中國統計學報
第38卷第3期
八十九年九月
231-241頁

千禧年香港觀鳥大賽資料分析

趙蓮菊　林淑萍　楊欣洲　葉兆輝*
清華大學統計研究所　*香港大學統計及精算學系

摘要

本文利用2000年香港觀鳥大賽二十隊參賽隊伍所觀察到鳥種之數據，來說明生態學中重複捕取模式在種類數估計上之實際應用。利用摺刀法與樣本涵蓋方法估計未被記錄的鳥種數量及全部鳥種數的信賴區間。

關鍵詞：重複捕取，鳥種，摺刀法，種類數估計，樣本涵蓋。

美國數學會分類索引：主要62P10；次要62G05。

1. 香港觀鳥大賽簡介

自1984年以來，世界自然（香港）基金會(World Wide Fund For Nature Hong Kong)每年舉辦一年一度之香港觀鳥大賽。2000年以前時間均在四月舉行，而2000年則第一次改在二月舉行。2000年的比賽於二月二十五日下午五時至二十六日下午五時止，共有20隊參加，其中18隊爲香港本地隊伍，另有來自廣州與北京的邀請隊。通常每隊有四個成員，在廿四小時之內走遍香港每個角落，把發現的野生鳥種記錄下來，記錄最多種類的隊伍贏得冠軍。記錄的鳥種只論“種類有無看到或聽到”而不管每一個種類的“數目多寡”。例如甲隊看到綠繡眼10隻，乙隊看到1隻，兩隊均可記錄此種類，而無須記錄隻數多寡。鳥種的記錄必須四人均看到或聽到才可計算，對於只聽到而沒有看到的鳥種須在記錄表上註明（每隊在賽前由主辦單位提供鳥名記錄單），若所聽到的鳥種是極稀有種或意料外之鳥種，則交由裁判審愼處理。每隊必須在記錄表中詳列看到該鳥種的時間與地點。

在2000年2月25日觀鳥大賽的開幕典禮中，並邀請1999年度香港小姐之冠軍、季軍及美腿小姐先作小型之“賽前賽”之觀鳥比賽，測試她們的觀鳥功力，她們必須在20分鐘內把記錄表上所列鳥種與典禮附近米埔水禽飼養池中所看到之鳥種“連連看”。1999香港小姐冠軍郭羨妮女士的感受是（見世界自然（香港）基金會2000年2月25日新聞稿，取自 http://www.wwf.org.hk/cpress/000225.htm）「觀鳥是富挑戰性及有趣的活動，由於雀鳥所處的位置可能隨時變動，辨識它們需靠高度的集中力。然而，觀賞雀鳥帶來的樂趣卻是盡在不言中」。未來倘若能以“鳥功”來作爲選“美”之評審準則以替代時下選美之“身材”、“美腿”標準，一定別有一番新的氣象！

2000年的觀鳥大賽由傻鳥隊(Birdbrains)記錄到154種而贏得冠軍，此隊由香港大學研究生物多樣性的Brian Morton教授領隊。20隊參賽隊伍共記錄到220種鳥類。有趣的是好幾隊皆以鳥名命名，例如傻鳥隊、翠鳥隊、鷹隊、夜鷹隊和白鶺鴒隊。世界自然（香港）基金會每年利用此觀鳥大賽來籌款，方式如下：在二十隊當中，任何民衆可選擇一個想贊助的隊伍，只要捐款200元港幣以上或該隊每發現一種鳥種至少捐款2元（例如某隊發現150種，則捐款爲300元以上），則有資格參加猜獎遊戲，贏者可得Pentax望遠鏡一

部。另外，個人捐款最高者可得Swarovski望遠鏡。以望遠鏡爲獎品當然是因爲望遠鏡是所有賞鳥者的必備用具之一。參賽的隊伍大部份是由企業贊助參加。此種募款在最近幾年均募得百萬港幣以上。世界自然（香港）基金會在過去十幾年間利用此項募款，致力多項保護環境及教育的工作，包括改善香港的米埔鳥類保護區之環境。米埔鳥類保護區每年棲息的鳥類約六萬隻，其中包括瀕臨絕種的黑面琵鷺（香港稱黑臉琵鷺），而每年至保護區內參加教育及活動的學生及一般人士超過四萬人。

