# 使用動差法估計共同物種數

## 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種物種， 其中，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個抽樣區塊，並且 表示第*i*物種在樣本中出現的區塊數量。則遵循總數為，機率為的二項分佈 (binomial distribution) 。在此，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 遵循二項分佈，且() 為機率密度函數 服從 的隨機變數。假設 為樣本中出現的區塊數正好為的平均機率，最終獲得以下樣本之物種出現區塊數的機率分佈如下：

又令表示在樣本中出現 個區塊的物種數，而。並根據樣本中物種出現次數的機率分佈，可知可以表示為：

依據上述式子，可獲得、 以及 ：

並根據柯西-施瓦茨不等式之概念與Good-Turing頻率公式 (Good, 1953, 2000) 得出近似式：。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 與 皆分別遵循二項分佈以及。並假設，；，，且機率密度函數分別為服從 與的與。將定義為樣本中出現的區塊數正好分別為 和 的平均機率。則：

其中與。令 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。藉此，可獲得、以及：

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |
|  | (3) |

將 設定為1，可藉由式 (1)、式 (2) 與式 (3) 成立以下近似值：

並經由化簡 可得：

|  |  |
| --- | --- |
|  | (4) |

以及經由求得：

並由於 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 () 服從Beta二項分佈時，則估計量 的偏差近似於 。

且由柯西-施瓦茨不等式可推得 ，並結合式 (4) 可得知 的偏差的下界近似為不等式：

故將 帶入式 (4)，且為了確保估計是的穩定與，將滿足 ，最終可獲得：

且在上述等式中皆表示為：若 時，則；而若 時，則表示為。

同理，可經由上述相同方式推導出與：

最後加入 以確保估計式的穩定，對其進行修正，最終得估計式：

其中、以及分別為：

且在上述等式中皆表示為：若 時，則；而若 時，則表示為。

## 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，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，在給定 的情況下，應服從超幾何分佈。又與有關，來自於，因此可推導出：

其中，為指標函數，表示若發生時，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 遵循超幾何分佈，且() 為服從 的隨機變數，故機率密函數 可表示為：

其中 與 ，且。

並且 的機率分布可依據 與 獲得 (Shen & He, 2008; Chiu, 2023)。設 為樣本中出現的區塊數正好為的平均機率，則：

令表示在個區塊中準確觀測到的物種數，而為在單群落樣本中出現個區塊數。故根據，可定義出 為：

依據上述式子可知、以及分別為：

並基於Good-Turing頻率公式與柯西-施瓦茨不等式之概念，針對單一群落樣本的估計得出近似式：。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 與 皆分別遵循超幾何分佈，並假設，；，，且機率密度函數分別為服從 與 的 與 。將 定義為樣本中出現的區塊數正好分別為 和 的平均機率，則：

其中，。

並又令 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 為樣本中觀測到的共同物種數量，。藉此，可獲得、以及：

|  |  |
| --- | --- |
|  | (5) |
|  | (6) |
|  | (7) |

依據式 (5) 與式 (6) 成立以下近似值。並將 設定為1，且假設：

由上述式子可得出：

|  |  |
| --- | --- |
|  | (8) |

將並將其代入 ，可得：

同理 也依此證明，可得：

|  |  |
| --- | --- |
|  | (9) |

並將該結果帶入可求得：

同時，在兩群群落樣本同時為、以及分別為：

|  |  |
| --- | --- |
|  | (10) |
|  | (11) |

並依據式 (10) 與式 (11) 成立以下近似值：

將式 (8)、(9) 代入 後，並為確保估計式穩定加入對估計式進行調整，最終得估計式：

其中、以及分別為：

|  |  |
| --- | --- |
|  | (12) |
|  | (13) |
|  | (14) |

在的基礎上，加入 的資訊對 的估計進行修正，依據式 (10) 與式 (11) 成立以下近似值：

並經由該式可推得出：

|  |  |
| --- | --- |
|  | (15) |

又經由可以得知，以及 分別為：

|  |  |
| --- | --- |
|  | (16) |
|  | (17) |

並依式 (16) 與式 (17) 成立以下近似式：

並由上述式子推得：

|  |  |
| --- | --- |
|  | (18) |

又可從式 (15) = 式 (18) 得：

並依公式解，且皆大於0，故得：

|  |  |
| --- | --- |
|  | (19) |

最終得估計式為：

其中，將式 (19) 的與 的結果分別帶入

中可分別求得 、 與 的估計式。