仰觀宇宙之大，俯察品類之盛:
統計在生物多樣性的應用

趙蓮菊, 2022統計學術研討會
淡江大學

前言

- 感謝淡江大學統計系楊文主任邀請
- 感謝淡江大學統計系所有工作人員辛勞;特別感謝李百靈教授及張雅梅教授安排與聯絡
- 在瘟疫蔓延時,特別感謝大家的參與聽講


青出於清(華)而勝於清
江山代有才人出的英才
Kaohsiung

大綱:

- 簡介生物多樣性
- 統計之父R. A. Fisher (費雪)在1943年最早的資料與統計模型, 及其後針對物種數估計的發展
- 電腦科學之父Alan M. Turing（圖靈）二次大戰時所發展的頻率理論及樣本涵蓋觀念
- （若時間許可）簡短介紹用樣本涵蓋標準化樣本之稀釋與外插統計方法

(因時間關係,大部分細節將省略)

仰觀宇宙之大

俯察品類之盛：生物的多樣1

俯察品類之盛:生物的多樣2

俯察品類之盛:生物的多樣3

變異數和生物多樣性

- 針對一組數值的資料，其變異數說明此資料變異分散的程度

- 當資料為生物類別和其存在的環境相關變數時，用以描述此生物資料的變異與分散即為生物多樣性。

生物多樣性

- 生物多樣性（biodiversity: biological diversity）一詞最早在1986年美國華盛頓區舉行之 “生物多樣性國家論壇”（*National Forum on Biological Diversity*) 提出最早出現該會議所出版的論文集的標題（Wilson and Peter 1988）

- 生物多樣性是指地球上所有生物的多樣化與變異性，它涵蓋了所有從基因、個體、族群、物種、群落、生態系到地景等各種層次的生命型式

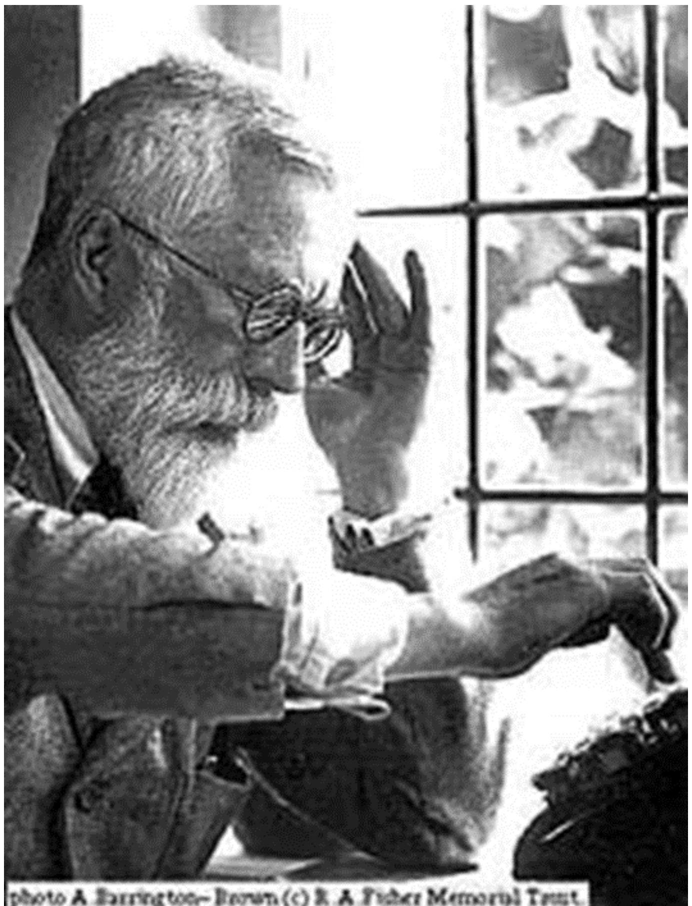
各種層次的生命型式


Biosphere
Biome
Ecosystem
生態系
Community
群落
Population
族群
Individual
個體
基因

IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services) 生物多樣性和生態系統服務跨國政府間科學政策平台2015 定義

- 各種層次的生命變異與多樣不僅展現在基因型的分類、生物外觀表型之分類、系統演化與親緣關係的分類、以及物種特質在生態系的功能分類
- 生物體之間交互作用(食物鏈或網路多樣性)的差異。
- 更涵蓋不同時間及空間中生物分佈的範圍及其豐富程度的變化。

Fisher, Corbet & Williams 1943 paper
Journal of Animal Ecology



THE RELATION BETWEEN THE NUMBER OF SPECIES AND THE NUMBER OF INDIVIDUALS IN A RANDOM SAMPLE OF AN ANIMAL POPULATION

BY R. A. FISHER (*Galton Laboratory*), A. STEVEN CORBET (*British Museum, Natural History*) AND C. B. WILLIAMS (*Rothamsted Experimental Station*)

(With 8 Figures in the Text)

