落樣本出現 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 $E(Q_{0+})$ 、 $E(Q_{1+})$ 以及 $E(Q_{2+})$ ：

$$\begin{aligned} E(Q_{0+}) &= S_{12} \times P(X_i = 0) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \times \left[\frac{\Gamma(\alpha_1)\Gamma(t_1+\beta_1)}{\Gamma(t_1+\alpha_1+\beta_1)} - \frac{\Gamma(\alpha_1)\Gamma(\beta_1+T_1)}{\Gamma(\alpha_1+\beta_1+T_1)} \right] \quad (5) \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned}$$

$$\begin{aligned} E(Q_{1+}) &= S_{12} \times P(X_i = 1) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \binom{t_1}{1} \left(\frac{\Gamma(\alpha_1+1)\Gamma(t_1+\beta_1-1)}{\Gamma(t_1+\alpha+\beta_1)} \right) \quad (6) \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned}$$

$$\begin{aligned} E(Q_{2+}) &= S_{12} \times P(X_i = 2) \times (1 - P(Y_i = 0)) \\ &= S_{12} \times G(\alpha_1, \beta_1, T_1) \times \binom{t_1}{2} \left(\frac{\Gamma(\alpha_1+2)\Gamma(t_1+\beta_1-2)}{\Gamma(t_1+\alpha+\beta_1)} \right) \quad (7) \\ &\quad \times \left\{ 1 - G(\alpha_2, \beta_2, T_2) \left[\frac{\Gamma(\alpha_2)\Gamma(t_2+\beta_2)}{\Gamma(t_2+\alpha_2+\beta_2)} - \frac{\Gamma(\alpha_2)\Gamma(\beta_2+T_2)}{\Gamma(\alpha_2+\beta_2+T_2)} \right] \right\} \end{aligned}$$

依據式 (5) 與式 (6) 成立以下近似值。並將 α 設定為 1，且假設 $T \gg \beta > 0$ ：

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

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

由上述式子可得出：

$$\beta_1 = \max \left(0, \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})} \right) t_1 \right) \quad (8)$$

將並將其代入 $E(Q_{0+}) = \left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1}$ ，可得：

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))}, & \text{if } \beta_1 > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q}, & \text{if } \beta_1 = 0 \end{cases}$$

同理 $E(Q_{+0})$ 也依此證明，可得：

$$\beta_2 = \max \left(0, \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})} \right) t_2 \right) \quad (9)$$

並將該結果帶入可求得：

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + q(E(Q_{+1}) - E(Q_{+2}))}, & \text{if } \beta_2 > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q}, & \text{if } \beta_2 = 0 \end{cases}$$

同時，在兩群群落樣本同時為 $E(Q_{00})$ 、 $E(Q_{11})$ 以及 $E(Q_{22})$ 分別為：

$$\begin{aligned} E(Q_{00}) &= S_{12} \times P(X_i = 0) \times P(Y_i = 0) \\ &= S_{12} G(\alpha_1, \beta_1, T_1) G(\alpha_2, \beta_2, T_2) \\ &\times \left[\frac{\Gamma(\alpha_1) \Gamma(t_1 + \beta_1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} - \frac{\Gamma(\alpha_1) \Gamma(\beta_1 + T_1)}{\Gamma(\alpha_1 + \beta_1 + T_1)} \right] \\ &\times \left[\frac{\Gamma(\alpha_2) \Gamma(t_2 + \beta_2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} - \frac{\Gamma(\alpha_2) \Gamma(\beta_2 + T_2)}{\Gamma(\alpha_2 + \beta_2 + T_2)} \right] \end{aligned} \quad (10)$$

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

並依據式 (10) 與式 (11) 成立以下近似值：

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

將式 (8)、(9) 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 $E(Q_{0+})$ 、 $E(Q_{+0})$ 以及 $E(Q_{00})$ 分別為：

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))}, & \text{if } \beta_1 > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q}, & \text{if } \beta_1 = 0 \end{cases} \tag{12}$$

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + q(E(Q_{+1}) - E(Q_{+2}))}, & \text{if } \beta_2 > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q}, & \text{if } \beta_2 = 0 \end{cases} \tag{13}$$

$$E(Q_{00}) = \left(\frac{t_1 - 1}{t_1} \frac{t_2 - 1}{t_2} \right) \times \left[\left(\frac{\beta_1}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1} \right] \times \left[\left(\frac{\beta_2}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2} \right] E(Q_{11}) \tag{14}$$

在 S_{wNew1} 的基礎上，加入 Q_{12} 、 Q_{21} 的資訊對 β_1 、 β_2 的估計進行修正，依據式 (10) 與式 (11) 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：

$$\begin{cases} \beta_1 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_2}{\beta_2 + t_2}\right) t_1 - t_1 \\ \beta_2 = \frac{E(Q_{11})}{E(Q_{22})} \left(\frac{t_1}{\beta_1 + t_1}\right) t_2 - t_2 \end{cases} \quad (15)$$

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：

$$\begin{aligned} E(Q_{12}) &= S_{12} \times P(X_i = 1) \times P(Y_i = 2) \\ &= S_{12} \times G_1(\alpha_1, \beta_1, T_1) G_2(\alpha_2, \beta_2, T_2) \\ &\quad \times \binom{t_1}{1} \binom{t_2}{2} \left(\frac{\Gamma(\alpha_1 + 1) \Gamma(t_1 + \beta_1 - 1)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 2) \Gamma(t_2 + \beta_2 - 2)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{aligned} \quad (16)$$

$$\begin{aligned} E(Q_{21}) &= S_{12} \times P(X_i = 2) \times P(Y_i = 1) \\ &= S_{12} \times G_1(\alpha_1, \beta_1, T_1) G_2(\alpha_2, \beta_2, T_2) \\ &\quad \times \binom{t_1}{2} \binom{t_2}{1} \left(\frac{\Gamma(\alpha_1 + 2) \Gamma(t_1 + \beta_1 - 2)}{\Gamma(t_1 + \alpha_1 + \beta_1)} \right) \left(\frac{\Gamma(\alpha_2 + 1) \Gamma(t_2 + \beta_2 - 1)}{\Gamma(t_2 + \alpha_2 + \beta_2)} \right) \end{aligned} \quad (17)$$

並依式 (16) 與式 (17) 成立以下近似式：

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

並由上述式子推得：

$$\begin{cases} \beta_1 = \frac{E(Q_{12})}{E(Q_{21})} \left(\frac{t_1}{t_2}\right) (t_2 + \beta_2) - t_1 \\ \beta_2 = \frac{E(Q_{21})}{E(Q_{12})} \left(\frac{t_2}{t_1}\right) (t_1 + \beta_1) - t_2 \end{cases} \quad (18)$$

又可從式 (15) = 式 (18) 得：

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

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

並依公式解 $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ ，且 β_1 、 β_2 皆大於0，故得：

$$\begin{cases} \beta_1^* = -t_1 + t_1 \sqrt{\frac{E(Q_{11})E(Q_{12})}{E(Q_{22})E(Q_{21})}} = \max\left(0, \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{12})}{E(Q_{22})E(Q_{21})}}\right)t_1\right) \\ \beta_2^* = -t_2 + t_2 \sqrt{\frac{E(Q_{11})E(Q_{21})}{E(Q_{22})E(Q_{12})}} = \max\left(0, \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{21})}{E(Q_{22})E(Q_{12})}}\right)t_2\right) \end{cases} \quad (19)$$

最終得估計式為：

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

其中，將式(19)的 β_1^* 與 β_2^* 的結果分別帶入

$$E(Q_{0+}) = \left(\frac{\beta_1^*}{t_1} + 1\right) \frac{(T_1 - t_1)}{T_1 + \beta_1^*}$$

$$E(Q_{+0}) = \left(\frac{\beta_2^*}{t_2} + 1\right) \frac{(T_2 - t_2)}{T_2 + \beta_2^*}$$

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

中可分別求得 $E(Q_{0+})$ 、 $E(Q_{+0})$ 與 $E(Q_{00})$ 的估計式。

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：

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

且在上述等式中 $(A)^-$ 皆表示為：若 $A \leq 0.5$ 時，則 $(A)^- = 0.5$ ；而若 $A > 0.5$ 時，則表示為 $(A)^- = A$ 。

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 $\begin{cases} \beta_1 = \max\left(0, \left(\frac{E(Q_{1+}) - E(Q_{2+})}{E(Q_{2+})}\right) t_1\right) \\ \beta_2 = \max\left(0, \left(\frac{E(Q_{+1}) - E(Q_{+2})}{E(Q_{+2})}\right) t_2\right) \end{cases}$ 。

$$E(Q_{0+}) = \begin{cases} \frac{t_1 - 1}{t_1} \frac{(1 - q)E(Q_{1+})^2}{E(Q_{2+}) + q(E(Q_{1+}) - E(Q_{2+}))}, & \text{if } \beta_1 > 0 \\ \frac{t_1 - 1}{t_1} \frac{1 - q}{q}, & \text{if } \beta_1 = 0 \end{cases}$$

$$E(Q_{+0}) = \begin{cases} \frac{t_2 - 1}{t_2} \frac{(1 - q)E(Q_{+1})^2}{E(Q_{+2}) + q(E(Q_{+1}) - E(Q_{+2}))}, & \text{if } \beta_2 > 0 \\ \frac{t_2 - 1}{t_2} \frac{1 - q}{q}, & \text{if } \beta_2 = 0 \end{cases}$$

4. S_{wNew2}

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

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

$$E(Q_{0+}) = \left(\frac{\beta_1^*}{t_1} + 1 \right) \frac{(T_1 - t_1)}{T_1 + \beta_1^*} ; E(Q_{+0}) = \left(\frac{\beta_2^*}{t_2} + 1 \right) \frac{(T_2 - t_2)}{T_2 + \beta_2^*}$$

並將 β_1^* 與 β_2^* 以 $\begin{cases} \beta_1^* = \max\left(0, \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{12})}{E(Q_{22})E(Q_{21})}}\right) t_1\right) \\ \beta_2^* = \max\left(0, \left(-1 + \sqrt{\frac{E(Q_{11})E(Q_{21})}{E(Q_{22})E(Q_{12})}}\right) t_2\right) \end{cases}$ 帶入上述式子。

