# 緒論

# 模型符號介紹與相關文獻回顧

## 符號定義

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

* ：混合群落中的總相異物種數。
* ：第群落的物種數，。
* ：兩群落的共同物種數。
* ：群落中物種出現的機率，即相對豐富度。。
* ：第群落中物種出現的機率，即相對豐富度。，。
* ：第一群落樣本，第物種出現的區塊數量。
* ：第二群落樣本，第物種出現的區塊數量。
* ：兩樣本中出現的共同物種數。。
* ：第群集的總區塊數量，。
* ：第群集的抽樣區塊數量，。
* ：抽樣比例，。
* ：群落樣本中物種的共出現在個區塊的機率。。
* ：第一群落樣本中物種的共出現在個區塊，且同時第二群落樣本中物種的共出現在個區塊的機率。，。
* ：樣本中剛好出現個區塊的物種數。，。
* ：樣本中在第一族群出現個區塊，並在第二族群出現個區塊的物種數。，。
* ：樣本中在第一族群出現次，並在第二族群出現的物種數。。
* ：樣本中在第一族群出現，並在第二族群出現次的物種數。。

## 相關文獻回顧

### 出現率數據

在大多數的生態調查研究中，抽樣為隨機且獨立的。且抽樣單位通常為陷阱、區塊與定時調查。在出現率數據的抽樣中，大多數的方法是將其中的研究區域劃分為多個面積大致相同的T個區塊，並從中隨機選擇特定的t個區塊做為抽樣樣本進行調查。對於不同類型的物種，準確計算每個抽樣區塊中出現的個體數往往是一件相對困難的事，在計算抽樣區塊中的無脊椎動物、植物或是微生物此情況更甚。因此在多數情況下，調查時僅記錄該物種在t個抽樣區塊中的出現率，即紀錄物種在該區塊出現與否。出現率數據由一組為t個抽樣區塊的樣本所組成，並記錄每個區塊中每種物種的出現或是未出現，以形成一個具有S行與t列的矩陣。其中，若是在第j個區塊中發現i物種，則計為1；反之若尚未觀測到該物種則計為0。

又被定義為該樣本中的出現頻率向量，，表示在該樣本中第i物種出現的總區塊數量。若則表示該抽樣樣本中並無觀測到該物種，且在樣本中觀測到的物種總數為，故。

並且，可令表示在出現頻率向量中出現k次的物種數，，且。故為在該樣本中僅出現在一個區塊的物種數，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，為在該樣本並未被觀測到的物種數。而真實的物種數，應為被觀測到的物種數與未被觀測到的物種數之總和。

### 取後放回之物種數估計

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

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

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

依據上一小節所述，出現率數據的樣本來自於依照物種抽樣區塊的出現率矩陣所組成。並可將該矩陣整理成出現頻率向量。應為服從伯努力分佈 (Bernoulli distribution) 的隨機變數，且當時機率為，而時機率為。則發生率矩陣的機率分佈為：

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

#### 單群落物種數估計

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

因此求得在樣本中未出現以及分別出現一次與兩次的物種數之期望值為：

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

固可求得：

最終將結果帶入 ，可知估計式 *Chao2*為：

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

#### 兩群落的共同物種數估計

在生態統計中，群落之間的共同物種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，而是必須針對群落間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 et al. (2009) 將*Chao2*的方法推廣至兩群落，建立一估計兩群落間存在的共同物種數之估計式。

假設在第一群落的樣本 () 與第二群落的樣本 () 中，分別有、個抽樣區塊。且兩群落的第i物種出現機率分別表示為、。則 表示在出現頻率向量中出現k個區塊，同時在中出現個區塊的物種數。可以表示為：

因此求得在樣本中分別未出現於兩群落的期望值為：

為求得下界估計式，因此需要與 的資訊：

又根據柯西-施瓦茨不等式之概念可以推導出：

固可求得：

同理可以求得：

而若是想要求得下界估計式，則需要與 的資訊：

又根據柯西-施瓦茨不等式之概念可以推導出：

最終得：

最終將結果帶入 ，可知估計式 *Pan*為：

### 取後不放回之物種數估計

相對於取後放回的抽樣方式，另一種在生態資料中常見的抽樣方法為取後不放回。在取後不放回的抽樣方法中，廣泛使用在林業調查中，依照所選區塊對樹木進行不重複取樣，或是用於陷阱或誘捕器的抽樣方式中，需要殺死個體的抽樣方法。