2. 千禧年資料分析

2000年20支隊伍及其發現鳥種的資料由世界自然（香港）基金會提供，此資料包含了220×20的一個矩陣（220爲發現之所有不同鳥種數，20爲隊伍數），若某種鳥種被某隊記錄，則記爲1，否則爲0。此20隊中，每一隊記錄到鳥種的數目如下（令n_i：第i隊記錄到的鳥種）。

表1　二十隊參賽隊伍各隊伍記錄之鳥種數目

i	1	2	3	4	5	6	7	8	9	10	
n_i	137	154	96	124	145	125	101	142	122	107	
i	11	12	13	14	15	16	17	18	19	20	總和
n_i	135	129	127	121	148	141	125	130	125	116	2550

此資料之頻率次數（f_j：恰被j隊記錄到的鳥種數）則在下列：

表2　二十隊記錄鳥種之頻率次數分布

j	1	2	3	4	5	6	7	8	9	10	
f_j	21	16	13	10	4	13	6	4	11	1	
j	11	12	13	14	15	16	17	18	19	20	總和
f_j	6	5	8	3	4	6	11	15	8	55	220

20隊共記錄了220種，主要的問題是：是否已完全記錄了當日所有香港的鳥種？亦即是否存在極稀有的種類，在此比賽的24小時之間並未被任何一隊所看到？

就資料的形式而言，由於對每一種類只記錄“有無看到”，因此整個資料矩陣類似一個20樣本之重複捕取資料。一般重複捕取是用在估計單一種類的鳥口總數問題，例如黑面琵鷺的總數。由於從外觀上無法分辨每一隻，因此通常須作記號或套腳環（稱爲繫放），用以辨別不同的每一隻，並記錄在每一次捕取樣本中是否捕到。但若在種類數的估計中，由於不同種類由外觀即可區別，因此不須捕取作記號，只須用望遠鏡觀看即可，而一般重複捕取的模式與估計便可直接應用在種類數的估計上，可參考Burnham & Overton(1979)，Bunge & Fitzpatrick (1993)與Boulinier等(1998)。

由於一般普遍鳥種幾乎被所有的參賽隊所記錄，而稀少種通常只被一、二隊記錄，或可能根本沒有被任何隊記錄。因此在表2之頻率數中，j愈大之f_j所代表的種類即爲普遍種，j愈小之f_j所代表者爲稀有種。由表2可知，有許多鳥種爲普遍種，但也有一些爲稀有種。因此對鳥種而言，豐盛程度有高有低，而被記錄的機率（即至少看到該鳥種某一隻的機率）因鳥種而異，此種不均勻性，在生態學上稱爲“heterogeneity”，對應的模式爲$\mathcal{M}_h$，亦即每一鳥種有不同的被記錄機率，但對每一隊而言，此機率相同。在此20隊中，由表1可知各隊“鳥功”不同，因此必須考慮各隊不同效應(team effects，在重複捕取上亦稱time effects，因爲通常是在不同時間捕取樣本的相對效應)，此種同時考慮heterogeneity與time effects之乘積模式稱爲$\mathcal{M}_{th}$。

在種類數估計問題中，$\mathcal{M}_h$與$\mathcal{M}_{th}$模式如下：假設有N個鳥種，K組參賽隊伍。令p_{ij}爲第i個鳥種被第j隊記錄的機率，$i=1,\ldots,N,\ j=1,\ldots,K$；$p_i$爲第$i$個鳥種之豐盛度，用以表現個體差異；$e_j$則爲第$j$隊的觀鳥功力，也就是任一種鳥，其是由第$j$個隊伍觀測得的機會，用以表現不同參賽隊之效果。在模式$\mathcal{M}_{th}$下，假設$0<p_{ij}=p_ie_j\le 1,\ i=1,\ldots,N,\ j=1,\ldots,K$；在模式$\mathcal{M}_h$下，令$e_1=\cdots=e_K=1$，簡化成$0<p_{ij}=p_i\le 1,\ i=1,\ldots,N,\ j=1,\ldots,K$。