5. $S_{wChao2.12}$

$$S_{wChao2.12} = D_{12} + \frac{Q_{1+}^2}{2k_1 Q_{2+} + r Q_{1+}} + \frac{Q_{+1}^2}{2k_2 Q_{+2} + r Q_{+1}} + \frac{Q_{11}^2}{4k_1 k_2 Q_{22} + 2k_1 r_2 Q_{21} + 2k_2 r_1 Q_{12} + r_1 r_2 Q_{11}}$$

其中 $k_i = \frac{t_j}{t_j - 1}$ 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 $R = \left\{ 1.96 \left[\log \left(1 + \frac{\widehat{var}(\widehat{S}_{12})}{(\hat{S}_{12} - D_{12})^2} \right) \right]^{\frac{1}{2}} \right\}$ 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. 變異係數

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

$\hat{\gamma}^2$ 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。

情況一：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，且 $t_1 = t_2$ 。

情況二：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，且 $t_1 = t_2$ 。

情況三：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 $t_2 = 2t_1$ 。

情況四：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， $q = \frac{t_j}{T_j}$ 。
- 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進行計算，在多種組合模擬下，在樣本大小為 10 的情況下，個模型的樣本涵蓋率介於 0.6 至 0.84 之間；而當樣本大小大於 10 時，任何模型的樣本涵蓋率皆大於 0.9。

且依據模擬結果可以得知 (Table 4.1-Table 4.4)，在第一種物種與區塊數的假設，四種模擬模型下，偏差的表現在兩種估計方法中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。而在樣本標準差 (Sample SE) 以及標準差估計 (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。且 *New* 的樣本標準差與標準差估計結果皆比 *Pan* 大。

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，因此將以均方根誤差 (root-mean-square error, RMSE) 以及 95% CI Coverage (95% confidence interval Coverage, 95% CI Coverage) 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；但在大樣本的抽樣下，*Pan* 的表現會略優於 *New*。另一方面，95% CI Coverage 也是能作為評估估計式優劣的標準之一。在 95% CI Coverage 的結果中 *New* 的結果皆略優於 *Pan*。

在第二種與第三種物種與區塊數的假設下所進行的模擬結果 (Table S. 1-Table S. 8)，無論是在何種估計值的結果，皆與第一種物種與區塊數的假設並無太大的差異，*New* 的表現皆明顯優於 *Pan*。但是，在針對第四種物種的假設下 (Table S. 9 - Table S. 12)，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，容易有高估的情況發生；而在 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% CI Coverage 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 取後放回的抽樣方式在情況一下，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。

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

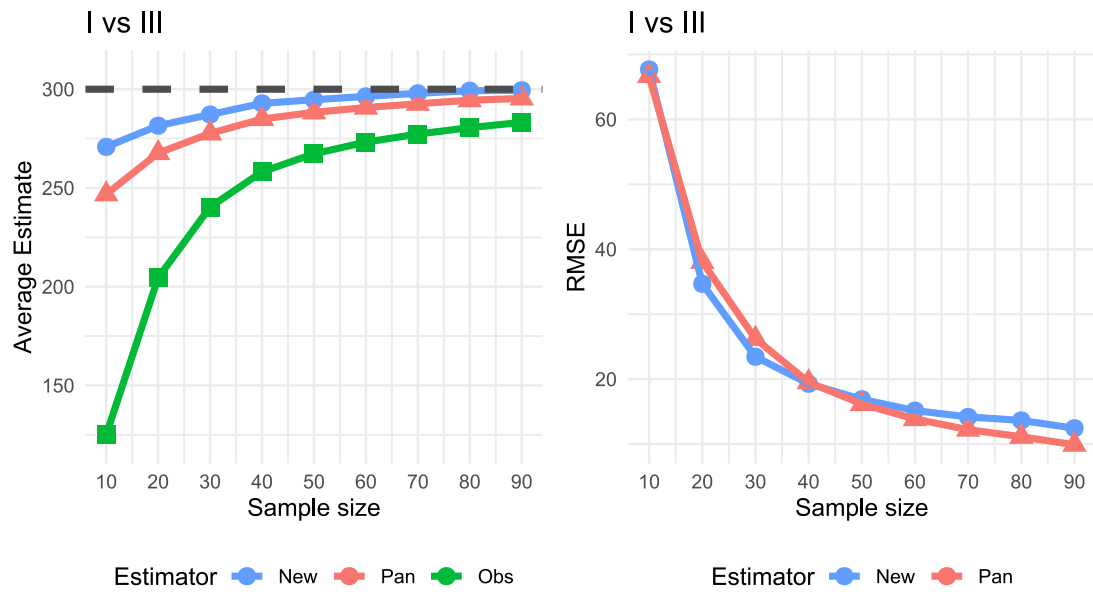


Figure 4.1 取後放回的抽樣方式在情況一下， 群落一為同質模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

Table 4.2 取後放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的估計結果。

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

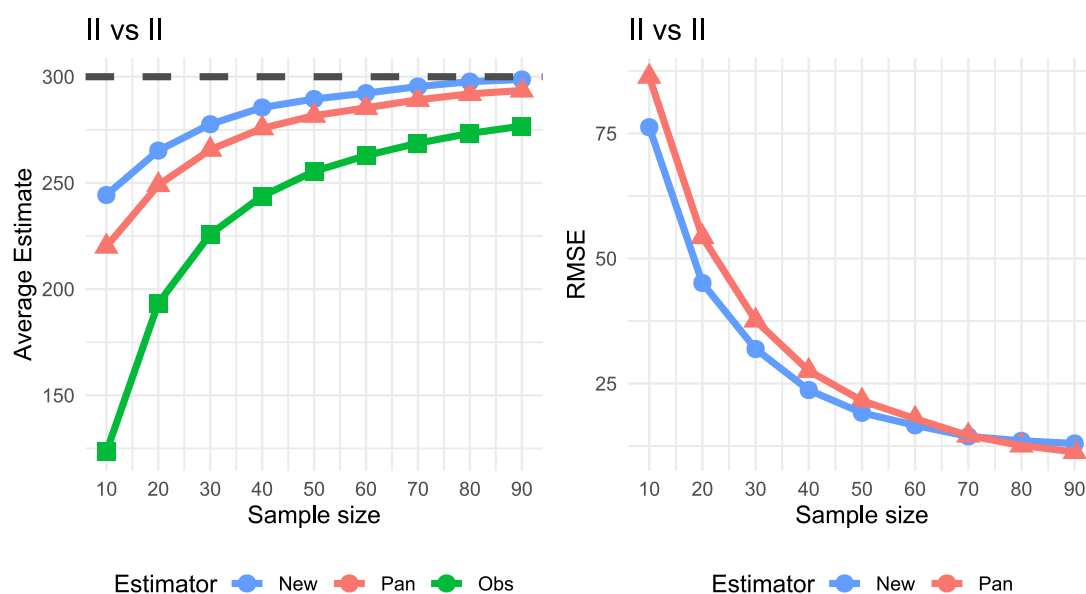


Figure 4.2 取後放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

Table 4.3 取後放回的抽樣方式在情況一下，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。

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

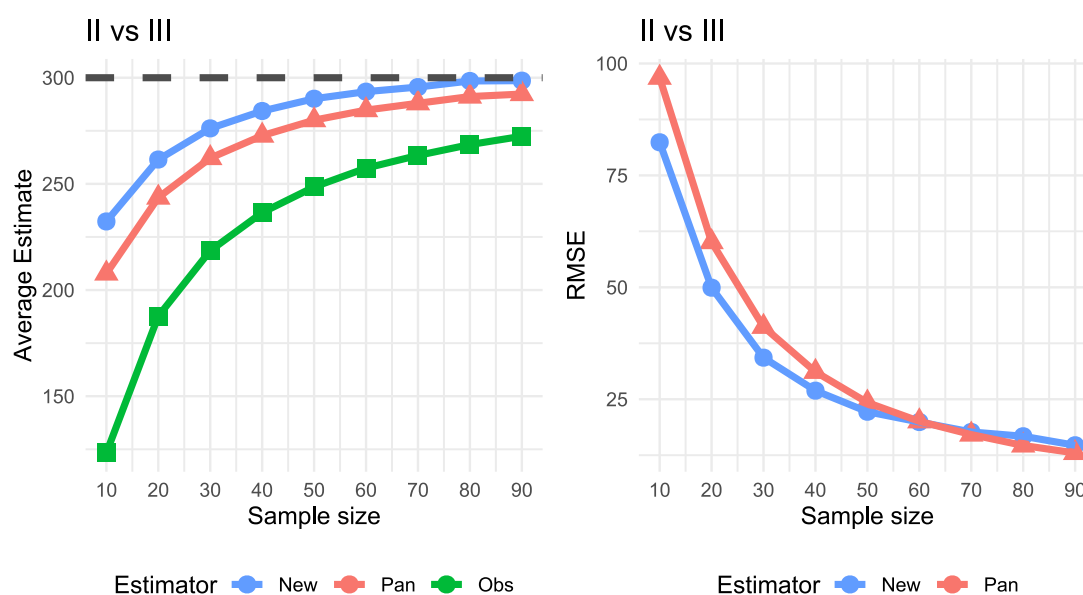


Figure 4.3 取後放回的抽樣方式在情況一下，群落一為均勻模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