CONTENTS

		PAGE
Part 1.	Results obtained with Malayan butterflies. By A. S. CORBET	42
Part 2.	Results obtained by means of a light-trap at Rothamsted. By C. B. WILLIAMS	44
Part 3.	A theoretical distribution for the apparent abundance of different species. By R. A. FISHER	54

PART 1. RESULTS OBTAINED WITH MALAYAN BUTTERFLIES

By A. STEVEN CORBET (*British Museum, Natural History*)

It is well known that the distribution of a series of biological measurements usually conforms to one of three types:

(*a*) the binomial distribution, where the frequencies are represented by the successive terms of the binomial $(q+p)^n$;

(*b*) the normal distribution, in which the results are distributed symmetrically about the mean or average value, and which is the special case of (*a*) when p and q are equal;

(*c*) the Poisson series, in which the frequencies are expressed by the series

$$e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\dots\right),$$

where m is the mean and e is the exponential base 2·7183.

The usual practice of calculating the arithmetic mean in a set of measurements of a biological nature, such as wing length of a butterfly, assumes a distribution showing no wide departure from normality, although it does not appear that this procedure has been vindicated.

It is the usual experience of collectors of species in a biological group, such as the Rhopalocera, that the species are not equally abundant, even under conditions of considerable uniformity, a majority being comparatively rare while only a few are common. As far as we are aware, no suggestion has been made previously that any mathematical relation exists between the number of individuals and the number of species in a random sample of insects or other animals. Recently, it has been found (Corbet, 1942) that, leaving out of account the commoner species of which no attempt was made to collect all individuals seen, the number of species S of butterflies of which n individual specimens were collected by a single collector in Malaya was given closely by the expression

$$S=C/n^m,$$

where C and m are constants.* When m is unity, as is the case with the Malayan collection, and has since been found to be a condition which obtains with collections of butterflies from Tioman Island and the Mentawi Islands in which the relation between S and n follows the above equation, the number of species of which 1, 2, 3, 4, ... specimens were obtained was very close to a series in harmonic progression. Thus, the series can be written

$$C\left(1+\tfrac{1}{2}+\tfrac{1}{3}+\dots\right).$$

Although this relation holds accurately with the rarer species, there is less agreement in the region of the common species; in fact, theoretical considerations preclude an exact relationship here.

Prof. Fisher (see Part 3) has evolved a logarithmic series which expresses accurately the relation between species and individuals in a random sample throughout the whole range of abundance:

$$S=n_1\left(1+\frac{x}{2}+\frac{x^2}{3}+\dots\right),$$

where S is the total number of species in the sample, n_1 the number of species represented by single specimens, and x is a constant slightly less than unity but approaching this value as the size of the sample is increased.

The total number of individual specimens, N, in the sample at all levels of abundance is given by $N=n_1/(1-x)$ and $N(1-x)/x$ is a constant α independent of the size of the sample. As the size of the sample is increased, n_1 approaches α.

The Fisher series has been established for all the entomological collections tested in which there was

* This equation may be written

$$\log S=\log C+m\log n,$$

so that the plot of the logarithms of S and n is a straight line.

Corbet收集馬來亞蝴蝶的資料

reason to believe that the collecting had been unselective (see Table 1). It is clear that when S and N are known, as is usually the case, the statistics n_1, x and α can be calculated. It is a curious fact that the number of uniques should approach a constant value with increasing size of collection. It is important to ascertain how far this type of distribution of individuals among species holds in other zoological groups, for it would appear that we have here an effective means of testing whether a collection has been made under conditions approaching random sampling or whether some degree of selection has been exercised, a consideration which is often of some importance in faunistic studies.

Table 1. *Entomological collections examined showing the relation between the numbers of species and individuals*

	Observations				Calculations			
	Total individuals	No. between 1 and 24	Total species	No. between 1 and 24	x	n_1 calc.	n_1 found	α
Malayan Rhopalocera	—	3306	620	501	0·997	135·05	118	135·47
Rhopalocera from Tioman Island (east coast of Malaya) (Malay collector, 1931)	157	—	41	—	0·887	15·96	19	18·00
Rhopalocera from Mentawi Islands (excluding Hesperiidae) (C. Boden Kloss and N. Smedley, 1924)	1,878	890	135	110	0·983	32·77	37	33·35
Karakorum Rhopalocera (Mme J. Visser-Hooft; *vide* Evans, 1927)	403	195 (1–28)	27	24 (1–28)	0·984	6·42	6	6·52
Mexican Elmidae (Col.) (H. E. Hinton; *vide* Hinton, 1940)	11,798	—	35	—	0·9998	4·72	4	4·72

How far the results obtained with any particular collection can be regarded as representative of the distribution of the same species group in the area in which the collection was made must obviously depend on the uniformity or otherwise of the conditions prevailing when the collection was made and to the extent to which these conditions are representative of the habitat. In an equatorial forest-clad island with no mountain heights above 2000 ft., conditions are very uniform as far as such orders as Leipidoptera are concerned; although even here some allowance must be made for the fact that collections of butterflies made in such regions are usually poor in the crepuscular species. In temperate climates, it is evident that results obtained during one period of the year are usually inapplicable to other seasons or to the year as a whole. It would appear that the results obtained with the moth trap at Harpenden (see Part 2) can be regarded as giving an accurate picture of the distribution frequencies of the phototropic moths in the area, and it is probable that the same is true of the collection of Mexican Elmidae.