在這種類型抽樣方法的單群落情況中，假設將該地區分為個大致相等的區塊，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，針對群落進行取後不放回之隨機抽樣，分別抽取*t*的區塊數，僅記錄每個採樣樣本中物種的發生率。又每個區塊物種存在的機率為，，為未知參數。模型假設在區塊中，物種 僅能在的目標區塊中被檢驗到，亦為未知參數，且。則在給定的條件下遵循參數和的零截尾二項分佈 (zero-truncated beta-binomial distribution) (Shen and He, 2008)：

|  |  |
| --- | --- |
|  |  |

當取後不放回之隨機抽樣從*T*個區塊中抽出*t*個區塊，並且每個樣本區塊中僅記錄物種的存在與否，以形成逐種樣本發生矩陣，在給定 的情況下，應遵循超幾何分佈 (hypergeometric distribution)：

|  |  |
| --- | --- |
|  |  |

#### 單群落物種數估計

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

因此求得在樣本中未出現以及分別出現一次與兩次的物種數之期望值為：

根據柯西-施瓦茨不等式 (Cauchy-Schwarz inequality) 之概念可以得：

將其中整理以便後續計算：

則不等式中的第二項可寫作：

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

故可將不等式整理為：

移項後得：

最終將結果帶入 ，可知估計式 *wChao2*為：

其中，，。

#### 兩群落的共同物種數估計

與取後放回的估計方法相似，在取後不放回的估計中也存在兩群落間的共同種估計需求。在此假設在第一群落的樣本 () 與第二群落的樣本 () 中，分別有、個抽樣區塊。且兩群落的第i物種出現機率分別表示為、。則 表示在出現頻率向量中出現k個區塊，同時在中出現個區塊的物種數。在給定與的情況下，可以表示為：

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的共同種，可得最終估計式為：

其中，，，

### 標準差估計

根據的漸近分布，其服從大小為以及機率為的多項分布 (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 delta 方法導出，表示為

|  |  |
| --- | --- |
|  |  |

其中

### 95%信賴區間

在此，物種豐富的信賴區間通過假符合對數常態分佈 (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：

|  |  |
| --- | --- |
|  |  |

其中 以此計算95%信賴區間的樣本涵蓋率 (95% confidence interval coverage rate, 95% CI Coverage) 。

# 以動差法修正共同物種數的估計

## 取後放回之抽樣方法的估計方式

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

為了評估物種檢測概率的異質性，應用生態學研究中廣泛使用的混合二項式模型。假設 遵循二項分佈，其中，，是機率密度函數為獨立同分佈的隨機變數。故假設服從 以確保物種相對豐富度的靈活性，獲得以下樣本之物種計數的邊際分佈如下：

又令表示在個區塊中準確觀測到的物種數，為在單群落樣本中出現次的區塊數。並根據柯西-施瓦茨不等式 (Cauchy-Schwarz inequality) 之概念與Good-Turing頻率公式 (Good, 1953, 2000) 的出近似式：。該近似式表示，稀有物種可以為未被觀測到的物種豐富度提供更多的估計資訊。

又根據樣本中物種計數的邊際機率分佈，可以表示為：

依據上述式子，可獲得為觀測以及出現一次至三次的物種豐富度期望值：

並將其推廣至兩群落，其中為兩樣本的物種豐富度正好分別為和的平均機率。

令表示在第一群落個區塊且在第二群落個區塊數中準確觀測到的物種數。則樣本中觀測到的共同物種數為。

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

並將 設定為1，且。成立以下近似值：

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

經化簡 求得。又化簡 之後，可得。隨後，依據不等式 = ，故將作為帶入 ，最終可獲得：

表示：若時，則；若時，則。

同理，可經由上述相同方式推導出：

|  |  |
| --- | --- |
|  |  |
|  |  |

並加入對估計式進行修正，最終得估計式

其中：

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

且表示：若時，則；若時，則。

## 取後不放回之抽樣方法的估計方式

依據2.2.3節所示，在目標區塊中檢視到物種i所存在的區塊數，其分布應服從一個二項分佈；且第i物種出現的區塊數所組成的出現頻率向量，在在給定 的情況下，服從一超幾何分佈。又與有關，來自於，因此可推導出：