Table 4.4 取後放回的抽樣方式在情況一下，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。

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

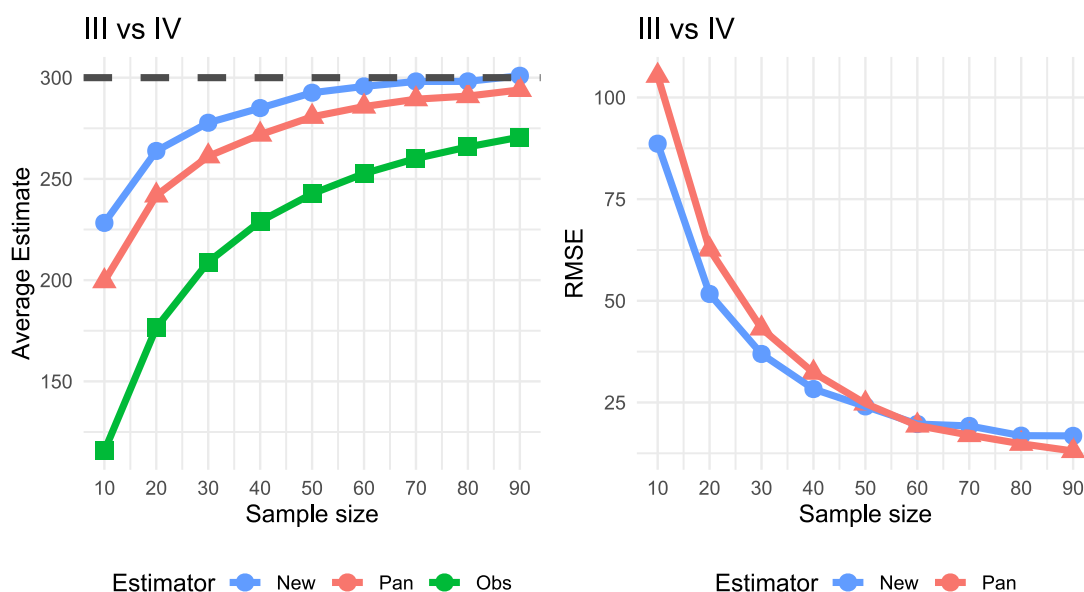


Figure 4.4 取後放回的抽樣方式在情況一下，群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

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 0.9。且模擬同質群落或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 (Table 4.5 - Table 4.8)，在偏差的表現上，在 *wNew1* 與 *wNew2* 小樣本的估計中，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則出現低估的表現。但在大樣本的情況下，所有群落組合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 *wChao2.12*。而在 95% CI Coverage 方面，由於 *wChao2.12* 已知為下界估計式，因此在 95% CI Coverage 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 *wChao2.12*。

在第二種與第三種物種與區塊數的假設底下，所進行的模擬結果，無論是在何種估計值的結果，皆與第一種物種與區塊數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有所差異，在取後不放回的模擬下，四種假設的結果皆無明顯差異 (Table S. 13 - Table S. 24)。

綜合以上論述可以得知，普遍而言在取樣比例小於 0.3 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，通常在其中一個群落的估計之樣本變異係數小於 0.4 時 ($\hat{y} < 0.4$)， $wNew1$ 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 0.3 時，取後不放回的估計結果才趨近穩定，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 0.3 時，以取後不放回的兩估計是對群落的共同種進行估計會獲得較好的結果。

Table 4.5 取後不放回的抽樣方式在情況一下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95%CI Coverage
0.1	wNew1	130.68	363.12	63.12	57.16	51.9	85.14	0.93
	wNew2		367.93	67.93	67.03	56.69	95.41	0.92
	wChao2.12		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	311.16	11.16	12.8	9.77	16.98	0.88
	wNew2		313.95	13.95	16.78	15	21.81	0.93
	wChao2.12		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	299.44	-0.56	6.72	7.5	6.74	0.98
	wNew2		297.72	-2.28	6.06	5.89	6.47	0.94
	wChao2.12		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
	wChao2.12		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 取後不放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的估計結果。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95%CI Coverage
0.1	wNew1	127.72	306.18	6.18	44.17	44.69	44.58	0.95
	wNew2		310.94	10.94	51.89	43.22	53.01	0.92
	wChao2.12		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	295.51	-4.49	14.16	14.3	14.85	0.95
	wNew2		302.6	2.6	19.56	17.19	19.72	0.93
	wChao2.12		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	297.26	-2.74	8.56	8.55	8.98	0.94
	wNew2		300.7	0.7	9.85	9.02	9.87	0.92
	wChao2.12		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	299.07	-0.93	5.17	4.97	5.25	0.94
	wNew2		299.6	-0.4	5.23	4.84	5.24	0.93
	wChao2.12		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 取後不放回的抽樣方式在情況一下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95%CI Coverage
0.1	wNew1	126.47	290.73	-9.27	43.62	38.75	44.58	0.92
	wNew2		299.38	-0.62	55.43	44.92	55.4	0.92
	wChao2.12		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	293.64	-6.36	15.81	14.89	17.03	0.94
	wNew2		297.21	-2.79	20.82	17.73	21	0.93
	wChao2.12		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	297.11	-2.89	8.34	9.2	8.82	0.97
	wNew2		297.61	-2.39	9	9.31	9.31	0.96
	wChao2.12		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	299.6	-0.4	5.32	5.33	5.34	0.94
	wNew2		299.04	-0.96	5.3	5.1	5.38	0.93
	wChao2.12		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 取後不放回的抽樣方式在情況一下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95%CI Coverage
0.1	wNew1	121.47	282.46	-17.54	44.72	41.29	48.02	0.94
	wNew2		290.68	-9.32	56.03	47.67	56.77	0.93
	wChao2.12		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	300.8	0.8	17.43	19.64	17.44	0.98
	wNew2		304.2	4.2	22.7	19.45	23.08	0.93
	wChao2.12		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	301.49	1.49	8.75	9.85	8.87	0.97
	wNew2		303.69	3.69	10.26	9.95	10.9	0.94
	wChao2.12		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	300.77	0.77	4.73	5.18	4.79	0.96
	wNew2		301.25	1.25	4.89	5.12	5.04	0.96
	wChao2.12		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

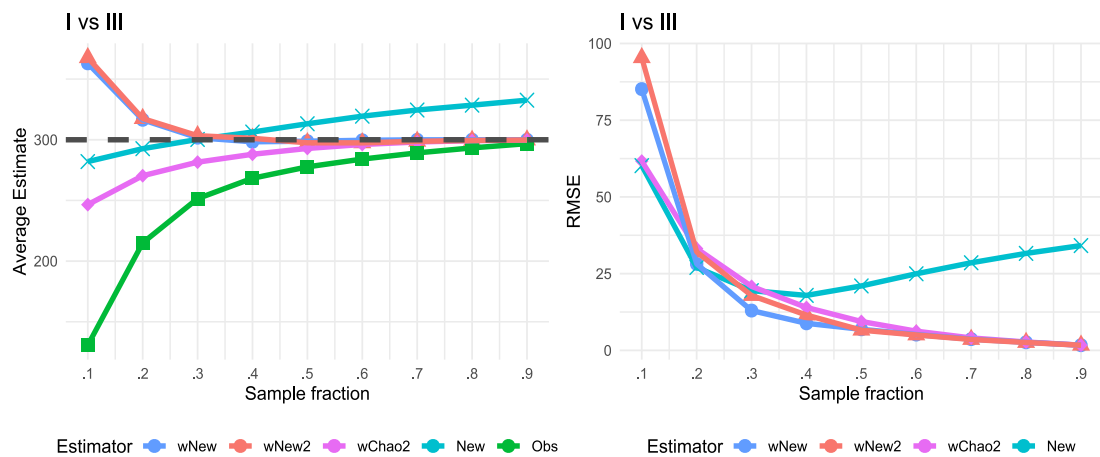


Figure 4.5 取後不放回的抽樣方式在情況一下，群落一為同質模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

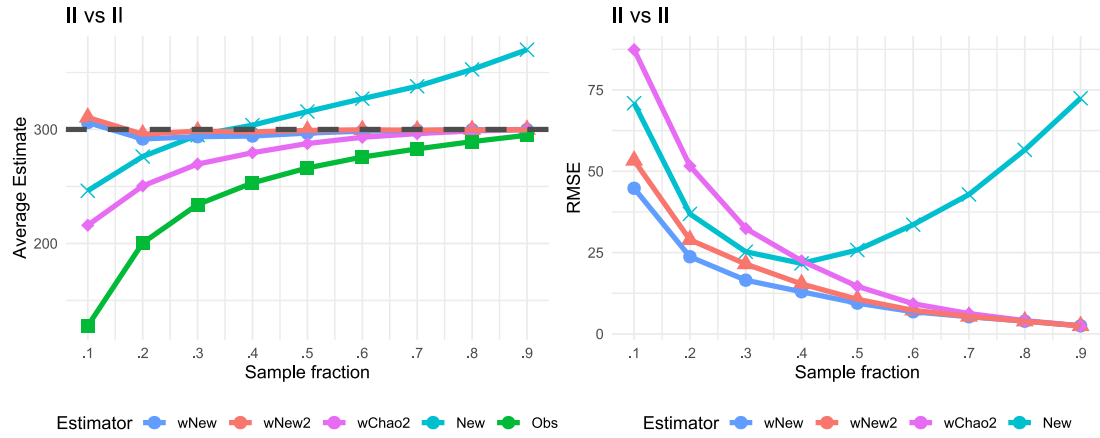


Figure 4.6 取後不放回的抽樣方式在情況一下，兩群落皆為均勻模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

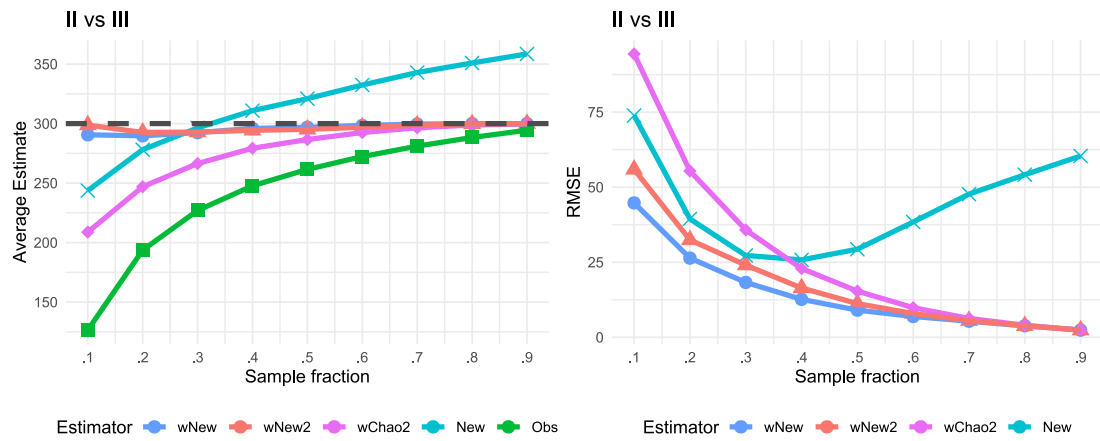


Figure 4.7 取後不放回的抽樣方式在情況一下，群落一為均勻模型，群落二為 Broken-stick 模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