With many collectors, and for a variety of reasons, the collecting of common species is discontinued once a certain number of specimens of these are obtained. In the case of the Malayan Rhopalocera cited, collecting of all individuals seen was not continued after 24 specimens had been taken. In such cases, we have the following information:

The total number of species under 25 individuals per species:

$$S_{(1-24)} = n_1\left(1 + \frac{x}{2} + \frac{x^2}{3} + \dots + \frac{x^{23}}{24}\right).$$

The total number of individuals at frequencies below 25 per species:

$$N_{(1-24)} = n_1\,(1 + x + x^2 + \dots + x^{23}).$$

Table 2, which gives the results obtained with the Malayan butterflies, is based on these considerations, and shows the very close relation between the observed and the calculated results.

Table 2. *Calculated and observed distribution frequencies of butterflies collected in Malaya*

The values in the second column are obtained from the Fisher series given on p. 42, taking $x = 0·997$.

n	S (calc.)	S (found)	Deviations
1	135·05	118	17·05
2	67·33	74	−6·77
3	44·75	44	0·75
4	33·46	24	9·46
5	26·69	29	−2·31
6	22·17	22	0·17
7	18·95	20	−1·05
8	16·53	19	−2·47
9	14·65	20	−5·35
10	13·14	15	−1·86
11	11·91	12	−0·09
12	10·89	14	−3·11
13	10·02	6	4·02
14	9·28	12	−2·72
15	8·63	6	2·63
16	8·07	9	−0·93
17	7·57	9	−1·43
18	7·13	6	1·13
19	6·74	10	−3·26
20	6·38	10	−3·62
21	6·06	11	−4·94
22	5·77	5	0·77
23	5·50	3	2·50
24	5·25	3	2·25
			0·82

Corbet馬來亞蝴蝶所有的資料

S_{obs} = 620物種, n = 9031隻

f_k: 資料中出現 k 次的物種數

Table S4.2. Abundance frequency counts of the Malayan butterfly survey (Fisher, Corbet & Williams 1943), S= 620 species, 9031 individuals, CV=1.435, H=5.736. The count f_k is the number of species represented by exactly k individuals in the survey

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
f_k	118	74	44	24	29	22	20	19	20	15	12	14	6	12	6

k	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
f_k	9	9	6	10	10	11	5	3	3	5	4	8	3	3	2

k	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
f_k	5	4	7	4	5	3	3	3	3	1	1	2	1	1	4

k	46	48	49	50	51	52	53	54	55	56	58	59	60	64	66
f_k	2	2	1	3	1	1	2	4	1	5	2	1	2	1	1

k	68	70	71	76	84	89	92	93	100	105	108	119	141	147	194
f_k	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1

R. A. Fisher (1890-1962)貢獻

數學家, 統計學家, 生物學家, 遺傳學家, 演化生物學家

取自wiki 英文版

Fisher's exact test, Fisher's inequality, Fisher's principle, Fisher's geometric model, Fisher's Iris data set, Fisher's linear discriminant, Fisher's equation, Fisher information, Fisher's method, Fisherian runaway, Fisher's fundamental theorem of natural selection, Fisher's noncentral hypergeometric distribution Fisher's z-distribution, Fisher transformation, Fisher consistency, F-distribution, F-test, Fisher–Tippett distribution, Fisher–Tippett–Gnedenko theorem, Fisher–Yates shuffle, Fisher–Race blood group system, Behrens–Fisher problem, Cornish–Fisher expansion, von Mises–Fisher distribution, family allowance, Wright–Fisher model…（續下頁）

R. A. Fisher (1890-1962)貢獻（續）

Ancillary statistic, Fiducial inference, Intraclass correlation
Infinitesimal model, Inverse probability, Lady tasting tea,
Null hypothesis, Maximum likelihood estimation, Neutral theory of molecular evolution, Particulate inheritance, Random effects model
Relative species abundance, Reproductive value, Sexy son hypothesis
Sufficient statistic, Analysis of variance, Variance

