### 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，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：