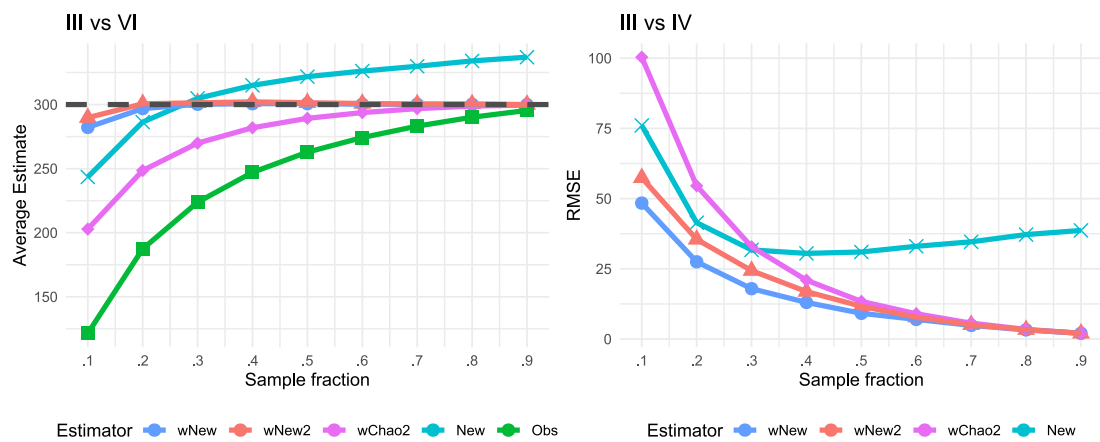


Figure 4.8 取後不放回的抽樣方式在情況一下，群落一為 Broken-stick 模型，群落二為對數常數模型之情況下的 AVG Estimate (左圖) 與 RMSE (右圖) 估計結果。

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) 和出現型 (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 = 0.08, CV = 1.32)，其中共存在 55 種物種；Big Wet 作為群落二使用 (Mean = 0.16, CV = 1.39)，其中共存在 65 種物種。Post-Big Wet 作為群落三使用 (Mean = 0.15, CV = 1.42)，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。在偏差結果方面，使用 *New* 的估計結果明顯表現較 *Pan* 佳。在估計式評估標準上，使用 RMSE 與 95% CI Coverage 進行評估兩種評估標準的結果表明，在多數情況下，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

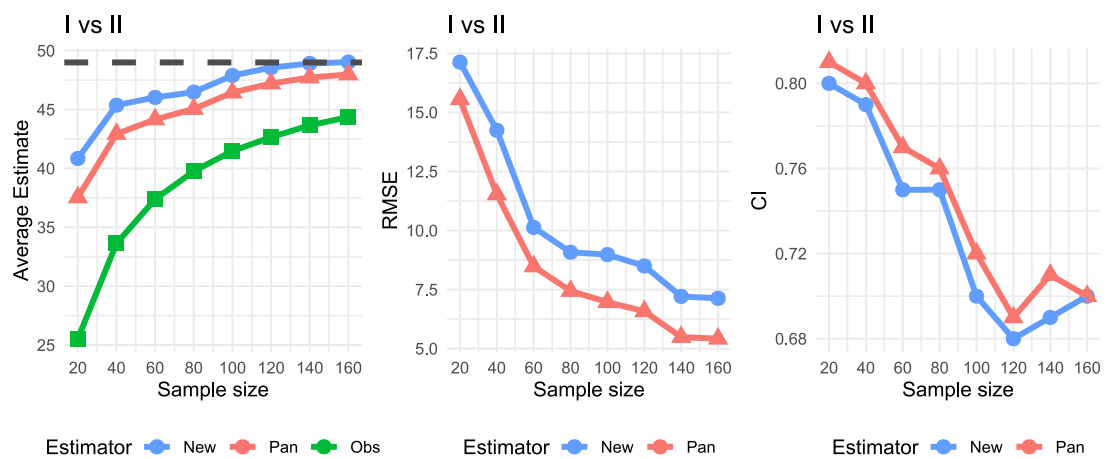


Figure 4.9 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Big Wet 估計結果。

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

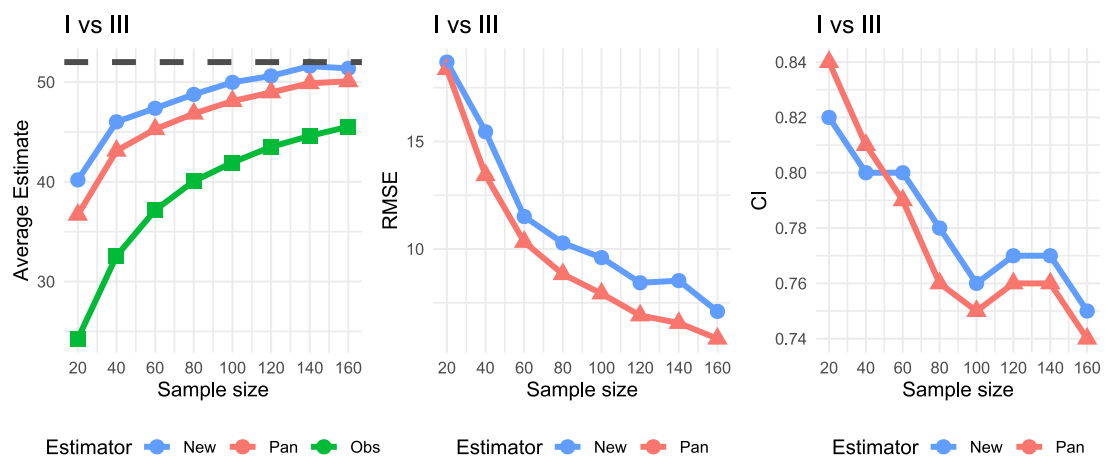


Figure 4.10 取後放回澳洲三種極端氣候鳥類資料集中的 Big Dry 與 Post-Big Wet 估計結果。

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

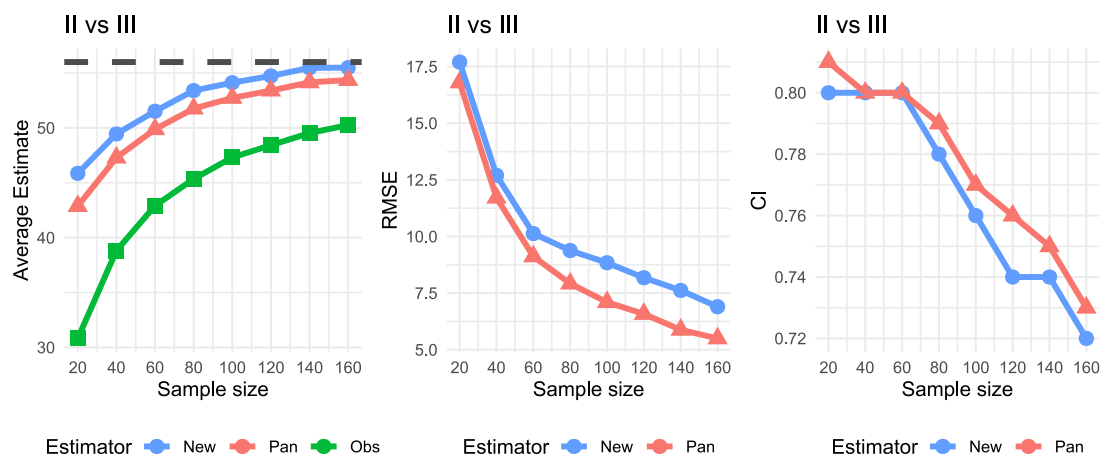


Figure 4.11 取後放回澳洲三種極端氣候鳥類資料集中的 Big Wet 與 Post-Big Wet 估計結果。

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，紀錄該地區樹種的生長狀況。將該地區分為 1250 個區塊進行調查，每塊區塊的大小為 400 平方公尺。以第一次調查 (1981 至 1983 年間) 與第八次調查 (2013 至 2016 年間) 的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 = 0.18, CV = 1.34)，第八次調查作為群落二 (Mean = 0.18, CV = 1.35)，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 *wNew1* 的估計方法在估計偏差方面，與真值差距最小。在抽樣比例為 0.1 時的估計標準差 (Est. SD)，與樣本標準差 (Sample SE) 存在較大的差異，而此差異會隨著抽樣比例增加差異逐漸減小。在估計式評估標準上，在兩種評估標準的結果表明，在小樣本中，*wNew1* 與 *wNew2* 的估計表現明顯較 *wChao2.12* 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		273.94	-10.06	23.46	26.09	25.51	0.97
	wNew2	225.24	298.91	14.91	46.77	34.39	49.07	0.92
	wChao2.12		253.31	-30.69	19.56	13.09	36.38	0.81
0.3	wNew1		283.48	-0.52	11.3	10.55	11.31	0.94
	wNew2	256.57	288.77	4.77	14.42	11.73	15.18	0.9
	wChao2.12		272.93	-11.07	9.05	7.26	14.3	0.84
0.5	wNew1		284.51	0.51	6.35	5.87	6.37	0.93
	wNew2	269.18	285.69	1.69	6.79	5.86	6.99	0.92
	wChao2.12		279.89	-4.11	5.77	4.84	7.08	0.84
0.7	wNew1		284.09	0.09	3.37	3.31	3.37	0.95
	wNew2	276.61	284.32	0.32	3.41	3.27	3.43	0.94
	wChao2.12		282.57	-1.43	3.3	3.01	3.59	0.87

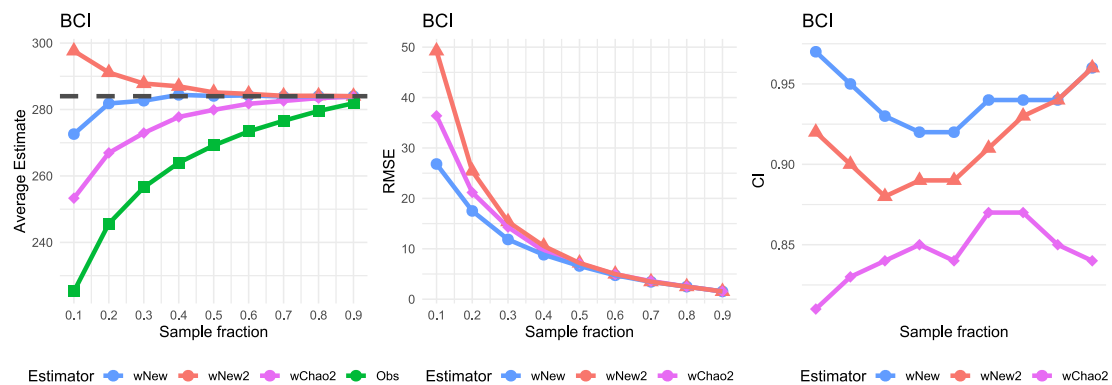


Figure 4.12 取後不放回 BCI 資料集中的兩群落估計結果。

4.7 討論

總結上述結果，在第 4 章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% CI Coverage 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% CI Coverage 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 *wChao2.12*。在 95% CI Coverage 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 *wChao2.12*