遺漏了 Fisher's alpha

遺漏了 Fisher's log series distribution

Corbet 當年提出的問題

- 假若再去馬來亞多看兩年, 會再多看到幾種沒看過的呢?
- 假若在馬來亞時間夠久, 大約有多少種蝴蝶呢？

物種數估計在許多學科的應用

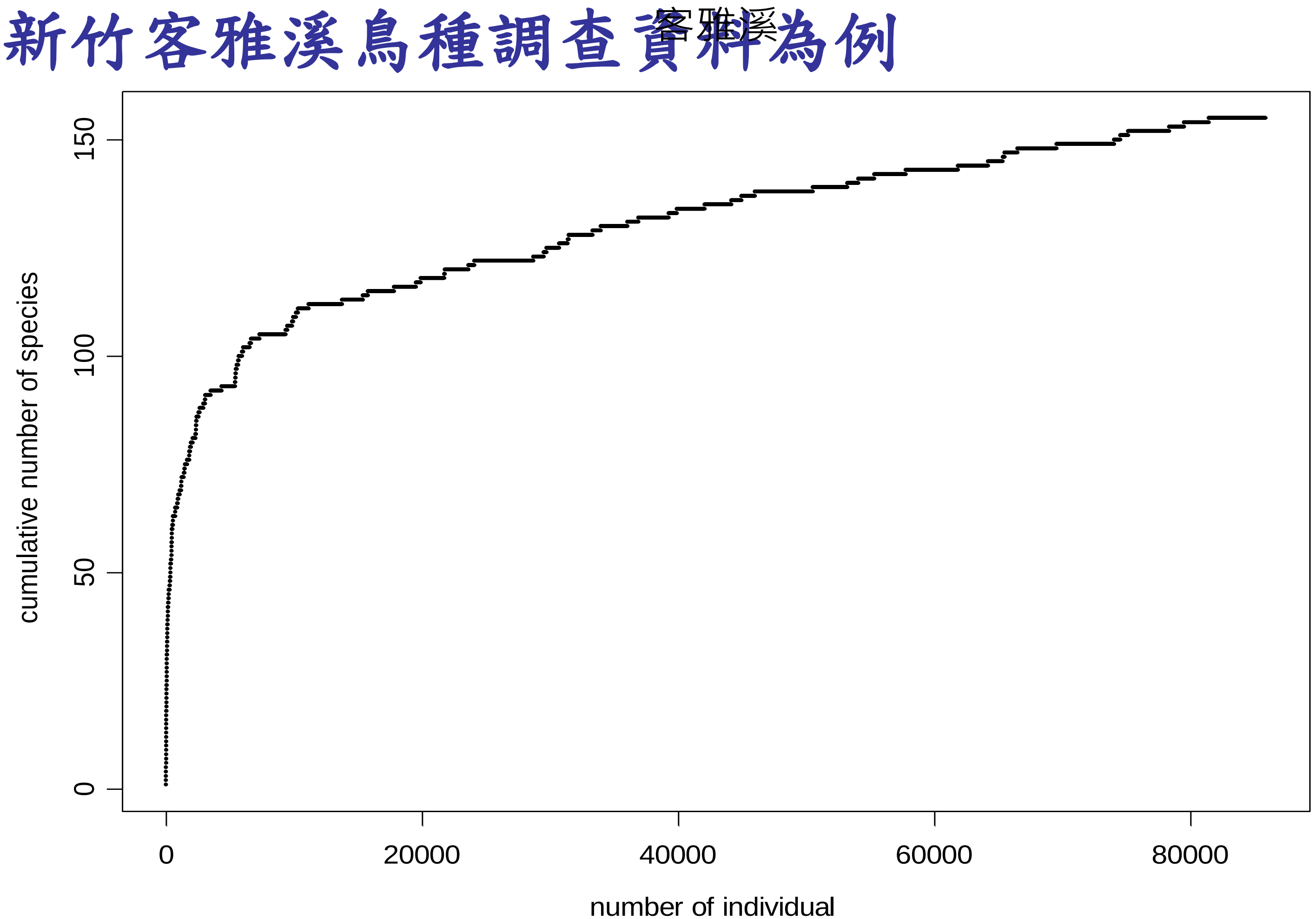
除了生物群落上的種類外尚有以下應用：

- 族群總數（重複補取資料）
- 醫學及流行病學上之患病人數
- 天文學中估計星球的數目
- 環境學中之污染物
- 軟體可靠度中的錯誤數
- 錢幣學或考古學上的鑄模或工具的種類
- 文學上作者之文字庫(不同的字)
- 中古世紀文學作品存留至今的比例

古典的做法:曲線配適法

配適物種累積曲線

以新竹客雅溪鳥種調查資料為例



Fisher's model: Poisson-gamma

S 未知物種

- 假設第i個物種出現為一Poisson(λ_i)分布
- λ_i 代表物種豐富度 $\lambda_1, \lambda_2, \ldots \lambda_S \sim gamma(a,b)$
- 第i個物種出現次數為 X_i
- 第i個物種出現 k 次機率為 ($k = 0, 1, \ldots$)

$$p_{a,b}(k) = E[I(X_i = k)] = \underset{\lambda}{E}\ \underset{X_i|\lambda}{E}\ I(X_i = k \mid \lambda_i)$$

$$= \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{1}{\Gamma(a) b^a} \lambda^{a-1} e^{-\frac{\lambda}{b}} d\lambda = \frac{\Gamma(a+k)}{k!\,\Gamma(a)} \cdot \left(\frac{1}{1+b}\right)^a \left(\frac{b}{1+b}\right)^k$$