|  |  |
| --- | --- |
|  |  |

其中，*I(A)*為指標函數，表示若出現*A*情況時，則該式為1，反之則即為0。並在此假設為一來自beta分佈的隨機樣本，故可將式子表示為：

|  |  |
| --- | --- |
|  |  |

其中。

後隨邊際分佈可通過獲得，故：

|  |  |
| --- | --- |
|  |  |

Chiu (2022) 基於Good-Turing頻率公式與柯西不等式 (Cauchy-Schwarz inequality) 之概念，針對單一群落的估計得出近似式：，其中為出現個區塊的物種數。從中可以得知，在物種估計時，採取出現較少次的物種，可以更多提供未出現物種的資訊，有助於縮小物種豐富度的估計結果。根據等式 ()，給出未觀測到的豐富度的期望值、唯一值和重複值：

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

並將上述概念推廣至兩群落，為兩樣本的物種豐富度正好分別為和的平均機率：

|  |  |
| --- | --- |
|  |  |

同時令為在樣本中第一群落出現次且第二群即出現次的區塊數，則為樣本中觀測到的共同物種數量，。

|  |  |  |
| --- | --- | --- |
|  |  | (10) |
|  |  | (12) |
|  |  | (13) |

將 設定為1，且。成立以下近似值：

|  |  |
| --- | --- |
|  |  |
|  |  |

得 ，代入 。得：

|  |  |
| --- | --- |
|  |  |

表示：若時，則等於；若，則等於。

同理 也依此證明，得：

|  |  |
| --- | --- |
|  |  |

表示：若時，則等於；若，則等於。

又：

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

並成立以下近似值：

|  |  |
| --- | --- |
|  |  |

將，代入 後，並加入對估計式進行調整，最終得估計式：

|  |  |
| --- | --- |
|  |  |

其中：

|  |  |
| --- | --- |
|  |  |

並在 () 中代入 。

|  |  |
| --- | --- |
|  |  |
|  |  |

表示：若時，則等於；若，則等於。

並在的基礎上，加入 對的估計進行修正，成立以下近似值：

|  |  |
| --- | --- |
|  |  |

經由式 () 與 式 () 推得出：

|  |  |
| --- | --- |
|  |  |

又，

|  |  |
| --- | --- |
|  |  |
|  |  |

並成立以下近似式：

|  |  |
| --- | --- |
|  |  |

由上推得：

|  |  |
| --- | --- |
|  |  |

又可從式 (18) = 式 (22) 得：

|  |  |
| --- | --- |
|  |  |
|  |  |

並依公式解，得：

|  |  |
| --- | --- |
|  |  |

最終得：

|  |  |
| --- | --- |
|  |  |

其中等於 (14)、等於 (15) 且 等於 (16)，且在 (14) 中的與使用帶入 (25) 計算。

此外，在Chao and Lin (2012) 中提出兩群集的取後不放回的共同種估計方式：

|  |  |
| --- | --- |
|  |  |

其中 與 。

# 模擬研究與討論

## 估計式整理

### 估計式整理

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。綜上所述，以下為不同估計方法之整理：

其中：

|  |
| --- |
|  |
|  |
|  |

且表示：若時，則；若時，則。

其中：

並在 的估計式中代入 。

表示：若時，則等於；若，則等於。

並在的估計式中代入 。而與的估計與相同。

其中 與 。

1. 標準差估計

根據的漸近分布，其服從大小為以及機率為的多項分布 (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 delta 方法導出，表示為

|  |  |
| --- | --- |
|  |  |

其中

1. 95%信賴區間
2. wBB1 and wBB2
3. Other

在估計式為下界估計式時所使用。在此，物種豐富的信賴區間通過假設符合對數常態分佈 (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：

|  |  |
| --- | --- |
|  |  |

其中 以此計算95%信賴區間的樣本涵蓋率 (95% confidence interval coverage rate, 95% CI Coverage) 。

## 模擬設定

## 模擬結果

### 取後放回之抽樣方法模擬結果

### 取後不放回與取後放回之抽樣方法模擬結果比較

### 討論

# 實例分析