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*, *wChao2.12*)。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：

- I. 山麓 (foothill, F) 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，且降雨集中於 11 月至隔年 4 月。在雨季會伴隨濃霧，濃霧可延長苔蘚植物的水合時間 (hydrated)。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。
- II. 下針葉林帶 (lower conifer, LC) 海拔 1200 至 2440 公尺，地形相對平坦，在該地區樹林、濕草地與裸岩地交錯分布。在下針葉林帶，主要的降水量以降雨和降雪為主，在夏季時非常乾燥，但在冬天時的積雪會在夏季初期氣候正開始變乾燥時，滲透到富含有機質的土壤中。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。
- III. 上針葉林 (upper conifer, UC) 海拔 2440 至 2750 公尺，氣溫較低針葉林區更低、生長季更短，該地區的樹林被溪流分隔。在平坦區域雖出現濕草地，但在裸岩區域土壤不足以支撐樹木生長，而使得陽光充足。大多數枯木，具有乾枯且堅硬的表面，而非腐爛而鬆軟，史的苔蘚不易生長在枯木上。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。

IV. 高地地區 (high country, H) 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且零星，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。在多數估計中，共同物種數的估計值或是標準差的估計結果中，*New* 所估計之結果高於 *Pan* 所估計，此情況與上一章所呈現之電腦模擬的結果相符 (錯誤! 找不到參照來源。)。

並經由 Jaccard 距離 (J_d) 量化四個群落之間的 *Beta* 多樣性結果可以得知 (錯誤! 找不到參照來源。)，山麓地區與下針葉林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區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區具有更高的群落相似性。其可能原因和 Wilson 與 Coleman (2022) 所提及，由於上針葉林因被溪流切割而不連續，且裸岩與枯木等皆為乾燥通風的基質，更較不易苔蘚生長，而使得該地區的群落相異性較大。

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) (錯誤! 找不到參照來源。)。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的結果，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
F & LC	New	80	158.39	33.81	(114.89, 256.13)
	Pan		129.11	19.11	(103.53, 182.52)
F & UC	New	21	53.13	17.61	(32.77, 108.72)
	Pan		41.03	11.42	(28.08, 77.68)
F & H	New	24	101.43	41.27	(53.06, 230.33)
	Pan		73.19	27.2	(41.88, 159.37)
LC & UC	New	43	62.42	16.19	(47.68, 123.66)
	Pan		58.36	8.48	(48.59, 85.21)
LC & H	New	53	91.03	28.46	(63.29, 193.54)
	Pan		85.36	14.62	(66.91, 128.3)

UC & H	New	30	41.25	12.27	(31.99, 93.77)
	Pan		41.7	7.11	(33.9, 65.12)

Table 5.2 取後放回之紅杉國家公園內苔蘚的混合群落中總相異物種數與 J_d 估計結果。

	Estimator	Estimate	J_d
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

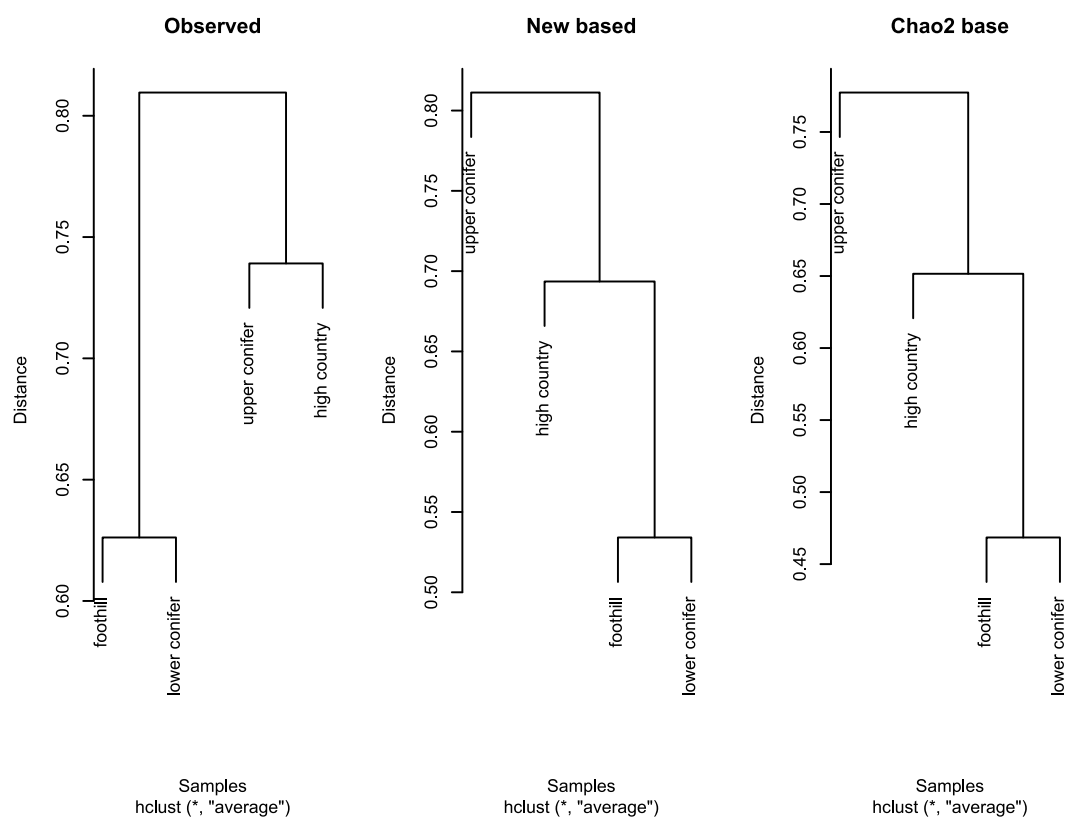


Figure 5.1 利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

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
F & LC	wNew1	80	135.45	19.2	(97.81 , 173.09)
	wChao2.12		116.69	12.11	(99.53 , 148.92)
F & UC	wNew1	21	50.41	16	(19.04 , 81.77)
	wChao2.12		46.94	12.97	(31.28 , 86.45)
F & H	wNew1	24	78.78	20.79	(38.03 , 119.53)
	wChao2.12		80.71	26.26	(47.91 , 158.54)
LC & UC	wNew1	43	66.27	9.61	(47.45 , 85.1)
	wChao2.12		56.64	6.61	(48.55 , 76.57)
LC & H	wNew1	53	90.06	13.4	(63.79 , 116.33)
	wChao2.12		75.78	8.92	(63.87 , 100.75)
UC & H	wNew1	30	46.27	7.3	(31.96 , 60.58)
	wChao2.12		39.19	5.19	(33.28 , 55.76)

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
F & LC	wNew1	80	111.47	9.14	(93.56 , 129.38)
	wChao2.12		103.86	7.52	(93.05 , 123.63)
F & UC	wNew1	21	36.07	6.3	(23.73 , 48.41)
	wChao2.12		34.25	5.72	(26.89 , 50.81)
F & H	wNew1	24	45.93	7.16	(31.9 , 59.96)
	wChao2.12		45.63	7.03	(35.62 , 64.25)
LC & UC	wNew1	43	57.89	5.68	(46.75 , 69.02)
	wChao2.12		53.18	4.77	(47.26 , 67.36)
LC & H	wNew1	53	75.25	7.49	(60.57 , 89.92)
	wChao2.12		69.22	6.07	(60.97 , 85.99)
UC & H	wNew1	30	40.57	4.63	(31.49 , 49.65)
	wChao2.12		37.12	3.91	(32.6 , 49.47)

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
F & LC	wNew1	80	95.76	4.91	(86.14 , 105.38)
	wChao2.12		93.58	4.58	(87.13 , 105.85)
F & UC	wNew1	21	28.39	3.35	(21.82 , 34.96)
	wChao2.12		27.87	3.27	(23.83 , 37.68)
F & H	wNew1	24	33.03	3.61	(25.95 , 40.11)
	wChao2.12		32.91	3.61	(28.16 , 43.13)
LC & UC	wNew1	43	51.61	3.59	(44.58 , 58.64)
	wChao2.12		49.85	3.31	(45.79 , 59.82)
LC & H	wNew1	53	64.6	4.23	(56.31 , 72.88)
	wChao2.12		62.73	3.87	(57.59 , 73.62)
UC & H	wNew1	30	36.14	3	(30.26 , 42.02)
	wChao2.12		34.87	2.77	(31.72 , 43.76)

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混合群落中總相異物種數與 J_d 估計結果。

	Estimator	Estimate			J_d		
		0.3	0.5	0.7	0.3	0.5	0.7
F & LC	Obs		214			0.63	
	MoRE	291.31	259.99	237.65	0.54	0.57	0.6
	wChao2	268.06	250.64	234.92	0.49	0.56	0.59
F & UC	Obs		164			0.87	
	MoRE	231.39	205.05	184.81	0.78	0.82	0.85
	wChao2	212.48	196.81	182.7	0.76	0.82	0.84
F & H	Obs		200			0.88	
	MoRE	274.34	245.6	223.28	0.71	0.81	0.85
	wChao2	250.35	235.22	220.71	0.69	0.8	0.85
LC & UC	Obs		172			0.75	
	MoRE	250.59	217.49	194.98	0.74	0.73	0.74
	wChao2	229.44	209.56	192.78	0.71	0.72	0.73
LC & H	Obs		201			0.74	
	MoRE	276.55	246.42	224.34	0.67	0.69	0.71
	wChao2	252.48	236.61	221.71	0.64	0.68	0.71
UC & H	Obs		115			0.74	
	MoRE	180.07	151.74	133.09	0.74	0.73	0.73
	wChao2	161.75	145.26	131.59	0.71	0.72	0.73