對$\mathcal{M}_h$模式而言，最常用的估計種類數之估計量爲由Burnham & Overton (1978)所提出之摺刀估計量(jackknife)。摺刀法考慮以觀測得之總相異種類數

M_{K+1} 爲估計式的基礎，$M_{K+1}=\sum_{j=1}^{K} f_j$，每次系統化地刪除$l$隊資料，重新估計。經證明可知，第$l$階之摺刀估計量$\hat{N}_{Jl}$爲頻率次數的一個線性組合，符號上可表示成

$$\hat{N}_{Jl}=\sum_{j=1}^{K} a_{lj} f_j \text{，}$$

上式中，常數a_{lj}是l階摺刀估計下頻率數爲j時對應之組合係數。細節可參見Burnham & Overton (1978, 1979)，以下僅列出$l=1,2$的情形：

$$\begin{aligned}\hat{N}_{J1} &= M_{K+1}+(\tfrac{K-1}{K})f_1 \text{，}\\ \hat{N}_{J2} &= M_{K+1}+(\tfrac{2K-3}{K})f_1-\tfrac{(K-2)^2}{K(K-1)}f_2 \text{，}\end{aligned}$$

即$a_{11}=(2K-1)/K,\ a_{12}=\cdots=a_{1K}=1;\ a_{21}=(3K-3)/K,\ a_{22}=(3K-4)/K(K-1),\ a_{23}=\cdots=a_{2K}=1$。在表3中，我們列出前二階之摺刀估計值及用重抽法(bootstrap resampling)計算得之標準差估計。重抽法的進行程序將在本節末加以說明。

由於摺刀估計通常有隨階次的增加而偏差降低，但變異增大的現象，Burnham & Overton (1978, 1979)進而提出內插摺刀估計量(interpolated jackknife)。其概念是透過逐次檢定

$$\begin{aligned}H_{0l} &: E(\hat{N}_{J,l+1}-\hat{N}_{J,l})=0 \text{，}\\ H_{1l} &: E(\hat{N}_{J,l+1}-\hat{N}_{J,l})\neq 0 \text{，}\end{aligned}$$

以決定出最佳的摺刀估計階次。在虛無假設下，檢定統計量

$$Z_l=\frac{\hat{N}_{J,l+1}-\hat{N}_{J,l}}{\{[M_{K+1}/(M_{K+1}-1)][\sum_{j=1}^{K}(a_{l+1,j}-a_{lj})^2 f_j-(\sum_{j=1}^{K}(a_{l+1,j}-a_{lj})f_j)^2/M_{K+1}]\}}$$

具有漸進標準常態分配，對應檢定之P值(P-value)爲P_l。透過上述的逐次檢定原理，由$l=1$開始檢定。則在顯著水準α下，可得$g=\min\{l:P_l>\alpha\}$爲最佳的摺刀估計階次，而內插摺刀估計爲

$$\begin{aligned}\hat{N}_J &= \hat{N}_{J1},\ g=1 \text{，}\\ \hat{N}_J &= c\hat{N}_{J,g}+(1-c)\hat{N}_{J,g-1},\ 1<g<5 \text{，}\\ \hat{N}_J &= \hat{N}_{J5},\ g\geq 5 \text{，}\end{aligned}$$

其中，內插係數$c=(0.05-P_{g-1})/(P_g-P_{g-1})$。在Boulinier(1998)之論文中推薦使用此估計量。對於我們分析的數據中，經計算可得$Z_1=0.982$，$P_1=0.326$，因此內插摺刀估計即為第一階摺刀估計。