負二項分布

Fisher's alpha

- For *gamma(a,b)*

  $mean = ab, \ var = ab^2, \ CV^2 = 1/a$

- Fisher let $a \to 0$ $\quad \theta = \frac{b}{1+b}$

$$P(X_i = k \mid X_i > 0) \to \frac{\theta^k}{k[-\log(1-\theta)]}$$

Fisher's alpha: 最早的diversity index

- 出現 k 次的物種數 $(k = 1, 2\ldots)$

$$E(f_k) \to S\frac{\theta^k}{k} \cdot \frac{1}{-\log(1-\theta)} \equiv \alpha\frac{\theta^k}{k}, \quad where \quad \alpha = \frac{S}{-\log(1-\theta)}$$

Fisher's log–series 分布

出現 1, 2, 3, .. k 次的物種數

$$\alpha\theta, \quad \frac{\alpha\theta^2}{2}, \quad \frac{\alpha\theta^3}{3}, \quad \ldots, \quad \frac{\alpha\theta^k}{k}, \ldots$$

如何從資料解參數??

Given data (S_{obs}, n), Fisher alpha can be solved by two equations

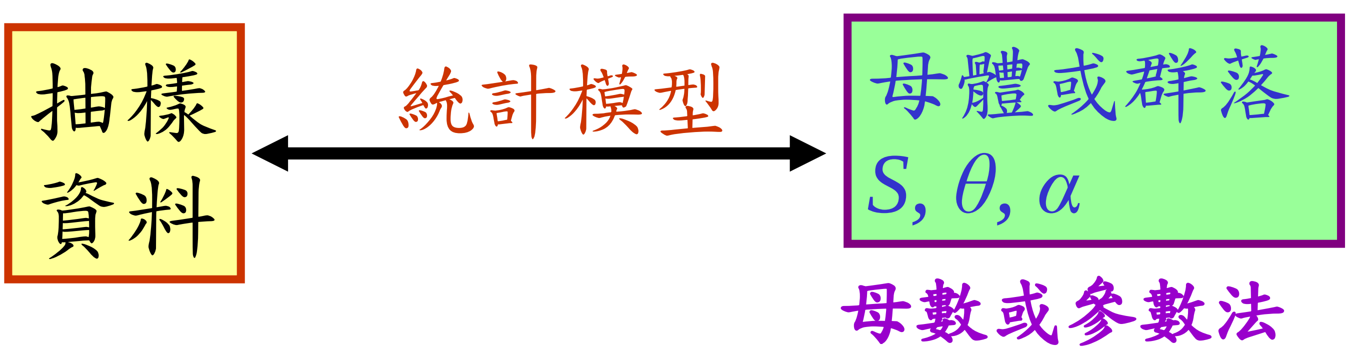
$$\alpha = \frac{S_{obs}}{-\ln(1-\theta)} \qquad n = \sum k f_k = \alpha \frac{\theta}{1-\theta}$$

- Fisher‘s alpha 僅由 S_{obs}, n 決定
- Fisher's model 無法估計物種數

統計模型應用於生物多樣性

抽樣模型：聯繫已知抽樣資料與未知母體參數

以管窺天：以一湯匙海水估整個大海的鹹度



物種數的無參數估計：$g(\lambda)$ 可為任何分布

- Poisson-mixture model

$$E(f_k) = S\int_0^\infty \lambda^k \frac{e^{-\lambda}}{k!} g(\lambda)d\lambda,$$

- Cauchy-Schwarz inequality

$$\int_0^\infty \frac{e^{-\lambda}}{0!} g(\lambda)d\lambda \times 2\int_0^\infty \lambda^2 \frac{e^{-\lambda}}{2!} g(\lambda)d\lambda \ge \left[\int_0^\infty \lambda \ \frac{e^{-\lambda}}{1!} g(\lambda)d\lambda\right]^2$$

$$\Leftrightarrow E(f_0)\times 2E(f_2) \ge [E(f_1)]^2 \quad \Rightarrow E(f_0) \ge \frac{[E(f_1)]^2}{2E(f_2)}$$

$$\hat{S}_{\text{Chao1}} = S_{obs} + \frac{f_1^2}{2 f_2}$$

Chao1 (1984) lower bound in Colwell & Coddington (1994)

統計模型

- 連續型模型 (時間固定)

  假設第i個物種出現為一Poisson(λ_i)分布

  λ_i代表物種豐富度

  $(X_1, X_2, \ldots, X_S \mid \sum_{i=1}^{S} X_i = n)$ follow a multinomial distribution with cell total n and cell

  probabilities $p_k = \lambda_k / \sum_{i=1}^{S} \lambda_i$, $k = 1, 2, \ldots, S$.