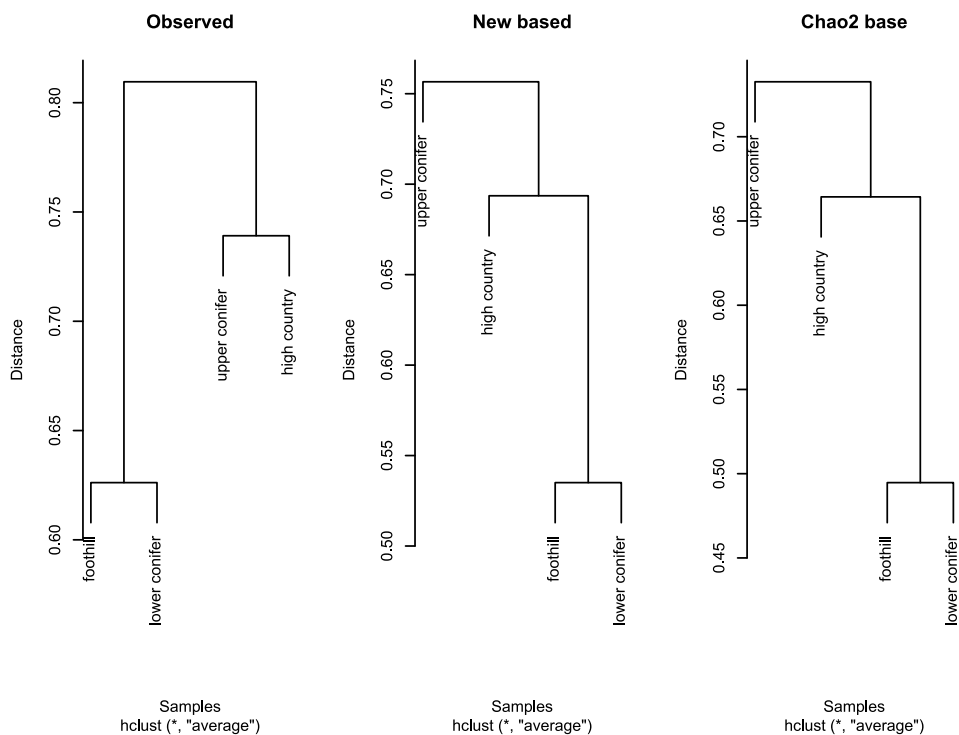


Figure 5.2 假設抽樣比例為 0.3 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

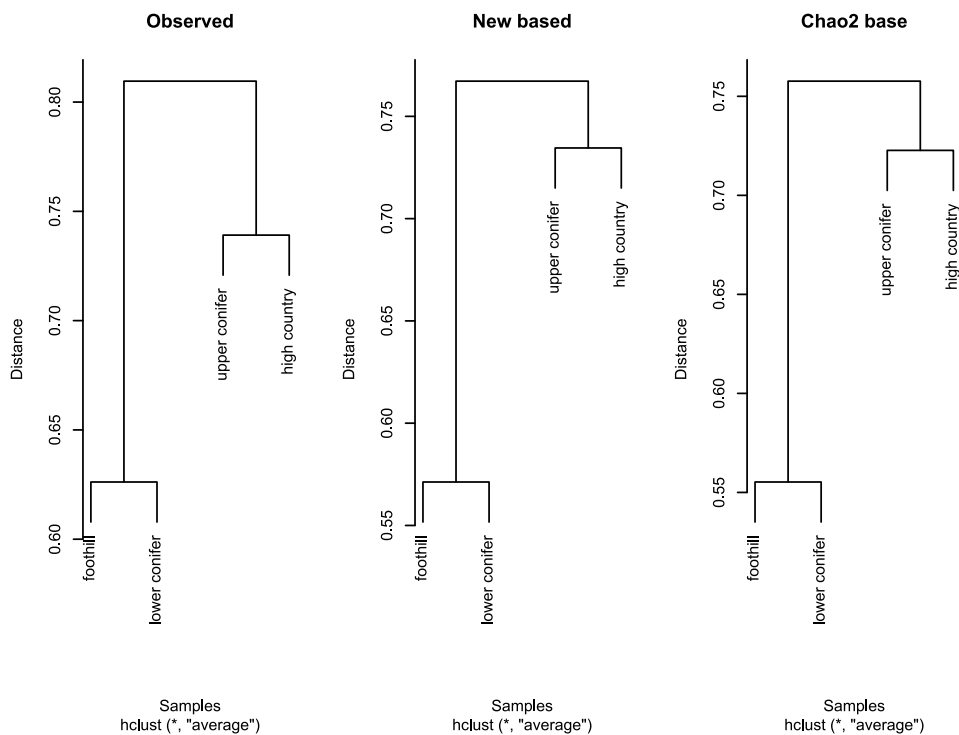


Figure 5.3 假設抽樣比例為 0.5 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

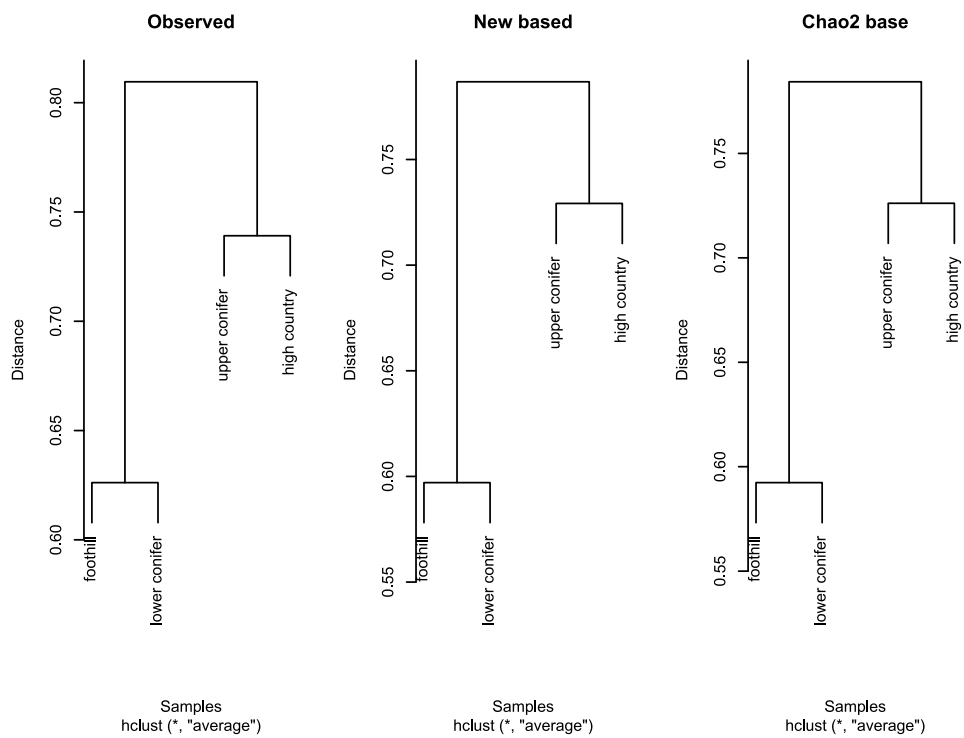


Figure 5.4 假設抽樣比例為 0.7 時，利用觀測物種數與分別使用兩種估計方法估計的物種數，繪製分類樹。

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 指標作為兩群落間 *Beta* 多樣性的量化指標。

透過第 4 章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，或是在針對估計式進行進一步的修正，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

針對的第 4 章電腦模擬結果，在不同抽樣方式，情況二至情況四的物種與區塊假設下之模擬。

物種與區塊數的假設

情況二：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，且 $t_1 = t_2$ 。

情況三：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 $t_2 = 2t_1$ 。

情況四：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。

Table S. 1 取後放回的抽樣方式在情況二下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	124.84	269.26	-30.74	61.37	46.09	68.61	0.86
	Pan		245.44	-54.56	40.48	27.76	67.92	0.82
30	New	238.61	284.7	-15.3	19.95	18.05	25.13	0.89
	Pan		275.22	-24.78	13.59	11.76	28.26	0.88
50	New	265.18	295.03	-4.97	17.2	15.4	17.9	0.87
	Pan		287.6	-12.4	11.82	10.15	17.13	0.87
70	New	275.28	299.83	-0.17	16.08	14.19	16.07	0.85
	Pan		293.29	-6.71	10.96	9.37	12.85	0.85

Table S. 2 取後放回的抽樣方式在情況二下，兩群落皆為均勻模型之情況下的估計結果。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	118.72	239.86	-60.14	54.85	40.74	81.37	0.87
	Pan		215.07	-84.93	33.16	22.92	91.17	0.83
30	New	222.32	276.8	-23.2	21.96	20.52	31.94	0.9
	Pan		264.23	-35.77	14.67	12.74	38.66	0.88
50	New	252.94	289.31	-10.69	17.5	16.05	20.5	0.88
	Pan		280.84	-19.16	12	10.54	22.61	0.88
70	New	267.23	295.18	-4.82	14.71	13.98	15.47	0.88
	Pan		288.57	-11.43	10.33	9.39	15.4	0.87

Table S. 3 取後放回的抽樣方式在情況二下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	119.37	229.87	-70.13	48.08	38.03	85.02	0.86
	Pan		204.97	-95.03	29.66	21.54	99.54	0.84
30	New	214.79	272.01	-27.99	24.65	21.57	37.29	0.88
	Pan		258.13	-41.87	16.59	13.56	45.03	0.87
50	New	244.86	289.81	-10.19	20.88	19.23	23.23	0.88
	Pan		278.69	-21.31	14.38	12.54	25.7	0.87
70	New	260.18	296.43	-3.57	18.26	17.32	18.59	0.89
	Pan		287.62	-12.38	12.68	11.42	17.72	0.88

Table S. 4 取後放回的抽樣方式在情況二下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	115.86	229.83	-70.17	50.19	40.94	86.26	0.87
	Pan		201.22	-98.78	30.98	22.83	103.52	0.85
30	New	209.58	275.15	-24.85	29.75	24.58	38.75	0.88
	Pan		259.48	-40.52	19.16	15.05	44.81	0.87
50	New	242.73	291.06	-8.94	21.86	20.18	23.61	0.88
	Pan		279.33	-20.67	14.79	12.93	25.41	0.89
70	New	259.41	297.41	-2.59	19.32	17.82	19.49	0.88
	Pan		288.54	-11.46	13.44	11.66	17.66	0.87

Table S. 5 取後放回的抽樣方式在情況三下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	151.89	291.68	-8.32	55.83	43.77	56.42	0.87
	Pan		270.85	-29.15	37.98	27.72	47.87	0.84
30	New	229.94	295.36	-4.64	25.16	21.78	25.57	0.9
	Pan		285.24	-14.76	17.63	14.62	22.99	0.89
50	New	262.44	300.04	0.04	18.16	16.1	18.15	0.89
	Pan		293.07	-6.93	12.65	10.84	14.42	0.88
70	New	276.46	300.61	0.61	14.48	12.96	14.49	0.86
	Pan		295.85	-4.15	10.32	8.84	11.12	0.85