作者之一(Chao, Lee & Jeng, 1992; Lee & Chao, 1994)曾提出針對模式$\mathcal{M}_h$與$\mathcal{M}_{th}$利用樣本涵蓋法之估計量對種類數進行推估。所謂「樣本涵蓋」，針對此二模式是指被記錄到的鳥種機率(或豐盛度)總和佔所有鳥種機率(或豐盛度)總和的比率，符號上可表示為

$$C=\frac{\sum_{i=1}^{N}p_i I[\text{第}i\text{個鳥種至少被其中一隊記錄}]}{\sum_{i=1}^{N}p_i},$$

上式中，$I[A]$為指示變數，事件A發生其值為1，若不發生其值為0。值得注意的是，樣本涵蓋會隨著樣本改變，因此其為含參數$\{p_1,\ldots,p_N\}$的隨機變數。由於樣本涵蓋一般容易估計且估計精確度高，而研究目的一種類數，卻是不易直接估計，因此可透過樣本涵蓋與種類數的關係來協助估計種類數。若各鳥種被發現的機會(或豐盛度)均等，亦即在條件$p_1=\cdots=p_N$下，一個很自然的種類數估計是

$$\hat{N}=\frac{M_{K+1}}{\hat{C}},$$

其中$\hat{C}=1-f_1/\sum_{j=1}^{K}jf_j$。此估計最早由Darroch & Ratcliff (1980)提出。但通常實際情形中，有些鳥種極易被發現（普遍種），有些卻鮮少被發現（稀有種），此時可透過「變化係數」γ來表現各鳥種間被發現機率的不均勻現象，變化係數可表示為$\gamma=[\sum_{i=1}^{N}(p_i-\bar{p})^2/N]^{1/2}/\bar{p}$，其中$\bar{p}=\sum_{i=1}^{N}p_i\Big/N$。當發現機率的不均勻性愈高時，變化係數也愈高，而不均勻性愈低時，變化係數則愈接近於0。Chao, Lee & Jeng (1992)，Lee & Chao (1994)證明在模式$\mathcal{M}_{th}$下，種類數估計量為

$$\hat{N}_s=\frac{M_{K+1}}{\hat{C}_s}+\frac{f_1}{\hat{C}_s}\hat{\gamma}_s^2,\ s=1,2, \tag{1}$$

上式中，樣本涵蓋估計$\hat{C}_1$及其偏差矯正估計$\hat{C}_2$分別為

$$\begin{aligned}\hat{C}_1 &= \hat{C}=1-f_1\Big/\textstyle\sum_{j=1}^{K}jf_j,\\ \hat{C}_2 &= 1-[f_1-2f_2/(K-1)]\Big/\textstyle\sum_{j=1}^{K}jf_j;\end{aligned}$$

變化係數平方之估計則為

$$\hat{\gamma}_s^2 = \max\left\{\frac{\hat{N}_{0,s}\sum_{j=2}^{K} j(j-1)f_j}{2\sum\sum_{j<k} n_j n_k} - 1, 0\right\},\ s = 1, 2,$$

其中，$\hat{N}_{0,s} = M_{K+1}/\hat{C}_s,\ s = 1, 2$。

當模式降爲$\mathcal{M}_h$時，(1)式仍爲適當的種類數估計，只是變化係數平方之估計可進一步化簡爲

$$\hat{\gamma}_s^2 = \max\left\{\frac{\hat{N}_{0,s}K\sum_{j=2}^{K} j(j-1)f_j}{(K-1)(\sum_{j=1}^{K} jf_j)^2} - 1, 0\right\},\ s = 1, 2。$$