- 離散型 (n 固定)

  $$(X_1, X_2, \ldots, X_S) \sim Multinomial(n; p_1, p_2, \ldots, p_S)$$

  p_i:相對豐富度或機率

Multinomail and Binomial-Mixture models:

- Multinomial model

$$(X_1, X_2, \ldots, X_S) \sim Multinomial(n; p_1, p_2, \ldots, p_S)$$

- Binomial-mixture model

$$E(f_k) = S \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} h(p)\,dp,$$

- Slightly different form

$$\hat{S}_{\text{Chao1}} = S_{obs} + \frac{(n-1)}{n} \frac{f_1^{\,2}}{2 f_2}$$

Assessing uncertainty under multinomial model

$$\hat{\text{var}}(\hat{S}_{\text{Chao1}}) = f_2\left[\frac{1}{4}\left(\frac{n-1}{n}\right)^2\left(\frac{f_1}{f_2}\right)^4 + \left(\frac{n-1}{n}\right)^2\left(\frac{f_1}{f_2}\right)^3 + \frac{1}{2}\left(\frac{n-1}{n}\right)\left(\frac{f_1}{f_2}\right)^2\right]$$

95% interval of S is obtained

$$[S_{obs} + (\hat{S}_{\text{Chao1}} - S_{obs})/R,\ S_{obs} + (\hat{S}_{\text{Chao1}} - S_{obs})R]$$

$$R = \exp\{1.96[1 + \hat{\text{var}}(\hat{S}_{\text{Chao1}})/(\hat{S}_{\text{Chao1}} - S_{obs})^2]^{1/2}\}$$

尚有許多其他估計方法

SpadeR 套件 (species richness) output

(2) SPECIES RICHNESS ESTIMATORS TABLE:

	Estimate	s.e.	95%Lower	95%Upper
Homogeneous Model	655.631	7.474	643.724	673.514
Homogeneous (MLE)	620.000	0.017	620.000	620.078
Chao1 (Chao, 1984)	714.071	22.663	679.057	769.844
Chao1-bc	712.030	22.204	677.738	766.689
iChao1 (Chiu et al. 2014)	737.064	13.563	713.354	766.795
ACE (Chao & Lee, 1992)	712.239	17.351	684.000	752.937
ACE-1 (Chao & Lee, 1992)	737.207	23.932	698.872	794.174
1st order jackknife	737.987	15.361	711.512	772.121
2nd order jackknife	781.985	26.604	737.654	843.021

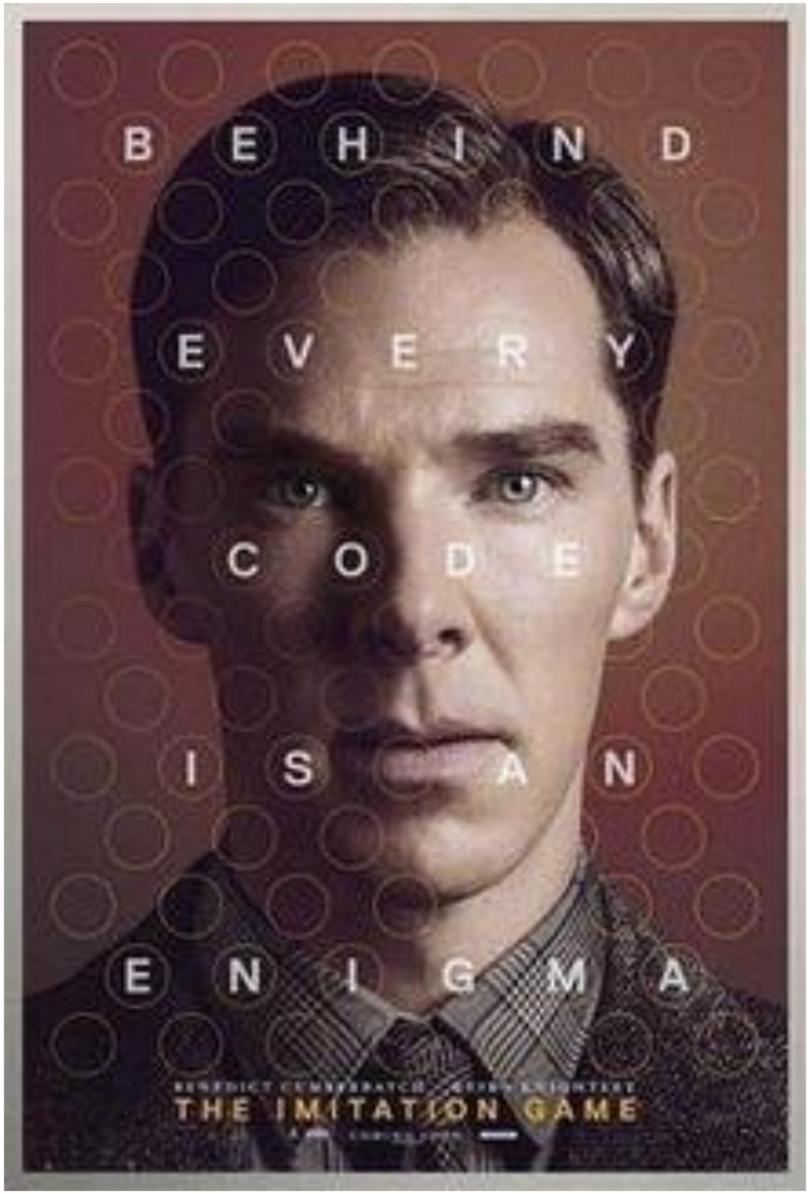
何時下界為不偏? (Chao et al. 2017)