Table S. 6 取後放回的抽樣方式在情況三下，兩群落皆為均勻模型之情況下的估計結果。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	154.86	253.01	-46.99	38.53	32.68	60.76	0.88
	Pan		233.32	-66.68	25.11	19.11	71.24	0.86
30	New	218.3	275.39	-24.61	22.88	21.29	33.59	0.92
	Pan		262.64	-37.36	15.29	13.13	40.36	0.9
50	New	244.6	285.93	-14.07	18.46	17.49	23.2	0.9
	Pan		276.47	-23.53	12.69	11.24	26.73	0.89
70	New	258.78	291.59	-8.41	15.85	15.3	17.94	0.9
	Pan		283.96	-16.04	11.14	10.1	19.53	0.88

Table S. 7 取後放回的抽樣方式在情況三下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	149.91	250.1	-49.9	39.71	34.31	63.76	0.89
	Pan		228.42	-71.58	25.27	19.96	75.9	0.87
30	New	210.79	275.31	-24.69	27.07	23.77	36.63	0.89
	Pan		260.27	-39.73	17.73	14.62	43.5	0.89
50	New	238.46	287	-13	21.37	19.78	25	0.9
	Pan		275.48	-24.52	14.5	12.64	28.48	0.88
70	New	253.86	291.79	-8.21	17.98	17.07	19.76	0.89
	Pan		283.04	-16.96	12.63	11.19	21.14	0.86

Table S. 8 取後放回的抽樣方式在情況三下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	145.97	243.51	-56.49	41.24	33.87	69.92	0.88
	Pan		219.83	-80.17	25.29	19.35	84.06	0.86
30	New	203.73	270.95	-29.05	28.84	24.61	40.92	0.88
	Pan		254.7	-45.3	18.87	15.08	49.07	0.87
50	New	231.27	282.68	-17.32	23.18	20.65	28.93	0.88
	Pan		270.41	-29.59	15.53	13.09	33.42	0.86
70	New	247.5	290.08	-9.92	19.83	18.4	22.16	0.89
	Pan		279.96	-20.04	13.62	11.94	24.23	0.88

Table S. 9 取後放回的抽樣方式在情況四下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	150.09	305.42	5.42	60.34	46.78	60.55	0.87
	Pan		283.4	-16.6	41.32	28.51	44.52	0.81
30	New	273.08	301.7	1.7	12.44	11.78	12.55	0.91
	Pan		297.95	-2.05	9.09	8.35	9.32	0.91
50	New	293.35	300.54	0.54	5.19	5.32	5.22	0.86
	Pan		299.45	-0.55	4.01	3.96	4.05	0.87
70	New	297.44	299.97	-0.03	3.08	3.08	3.08	0.78
	Pan		299.4	-0.6	2.41	2.27	2.48	0.77

Table S. 10 取後放回的抽樣方式在情況四下，兩群落皆為均勻模型之情況下的估計結果。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	189.35	292.07	-7.93	35.51	30.45	36.37	0.89
	Pan		278.38	-21.62	24.72	18.7	32.83	0.85
30	New	284.54	301	1	8.81	8.7	8.86	0.9
	Pan		298.27	-1.73	6.39	6.11	6.61	0.88
50	New	296.12	300.51	0.51	4.03	4.08	4.06	0.83
	Pan		299.73	-0.27	3.09	3.05	3.1	0.82
70	New	298.58	300.06	0.06	2.09	2.24	2.09	0.77
	Pan		299.73	-0.27	1.66	1.69	1.68	0.75

Table S. 11 取後放回的抽樣方式在情況四下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	179.09	280.11	-19.89	39.18	31.17	43.92	0.86
	Pan		262.25	-37.75	25.55	18.42	45.58	0.83
30	New	274.47	300.73	0.73	13.23	11.92	13.24	0.88
	Pan		295.92	-4.08	9.39	8.01	10.23	0.86
50	New	291.63	301.1	1.1	6.6	6.55	6.69	0.87
	Pan		299.63	-0.37	5.02	4.78	5.03	0.86
70	New	296.42	300.37	0.37	3.98	4.04	3.99	0.81
	Pan		299.7	-0.3	3.13	3.06	3.14	0.81

Table S. 12 取後放回的抽樣方式在情況四下，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。

Sample size	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
10	New	173.62	275.55	-24.45	42.65	33.45	49.14	0.87
	Pan		252.91	-47.09	26.06	18.94	53.82	0.85
30	New	267.04	299.99	-0.01	15.73	14.21	15.72	0.89
	Pan		294.02	-5.98	10.84	9.28	12.38	0.88
50	New	287.71	300.53	0.53	8.33	7.85	8.34	0.88
	Pan		298.39	-1.61	6.08	5.6	6.29	0.87
70	New	294.25	300.18	0.18	4.78	5.08	4.78	0.85
	Pan		299.23	-0.77	3.72	3.81	3.79	0.85

Table S. 13 取後不放回的抽樣方式在情況二下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	130.68	363.05	63.05	57.26	51.85	85.15	0.92
	wNew2		367.87	67.87	67.21	56.68	95.49	0.91
	wChao2		246.5	-53.5	30.8	24.32	61.73	0.87
	New		281.9	-18.1	57.45	45.58	60.21	0.89
0.3	wNew1	251.46	311.01	11.01	12.82	9.77	16.9	0.87
	wNew2		313.66	13.66	16.61	14.97	21.5	0.93
	wChao2		281.44	-18.56	9.7	9.14	20.94	0.91
	New		300.09	0.09	19.47	19.1	19.46	0.93
0.5	wNew1	277.73	299.59	-0.41	6.85	7.51	6.86	0.96
	wNew2		297.84	-2.16	6.14	5.89	6.5	0.94
	wChao2		293	-7	6.28	5.83	9.4	0.89
	New		313.52	13.52	16.8	16.91	21.56	0.91
0.7	wNew1	289.07	299.79	-0.21	3.77	4.05	3.77	0.97
	wNew2		298.57	-1.43	3.41	3.51	3.7	0.96
	wChao2		298.03	-1.97	3.72	3.73	4.2	0.9
	New		324.13	24.13	14.71	17.16	28.25	0.94

Table S. 14 取後不放回的抽樣方式在情況二下，兩群落皆為均勻模型之情況下的估計結果。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	127.75	307.14	7.14	45.84	44.82	46.37	0.96
	wNew2		310.97	10.97	51.61	43.1	52.74	0.93
	wChao2		216.34	-83.66	24.65	19.44	87.21	0.87
	New		246.24	-53.76	46.83	38.83	71.28	0.9
0.3	wNew1	234.05	295.24	-4.76	13.56	14.3	14.37	0.96
	wNew2		301.76	1.76	18.12	16.94	18.2	0.95
	wChao2		269.31	-30.69	10.75	9.88	32.52	0.91
	New		294.08	-5.92	22.85	21.09	23.6	0.9
0.5	wNew1	266.21	297.06	-2.94	8.33	8.51	8.83	0.95
	wNew2		300.6	0.6	9.74	8.98	9.75	0.94
	wChao2		287.46	-12.54	7.63	6.76	14.68	0.9
	New		314.96	14.96	20.43	19.06	25.31	0.91
0.7	wNew1	282.98	299.11	-0.89	5.14	4.97	5.21	0.95
	wNew2		299.58	-0.42	5.23	4.84	5.24	0.93
	wChao2		296.37	-3.63	5.05	4.55	6.22	0.89
	New		338.29	38.29	20.47	21.34	43.42	0.94

Table S. 15 取後不放回的抽樣方式在情況二下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	126.46	290.65	-9.35	43.42	38.73	44.39	0.92
	wNew2		299.29	-0.71	55.46	44.91	55.44	0.92
	wChao2		208.76	-91.24	24.23	19.12	94.4	0.87
	New		243.6	-56.4	47.84	38.78	73.94	0.88
0.3	wNew1	226.84	293.38	-6.62	16.14	14.9	17.43	0.94
	wNew2		296.97	-3.03	21.04	17.76	21.24	0.92
	wChao2		266.22	-33.78	12.58	10.78	36.04	0.88
	New		296.13	-3.87	27.57	24.19	27.82	0.89
0.5	wNew1	261.66	297.43	-2.57	8.65	9.23	9.02	0.96
	wNew2		298.04	-1.96	9.33	9.33	9.53	0.95
	wChao2		286.87	-13.13	7.84	7.46	15.29	0.9
	New		321.63	21.63	21	22.07	30.14	0.94
0.7	wNew1	281.01	299.5	-0.5	5.31	5.33	5.33	0.95
	wNew2		298.91	-1.09	5.21	5.09	5.32	0.94
	wChao2		296.41	-3.59	5.23	4.89	6.34	0.89
	New		342.5	42.5	20.79	23.2	47.31	0.95

Table S. 16 取後放回的抽樣方式在情況二下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	121.46	282.28	-17.72	44.81	41.19	48.17	0.94
	wNew2		290.34	-9.66	55.69	47.56	56.5	0.93
	wChao2		202.71	-97.29	25.03	19.93	100.45	0.86
	New		243.35	-56.65	50.83	41.75	76.09	0.87
0.3	wNew1	223.8	300.92	0.92	17.6	19.65	17.62	0.97
	wNew2		304.37	4.37	22.81	19.48	23.22	0.93
	wChao2		270.02	-29.98	13.43	11.79	32.85	0.9
	New		305	5	31.71	27.54	32.09	0.9
0.5	wNew1	262.84	301.43	1.43	8.93	9.83	9.04	0.97
	wNew2		303.64	3.64	10.47	9.93	11.08	0.94
	wChao2		289.22	-10.78	8.19	7.55	13.54	0.9
	New		321.8	21.8	22.39	22.04	31.24	0.92
0.7	wNew1	283.23	300.86	0.86	4.84	5.18	4.92	0.96
	wNew2		301.34	1.34	5	5.12	5.18	0.95
	wChao2		296.91	-3.09	4.79	4.59	5.7	0.91
	New		330.09	30.09	17.81	19.61	34.97	0.95