本文中，有關種類數估計之標準差，將利用重抽法加以計算。爲方便說明，在此僅以三組參賽隊伍（$K = 3$）爲例，說明重抽法計算某一估計量$\hat{N}$的標準差。令I_j爲鳥種是否曾被第j組參賽隊伍記錄的指標，$I_j = 1$表示該鳥種曾被第j隊記錄，$I_j = 0$表示未曾被記錄，因此鳥種的記錄歷史可以用(I_1, I_2, I_3)來表示。現令$Z_{I_1I_2I_3}$表示記錄結果爲(I_1, I_2, I_3)的鳥種數目，例如：Z_{101}表示被第一隊與第三隊記錄過的鳥種數目。因此共有七個記錄值$(Z_{100}, Z_{010}, Z_{001}, Z_{110}, Z_{101}, Z_{011}, Z_{111})$，以及從未被記錄到的鳥種數目$Z_{000}$。但由於資料中$Z_{000}$未知，考慮用$\hat{Z}_{000} = \hat{N} - M_4$估計之。假設$(Z_{000}, Z_{100}, Z_{010}, Z_{001}, Z_{110}, Z_{101}, Z_{011}, Z_{111})$爲一組來自多項分配

$$Multinomial(\hat{N};\ \frac{\hat{Z}_{000}}{\hat{N}}, \frac{Z_{100}}{\hat{N}}, \frac{Z_{010}}{\hat{N}}, \frac{Z_{001}}{\hat{N}}, \frac{Z_{110}}{\hat{N}}, \frac{Z_{101}}{\hat{N}}, \frac{Z_{011}}{\hat{N}}, \frac{Z_{111}}{\hat{N}})$$

的隨機變數向量，依此分配重複抽取2000組重抽樣本(replications)，分別計算這2000組重抽樣本之種類數估計$(\hat{N}_1, \ldots, \hat{N}_{2000})$，並依此結果計算其樣本標準差以作爲總類數估計標準差之估計。同樣過程可推廣至任意多數目之隊伍。

在信賴區間的計算上，Chao (1989) 將$\log(\hat{N} - M_{K+1})$視爲一具有漸進常態分配的隨機變數，而提出種類數估計$\hat{N}$的95%信賴區間公式

$$[M_{K+1} + (\hat{N} - M_{K+1})/A\ ,\ M_{K+1} + (\hat{N} - M_{K+1})A],$$

其中，

$$A = \exp(1.96 \times \{\log[1 + \hat{V}(\hat{N})/(\hat{N} - M_{K+1})^2]\}^{1/2}),$$

而$\hat{V}(\hat{N})$爲$\hat{N}$之變異數估計。本文中採用上述的估計，好處之一爲此信賴區間之下界一定會超過被記錄的相異種類數。

在表3中除了列出模式$\mathcal{M}_h$中前二階摺刀估計量及其標準差外，並同時列出樣本涵蓋法在模式$\mathcal{M}_{th}$下之二個估計值（見式(1)）。由於樣本涵蓋法所求得之估計量對模式$\mathcal{M}_h$與$\mathcal{M}_{th}$各項結果差不多，故在表3中只列出模式$\mathcal{M}_{th}$之值。表3中並列有利用2000次重抽樣本所得之標準差及95%信賴區間，同時並列出二估計量對應之變化係數之值。由變化係數估計值均爲0.62可知，明顯的不均勻性存在而不能忽略。（若我們忽略此不均勻性，則估計值$\hat{N} = M_{K+1}/\hat{C} = 220/(1-21/2550) = 222$極可能低估了眞正的種類數。）

表3　種類數估計及標準差估計

模式	估計方法	變化係數估計值	估計值	95%信賴區間
$\mathcal{M}_h$	$\hat{N}_{J1}$		240（標準差6.3）	(231, 256)
	$\hat{N}_{J2}$		245（標準差10.3）	(232, 275)
$\mathcal{M}_{th}$	$\hat{N}_1$	0.62	230（標準差4.4）	(225, 243)
	$\hat{N}_2$	0.62	230（標準差4.6）	(224, 244)

3. 結論與討論

由表3我們可知，第一階摺刀法估計量顯示約仍有20種鳥種在此次觀鳥大賽中未被記錄到，全部種類數95%信賴區間爲（231, 256）。而樣本涵蓋則顯示約有10種未被記錄到，本文介紹之種類數的二個估計量的95%信賴區間分別得到（225, 243）與（224, 244）。

