**模型模擬設定**

**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：

1. 同質性模型 (homogeneity model)， ， ()
2. 均勻模型 (uniform model)， ， ()，其中服從均勻分佈，為調整常數。
3. Broken-stick 模型， ， ()，其中服從指數函數分佈，為調整常數。
4. 對數常態模型 (log-normal model)，， () ，其中服從對數常態分佈，為調整常數。

**物種與區塊數的假設**

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

1. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)。且兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用。
2. 假設總物種數皆為700種物種 (700)，其中群落一與第二群落分別存在400與600種物種 (400，600)，兩群落間共有300種共同種(300)。且兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用。
3. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)；兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用，而。
4. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)。且兩族群分別為100與200區塊 (100，200)，並從中依比例抽取與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 1000次。

**表格中名詞定義**

* q：為抽樣比例，。
* 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：次模擬之平均估計值。
* Bias：次模擬之偏差 (bias)，。
* Sample SE：次模擬之樣本標準差。
* Est. SD：次模擬之標準差估計值。
* RMSE：次模擬估計量之樣本均方根誤差。
* 95% CI Coverage：次模擬中理論值於信賴區間的比例。

**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，在第一種物種與區塊數的假設下，在偏差的表現上， 在四種模擬情況下，兩估計方法下所估計的群落共同種，皆有低估的情況發生。尤其在小樣本時低估情況更加明顯。平均而言，在兩種估計方法的偏差結果中皆顯示，*New*所估計之群落共同種相較於原有的*Pan*方法在偏差的估計表現上更為優秀。在樣本標準差 (Sample SE) 以及標準差估計 (Est. SD) 的結果中，可以發現無論在何種模擬情形下兩種方法皆有低估的情況發生。然而，單憑偏差與標準差的估計的結果無法從中判定估計式的好壞，在評估一個估計方法的優劣之時，還需要考慮該估計量的準確度以及精確度，因此將以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，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，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底下，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。

**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，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*。 而在95% CI Coverage方面，由於*wChao2*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*。

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。

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 (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。

**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1000次的重複抽樣，以評斷估計式的結果優劣。

**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。

**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。將該地區非為1250個區塊進行調查，每塊區塊的大小為400平方公尺。以第一次調查 (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*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。

**討論**

總結上述結果，在**錯誤! 找不到參照來源。**中使用了電腦模擬建立不同模型假設下的四種群落搭配四種不同的物種分配方式，以及分別使用兩筆真實資料作為群落母體。並從中隨機以取後放回以及取後不放回兩種不同的抽樣方式，重複抽取1000次，以計算平均的估計值、偏差、樣本標準差、族群標準差、RMSE以及95% 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*。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*表現更好。隨著樣本增加，三種估計方法的RMSE沒有明顯差異，但在95% CI Coverage方面，*wNew1*在小樣本中表現明顯優於其他兩者。

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

1. **實例分析**
   1. **紅杉國家公園內苔蘚資料**

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

1. 山麓 (F, foothill) 海拔1200公尺以下，該地區以地中海型氣候為主要特徵：夏季炎熱乾燥；冬季則為涼爽，通常都在冰點以上。降水形式以降雨為主。其中包含67個區塊以及132種物種 (Mean = 0.08，CV = 1.17)，且僅出現在單一區塊的物種具有48種。該地區的樣本覆蓋率為0.91。
2. 下針葉林帶 (LC, lower conifer) 海拔1200至2440公尺，在該地區主要的植被有巨型紅杉、雪松與白冷杉等。在下針葉林帶，主要的降水量以降雨和降雪為主。該地區氣候在夏季時非常乾燥，但在冬天時的積雪會滲透到土壤中變得相對涼爽。其中包含100個區塊以及162種物種 (Mean = 0.05，CV = 1.17) ，且僅出現在單一區塊的物種具有59種。該地區的樣本覆蓋率為0.93。
3. 上針葉林 (UC, upper conifer) 海拔2440至2750公尺，該地區氣溫較低，生長季較短。該地區被河流切割，與其他地區相比土壤相對較少，裸岩較多，且腐木會迅速乾枯，較不利於苔癬生長。其中包含17個區塊以及53種物種 (Mean = 0.11，CV = 0.66) ，且僅出現在單一區塊的物種具有28種。該地區的樣本覆蓋率為0.74。
4. 高地地區 (H, high country) 海拔2750公尺以上，該地區冬季寒冷且漫長，夏季短而乾燥，大部分的水分來自於夏季融雪。植被相較其他區域更矮小，土壤為裸岩與碎石居多。但因融雪灌溉，因此該地潮濕的棲息地不斷增加，包括溪流、季節性滲漏 (seepages)、湖泊、草地和沼澤等地形。其中包含69個區塊以及92種物種 (Mean = 0.07，CV = 1.28) ，且僅出現在單一區塊的物種具有39種。該地區的樣本覆蓋率為0.91。

**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，估計兩群落之間的共同物種數，在多數估計，無論是在共同物種數的估計值或是標準差的估計結果中，*BB*所估計之結果高於*Pan*所估計之，此情況與上一章所呈現之電腦模擬的結果相符。

並經由Jaccard 指數量化四群落間的Beta多樣性結果可以得知，山麓地區與下針葉林地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。

繪製在未使用估計式修正僅使用觀測值作為群落物種數，與使用估計式修 正群落物種數後之估計結果的樹狀結構圖 (dendrogram) 。在分類樹結果中，可以看到，僅由觀測值作為依據所繪製之分類樹結果，將山麓地區與下針葉林分為一類群，而上針葉林與高地地區做為第二類群；但在估計出的結果中，則是在山麓地區與下針葉林具有最近的距離，隨後與該群相近的為高地地區，而在上針葉林則與其餘三群落具有更遠的距離。在使用估計式修正群落內物種數所量化之Beta多樣性結果，相較於直接使用觀測物種數所估計之結果之間會有所不同。

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

**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 。

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。

# 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近一步的修正，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。