Table S. 17 取後不放回的抽樣方式在情況三下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	159.06	373.25	73.25	47.31	39.17	87.18	0.91
	wNew2		382.25	82.25	70.75	56.44	108.47	0.92
	wChao2		271.82	-28.18	27.94	23.77	39.67	0.9
	New		308.83	8.83	56.2	44.22	56.86	0.9
0.3	wNew1	243.7	330.38	30.38	13.99	12.84	33.45	0.94
	wNew2		334.38	34.38	18.56	17.4	39.06	0.94
	wChao2		289.62	-10.38	11.8	10.95	15.71	0.92
	New		312.97	12.97	23.34	21.64	26.69	0.92
0.5	wNew1	277.54	315.42	15.42	7	7.6	16.93	0.96
	wNew2		316.22	16.22	8	8.6	18.09	0.96
	wChao2		296.67	-3.33	6.05	5.99	6.9	0.91
	New		322.72	22.72	18.08	18.55	29.03	0.94
0.7	wNew1	291.93	304.46	4.46	3.47	4.16	5.65	0.98
	wNew2		304.6	4.6	3.6	4.29	5.84	0.98
	wChao2		299.35	-0.65	3.02	3.18	3.09	0.91
	New		329.92	29.92	14.28	18.41	33.14	0.96

Table S. 18 取後不放回的抽樣方式在情況三下，兩群落皆為均勻模型之情況下的估計結果。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	160.17	302.73	2.73	33.67	28.43	33.77	0.9
	wNew2		308.52	8.52	40.33	35.82	41.2	0.93
	wChao2		233.84	-66.16	20.05	16.93	69.13	0.89
	New		262.46	-37.54	37.84	32.68	53.29	0.9
0.3	wNew1	225.7	294.97	-5.03	16.21	14.21	16.97	0.93
	wNew2		301.87	1.87	22.21	19.5	22.28	0.93
	wChao2		264.79	-35.21	12.57	10.53	37.39	0.87
	New		289	-11	26.09	21.81	28.3	0.88
0.5	wNew1	254.95	296.39	-3.61	10.88	10	11.46	0.93
	wNew2		301.25	1.25	14.7	12.97	14.75	0.93
	wChao2		280.67	-19.33	9.33	7.9	21.46	0.87
	New		307.28	7.28	22.1	19.68	23.26	0.89
0.7	wNew1	271.15	297.59	-2.41	8.2	7.62	8.54	0.92
	wNew2		300.09	0.09	9.54	8.39	9.54	0.92
	wChao2		289.4	-10.6	7.38	6.26	12.91	0.87
	New		331.18	31.18	24.84	23.16	39.86	0.91

Table S. 19 取後不放回的抽樣方式在情況三下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	153.34	299.93	-0.07	36.54	28.54	36.53	0.89
	wNew2		306.95	6.95	48.58	41.09	49.05	0.93
	wChao2		230.42	-69.58	21.62	17.94	72.86	0.88
	New		266.17	-33.83	45.96	36.79	57.05	0.89
0.3	wNew1	219.68	295.52	-4.48	16.77	15.01	17.35	0.92
	wNew2		300.87	0.87	23.68	20.31	23.68	0.94
	wChao2		263.91	-36.09	12.81	11.37	38.29	0.89
	New		296.76	-3.24	28.32	25.95	28.49	0.9
0.5	wNew1	251.18	296.83	-3.17	10.52	10.46	10.98	0.95
	wNew2		299.4	-0.6	13.31	12.65	13.32	0.94
	wChao2		280.36	-19.64	8.98	8.37	21.6	0.91
	New		315.41	15.41	22.86	23.04	27.55	0.94
0.7	wNew1	269.77	297.88	-2.12	7.64	7.88	7.92	0.96
	wNew2		299	-1	8.42	8.35	8.48	0.96
	wChao2		289.27	-10.73	6.85	6.43	12.73	0.89
	New		331.2	31.2	20.99	22.91	37.59	0.95

Table S. 20 取後放回的抽樣方式在情況三下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	154.28	288.19	-11.81	33.85	30.26	35.84	0.92
	wNew2		296.9	-3.1	44.25	37.21	44.34	0.93
	wChao2		225.06	-74.94	20.74	16.82	77.76	0.88
	New		262.05	-37.95	42.71	35.21	57.12	0.88
0.3	wNew1	218.34	295.64	-4.36	18.74	17.87	19.23	0.94
	wNew2		300.33	0.33	25.28	21.32	25.27	0.94
	wChao2		263.33	-36.67	13.56	11.91	39.1	0.9
	New		295.2	-4.8	30.68	25.92	31.03	0.9
0.5	wNew1	248.75	297.2	-2.8	12.83	12.29	13.13	0.94
	wNew2		298.45	-1.55	15.46	13.78	15.53	0.94
	wChao2		279.09	-20.91	10.32	9.11	23.32	0.88
	New		305.35	5.35	23.07	21.46	23.67	0.9
0.7	wNew1	266.82	298.09	-1.91	8.5	9.04	8.71	0.96
	wNew2		298.04	-1.96	9.71	9.4	9.9	0.94
	wChao2		287.74	-12.26	7.36	7.17	14.3	0.91
	New		310.82	10.82	17.42	18.4	20.5	0.92

Table S. 21 取後不放回的抽樣方式在情況四下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	159.21	379.59	79.59	48.65	39.61	93.27	0.89
	wNew2		391.99	91.99	92.69	60.95	130.55	0.92
	wChao2		272.87	-27.13	28.1	24.33	39.05	0.9
	New		304.06	4.06	51.66	42.84	51.79	0.9
0.3	wNew1	271.94	314.85	14.85	10.24	8.33	18.03	0.89
	wNew2		317.52	17.52	12.44	11.8	21.48	0.94
	wChao2		291.25	-8.75	7.89	7.03	11.78	0.88
	New		300.94	0.94	14.23	13.25	14.25	0.9
0.5	wNew1	288.99	300.51	0.51	4.81	5.12	4.83	0.96
	wNew2		299.28	-0.72	4.15	3.99	4.21	0.93
	wChao2		296.71	-3.29	4.41	4.11	5.51	0.86
	New		305.96	5.96	10.79	10.9	12.32	0.88
0.7	wNew1	294.62	299.87	-0.13	2.75	2.84	2.75	0.95
	wNew2		299.3	-0.7	2.53	2.46	2.62	0.95
	wChao2		298.96	-1.04	2.71	2.59	2.9	0.85
	New		310.06	10.06	9.8	10.72	14.04	0.91

Table S. 22 取後不放回的抽樣方式在情況四下，兩群落皆為均勻模型之情況下的估計結果。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	157.26	306.13	6.13	36.13	29.39	36.63	0.88
	wNew2		309.38	9.38	44.84	37.29	45.78	0.92
	wChao2		231.82	-68.18	21.15	17.38	71.38	0.88
	New		256.95	-43.05	38.25	32.16	57.57	0.88
0.3	wNew1	250.49	298.19	-1.81	12.04	10.69	12.17	0.92
	wNew2		303.26	3.26	16.27	15.34	16.59	0.94
	wChao2		277.91	-22.09	9.81	8.74	24.16	0.9
	New		295.75	-4.25	19.08	17.75	19.53	0.9
0.5	wNew1	275.77	299.32	-0.68	7.08	6.74	7.11	0.95
	wNew2		299.32	-0.68	7.23	6.97	7.25	0.94
	wChao2		291.99	-8.01	6.51	5.92	10.32	0.89
	New		311.21	11.21	15.68	15.64	19.27	0.92
0.7	wNew1	288.22	299.73	-0.27	4.17	4.08	4.18	0.94
	wNew2		298.94	-1.06	3.96	3.81	4.1	0.94
	wChao2		297.66	-2.34	4.12	3.82	4.74	0.88
	New		323.15	23.15	14.83	16.28	27.49	0.93

Table S. 23 取後不放回的抽樣方式在情況四下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	156.58	300.09	0.09	35.8	27.9	35.78	0.89
	wNew2		310.17	10.17	52.07	42.92	53.03	0.94
	wChao2		229.88	-70.12	20.99	17.59	73.19	0.89
	New		256.18	-43.82	40.83	33.23	59.88	0.89
0.3	wNew1	246.82	295.82	-4.18	13.73	10.71	14.35	0.88
	wNew2		305.03	5.03	20.11	17.09	20.73	0.92
	wChao2		275.39	-24.61	11.32	8.99	27.09	0.86
	New		295.65	-4.35	23.25	19.35	23.65	0.86
0.5	wNew1	272.32	298.08	-1.92	7.63	7.05	7.86	0.93
	wNew2		298.96	-1.04	8.02	7.79	8.09	0.95
	wChao2		290.34	-9.66	6.92	6.29	11.88	0.89
	New		314.46	14.46	17.74	17.83	22.88	0.92
0.7	wNew1	286.38	299.73	-0.27	4.58	4.43	4.59	0.94
	wNew2		299.05	-0.95	4.42	4.23	4.52	0.94
	wChao2		297.47	-2.53	4.49	4.15	5.15	0.89
	New		330.72	30.72	17.57	19.77	35.39	0.94

Table S. 24 取後放回的抽樣方式在情況四下，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。

q	Estimator	Obs	AVG Estimate	Bias	Sample SE	Est. SD	RMSE	95% CI Coverage
0.1	wNew1	172.25	289.62	-10.38	28.85	25.08	30.65	0.92
	wNew2		297.81	-2.19	44.89	37.64	44.92	0.93
	wChao2		232.28	-67.72	17.57	15.58	69.96	0.9
	New		256.73	-43.27	32.13	29.1	53.89	0.9
0.3	wNew1	250.99	297.87	-2.13	13.14	12.23	13.31	0.93
	wNew2		301.84	1.84	17.17	16.48	17.26	0.95
	wChao2		279.12	-20.88	10.51	9.23	23.38	0.88
	New		298.74	-1.26	21.2	19.82	21.23	0.89
0.5	wNew1	274.96	298.83	-1.17	6.95	7.14	7.04	0.96
	wNew2		296.82	-3.18	6.29	6.43	7.04	0.96
	wChao2		291.84	-8.16	6.42	6.12	10.38	0.9
	New		313.39	13.39	17	17.48	21.63	0.92
0.7	wNew1	287.79	299.7	-0.3	3.97	4.2	3.98	0.95
	wNew2		298.47	-1.53	3.6	3.74	3.9	0.95
	wChao2		297.65	-2.35	3.94	3.91	4.58	0.9
	New		325.28	25.28	15.77	17.75	29.79	0.93