在香港彈丸之地，「米埔鳥類保護區」之設立誠然已是難能可貴，這也提供世界極度瀕危之黑面琵鷺的一個重要棲息地。在2000年時米埔保護區記錄到黑面琵鷺178隻。主辦香港觀鳥大賽的世界自然（香港）基金會亦將黑面琵鷺的保護與研究視爲其重要保育工作之一。尤其近年來，很多發展商利用不同方法與理由做藉口，不斷向米埔溼地和附近的地方大興土木，對溼地的自然環境構成莫大的威脅。再者，爲了改善大陸內地與香港的交通，當地政府將會興建一條橫跨溼地的鐵路，米埔溼地和鳥類能否繼續生存，頓成疑

問？令人扼腕悲嘆。

而在台灣情況更危急，每年在台灣七股溼地度冬的黑面琵鷺是全世界最多的族羣，黑面琵鷺如此愛戀台灣，而回報它們的是不僅無法提供一個「保護區」，工業用地更節節進逼它們的棲息地，其處境已是危在旦夕。

常有人們提出「鳥和人究竟誰重要？」，其言下之意，好像隱喻人的生存當然重於鳥。然而，人與鳥都是自然界共生的一環，「今日鳥類、明日人類」，今日鳥類面臨棲息地的破壞與強奪，當鳥類無法在該土地生存時，整個生態結構破壞，過了不久，人類也無法在該土地生存了。惟有尊重自然，尊重生命，各類生物才能和諧地在大自然中永續發展。

致謝詞：感謝主編何淮中教授的邀稿及審查者的修改意見。本文所討論的資料係由世界自然（香港）基金會提供，感謝所有參與此次觀鳥大賽者羣策羣力貢獻鳥種資料，以及朱麗珊（Maxine Chu）女士與Michael Chalmers先生花費寶貴時間整理提供全部資料，謹致謝忱，並向世界自然（香港）基金會多年來積極維護自然環境生態的目標與努力致敬。

參考文獻

Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E. and Pollock, K. H. (1998). Estimating Species Richness: The Importance of Heterogeneity in Species Detectability. *Ecology* **79**, 1018-1028.

Bunge, J. and Fitzpatrick, M. (1993). Estimating the Number of Species: A Review. *Journal of the American Statistical Association* **88**, 364-373.

Burnham, K. P. and Overton, W. S. (1978). Estimation of the Size of a Closed Population When Capture Probabilities Vary among Animals. *Biometrika* **65**, 625-633.

Burnham, K. P. and Overton, W. S. (1979). Robust Estimation of Population Size When Capture Probabilities Vary among Animals. *Ecology* **60**, 927-936.

Chao, A. (1989). Estimating Population Size for Sparse Data in Capture-Recapture Experiments. *Biometrics* **45**, 427-438.

Chao, A., Lee, S.-M. and Jeng, S.-L. (1992). Estimating Population Size for Capture- Recapture Data When Capture Probabilities Vary by Time and Individual Animal. *Biometrics* **48**, 201-216.

Darroch, J. N. and Ratcliff, D. (1980). A Note on Capture-Recapture Estimation. *Biometrics* **36**, 149-153.

Lee S.-M. and Chao, A. (1994). Estimating Population Size via Sample Coverage for Closed Capture-Recapture Models. *Biometrics* **50**, 88-97.

[民國89年5月25日收稿，民國89年8月9日接受。]

analysis of Hong Kong Big Bird Race data for the year of

Anne Chao, Shu-Ping Lin, Hsin-Chou Yang and Paul S. F. Yip*
Institute of Statistics, National Tsing Hua University
*** Department of Statistics and Actuarial Science, Hong Kong University**

ABSTRACT

The species data of Hong Kong Big Bird Race for the year of 2000 are used to illustrate the practical application of ecological capture-recapture models to species estimation. The jackknife and sample coverage methods are used to estimate the number of non-recorded species in the race and an associated confidence intervals.

Key words and phrases: Bird species, capture-recapture, jackknife, sample coverage, species estimation.

AMS 1991 subject classifications: Primary 62P10; secondary 62G05.