- ❑ 若mixture is degenerate (所有物種豐富度或機率為相同)但除族群估計外此模型無多大用處
- ❑ 應用Good-Turing 頻率理論: 模型假設可放鬆許多,而適合許多實際估計問題

Alan M. Turing 圖靈 (1912-1954):

- 電腦科學與人工智慧之父
- 二戰英雄，圖靈領銜的科學團隊，成功破譯德軍極為嚴密的「恩尼格瑪」（Enigma）密碼，致使盟軍在大西洋等重要戰役中擊敗納粹，加速二戰的終結
- 圖靈在 1946 年喬治六世國王頒受大英帝國勳章，以表彰他在戰時的貢獻

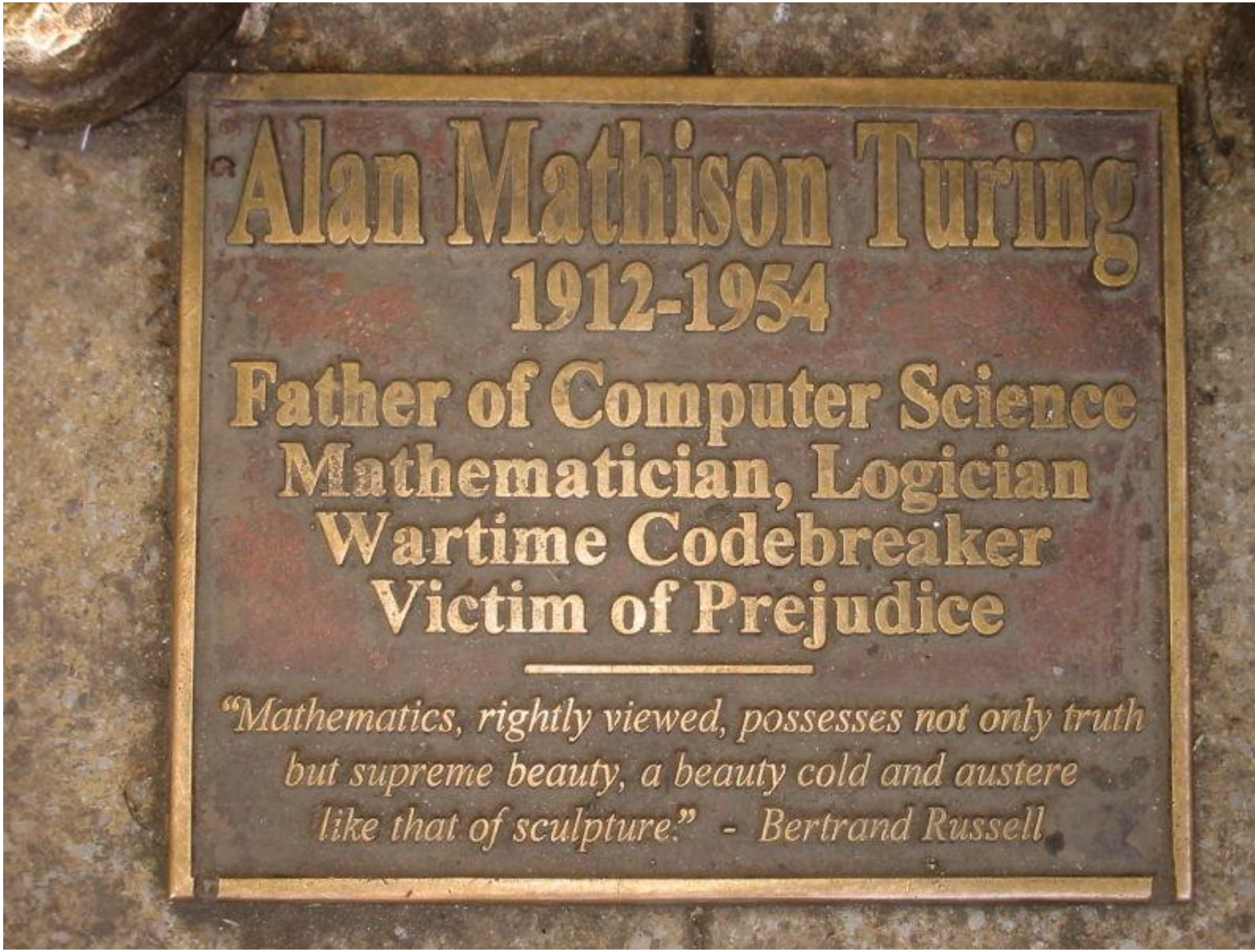
模仿/解碼遊戲 The Imitation Game



班奈狄克·康柏拜區
任男主角

圖靈的紀念雕像
英國曼徹斯特

2016 十二月
訪問圖靈的紀念雕像
獻上蘋果



Turing memorial statue plaque in Sackville Park, Manchester, UK

