

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，其中一項常見的方式便是使用有母數的估計方式建立估計式。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 Beta 二項分佈的模型假設，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：物種豐富度、共同物種、Beta 二項分佈、動差法

Abstract

The proportion of shared species between two communities can be a simple and intuitive indicator of beta diversity. In many past studies, various estimation methods for the number of shared species between communities have been proposed for different sampling methods, survey techniques, or data types. One common approach in constructing estimation methods is to use parametric estimation methods to develop estimation equations. Additionally, many studies have focused on estimating the number of species within a single community, assuming that the proportion of species in appearance data blocks follows a beta-binomial distribution, and have used the moment method to construct estimation equations, all of which have achieved good estimation results. Therefore, based on the assumption of the beta-binomial distribution model, this study uses the moment method to construct three estimation equations for both samplings with replacement and without replacement. Multiple computer simulations evaluate the effectiveness and stability of the estimation equations. In terms of results, compared to existing methods for estimating shared species, the proposed method in this study yields smaller biases. It demonstrates better performance in terms of RMSE and 95% confidence interval coverage. Finally, the proposed estimation equations are applied to empirical data to compare the results with existing estimation equations. The estimation results from this study are higher than those from existing estimation equations when calculating shared species between communities.

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

目錄

摘要.....	i
Abstract.....	ii
目錄.....	iii
圖目錄.....	iv
表目錄.....	v
第 1 章 緒論.....	1
第 2 章 模型符號介紹與相關文獻回顧.....	4
2.1 符號定義.....	4
2.2 相關文獻回顧.....	6

圖目錄

找不到圖表目錄。

表目錄

找不到圖表目錄。

找不到圖表目錄。

第1章緒論

物種豐富度，即在某一特定群落中所存在的物種數，在生態多樣性的研究中，物種豐富度是一種最為簡單且直觀的指標，特別是在維持多種生態功能，以及在高多樣性的保護與管理上。此外，也有研究表明，物種豐富度與生態系統的功能性呈正向關係 (Shmida et al., 1985; Maestre et al., 2012)。故在群落物種豐富度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豐富度便成為一大難題。

在過去文獻中針對物種豐富度所進行的估計方式可依據所蒐集的資料型態大致分為兩種：個體豐富度資料與物種出現與否個體資料。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；而物種出現與否之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。因此在物種豐富度估計方法的推導上，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。例如在物種豐富度估計中常見的下界估計方法 (Chao 1984; 1987) 以及 jackknife 估計方法 (Burnham & Overton, 1978; 1979)。另一方面，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) 等方式進行估計。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豐富度之外，往往也需要針對多個群落之間進行比較。這些群落可能是一保護區隨時間的變化，或是不同海拔或緯度的生態差異等。在這其中，共同物種豐富度，即共同物種數，便是建構群落之間相似性以及 *Beta* 多樣性重要的基礎之一。因此，估計多個群落間存在的共同物種豐富度也是一樣重要的議題。在過去，已有許多研究針對不同的抽樣方式，依據或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，根據物種出現與否資，結合 Beta 二項式模型 (Beta-binomial model) 改善的估計的準確度以進行共同物種數的估計。

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。 $k = 1, 2, \dots, S_j$ ， $j = 1, 2$ ， $Q_{k(j)} = \sum_{i=1}^{S_j} I(X_i = k)$ 。
- 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

在大多數的物種數估計的研究中的理論假設皆為隨機且獨立的抽樣，但在通常情況下，個體抽樣的方式較難符合該假設，而相較於個體抽樣，區塊抽樣較能符合該理論假設。區塊抽樣單位通常為陷阱、區塊與固定時間調查。在出現型數據的抽樣中，大多數的方法是將其中的研究區域劃分為多個面積大致相同的區塊，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 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的矩陣 $[W_{ij}]$ 。其中，若是 i 物種出現在 j 區塊中，則計為 1；反之若尚未觀測到該物種則計為 0。

以第一群落為例， X_i 被定義為落樣本中第 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 物種出現的總區塊數量，故 $0 < X_i \leq t_1$ 。

並且，可令 $Q_{k(1)}$ 表示在出現頻率向量 X_i 中出現 k 次的物種數， $Q_{k(1)} = \sum_{i=1}^{S_1} I(X_i = k)$ ，且 $k = 1, 2, \dots, t_1$ 。故 $Q_{1(1)}$ 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， $Q_{0(1)}$ 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。

2.2.2 取後放回之抽樣方式

在生態調查的研究中，物種數或稱物種豐富度是最直接呈現多樣性的指標之一。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豐富度的結果。

對於出現型數據所開發之物種豐富度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。一般而言，傳統的捉放法是藉由該物種在樣本中所佔比例估計個體數，而在物種豐富度的估計中，可將捉放法「個體數」對應至「物種數」，以估計群落中的物種數作為物種豐富度的指標所使用。

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

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：

$$P(W_{ij} = w_{ij} | \pi_{i(1)}) = \prod_{i=1}^{S_1} \prod_{j=1}^{t_1} \pi_{i(1)}^{w_{ij}} (1 - \pi_{i(1)})^{1-w_{ij}}$$

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

$$P(X_i = x_i | \pi_{i(1)}) = \sum_{i=1}^{S_1} \binom{t_1}{x_i} \pi_{i(1)}^{x_i} (1 - \pi_{i(1)})^{t_1-x_i}$$

單群落物種數估計

無母數的物種豐富度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，不對物種豐富度或者物種出現機率的分布進行假設。Chao
(1987) 針對出現型數據建立物種豐富度的無母數估計模型 *Chao2*。以第一群落
作為單群落的例子，根據樣本中物種計數的機率分佈， $E(Q_{k(1)})$ 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豐富度。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 等人 (2014) 應用 Good-Turing 頻率公式，加入
 $Q_{3(1)}$ 與 $Q_{4(1)}$ 的資訊對其進行修正，得到更為準確的下界估計式 *iChao2*：

$$\hat{S}_{iChao2} = \hat{S}_{Chao2} + \frac{t_1 - 3}{4t_1} \frac{Q_{3(1)}}{Q_{4(1)}} \times \max \left(Q_{1(1)} - \frac{t_1 - 3}{2(t_1 - 1)} \frac{Q_{2(1)}Q_{3(1)}}{Q_{4(1)}}, 0 \right)$$

兩群落共同物種數估計

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。為此，當在比較兩群落之間的物種豐富度時，並非僅考慮單一群落的
物種豐富度，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，並加上已存在於樣本中的共同
物種數，作為修正的共同物種數所使用。

Pan 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，最終可以推導出估計式 *Pan*：

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

2.2.3 取後不放回之抽樣方式

相對於取後放回的抽樣方式，另一種在生態資料中常見的抽樣方法為取後不放回，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。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，則被紀錄為存在，反之則為不存在，最終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)：

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

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：

$$\hat{S}_{wChao2} = S_{obs} + \frac{E(Q_{1(1)})^2}{w2E(Q_{2(1)}) + rE(Q_{1(1)})}$$

其中， $w = \frac{t_1}{t_1 - 1}$ ， $r = \frac{q_1}{1 - q_1}$ 。

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

2.2.4 標準差估計

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

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

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

2.2.5 95%信賴區間

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

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

2.2.6 樣本涵蓋率

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

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

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

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

2.2.7 Jaccard 指標

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：

$$J_{index} = \frac{S_{12}}{S}$$

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：

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

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。