電腦科學之父, 數學家, 邏輯學家, 戰時密碼破解者, 偏見下的犧牲者

數學包含了真實與極致的美麗, 一種如雕像冷酷且嚴峻的美麗

圖靈:偏見下的犧牲者

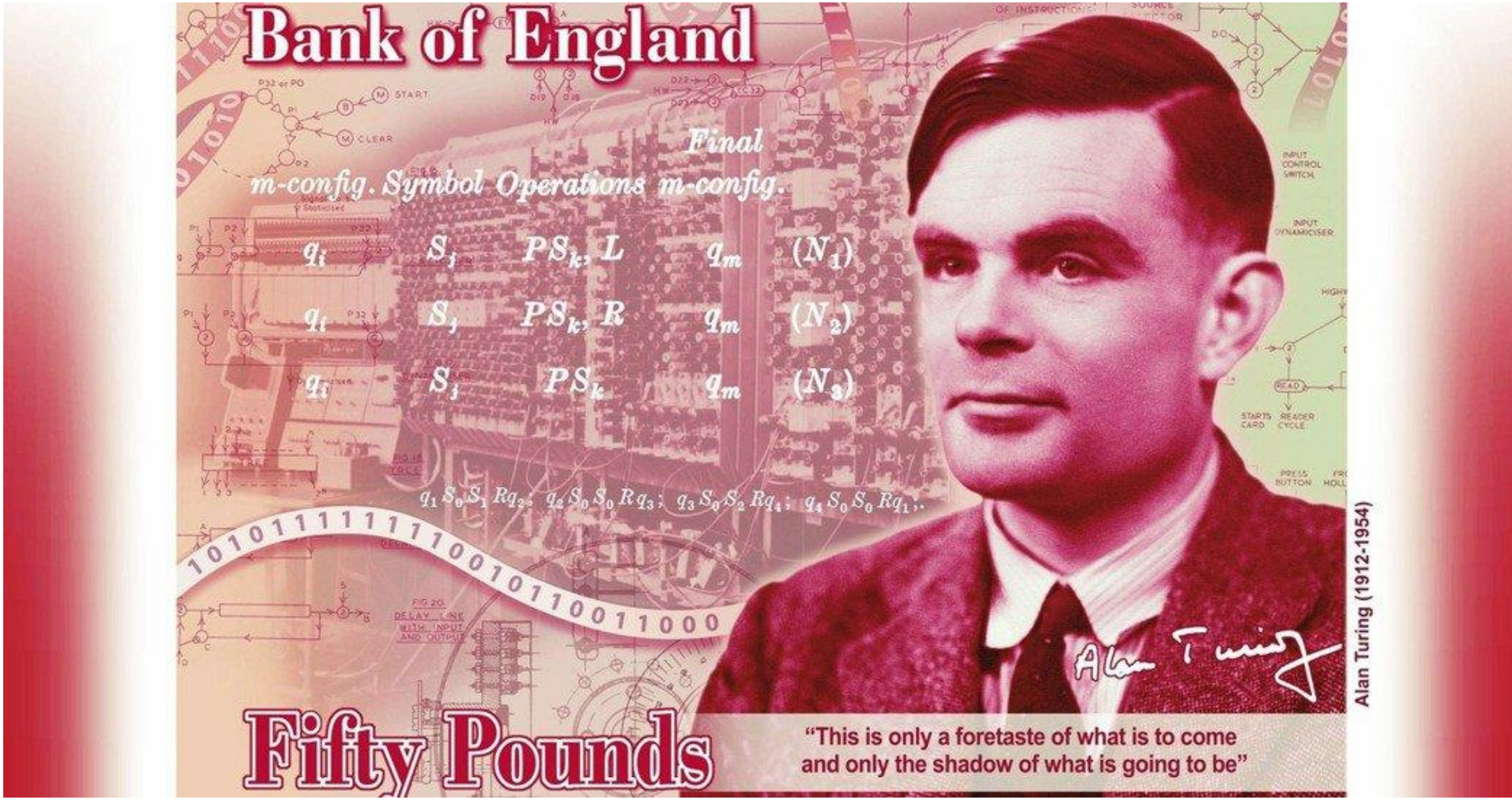
- 圖靈是生不逢時的同性戀者
- 1952年，他的同性伴侶闖進圖靈的房子行竊，但是英國警方的調查結果使得他被控以「明顯的猥褻和性顛倒行為」罪。他因此被判定有罪。
- 審判後，他給予兩個選擇：坐牢或女性荷爾蒙注射（即化學去勢）他選擇了後者
- 1954年，圖靈吃沾氰化物的蘋果而自殺死亡。

蘋果--如同夏娃反抗上帝的旨意?
賈伯斯蘋果電腦商標有關?

平反「同性戀之罪」

- 英國政府一直到才在2009年由時任首相布朗撰文正式道歉
- 2013年，英國女王特赦圖靈「同性戀之罪」，也進一步鼓舞、催生了2017年生效的《艾倫·圖靈法》，赦免曾因同性戀性而遭定罪的男同性戀者，約有近49000人因此獲得法律上遲來的平反。
- 自1966年以來，圖靈獎被廣泛認為是電腦科學世界的最高榮譽，相當於諾貝爾獎

2021.6.23年發行50英鎊紀念鈔票

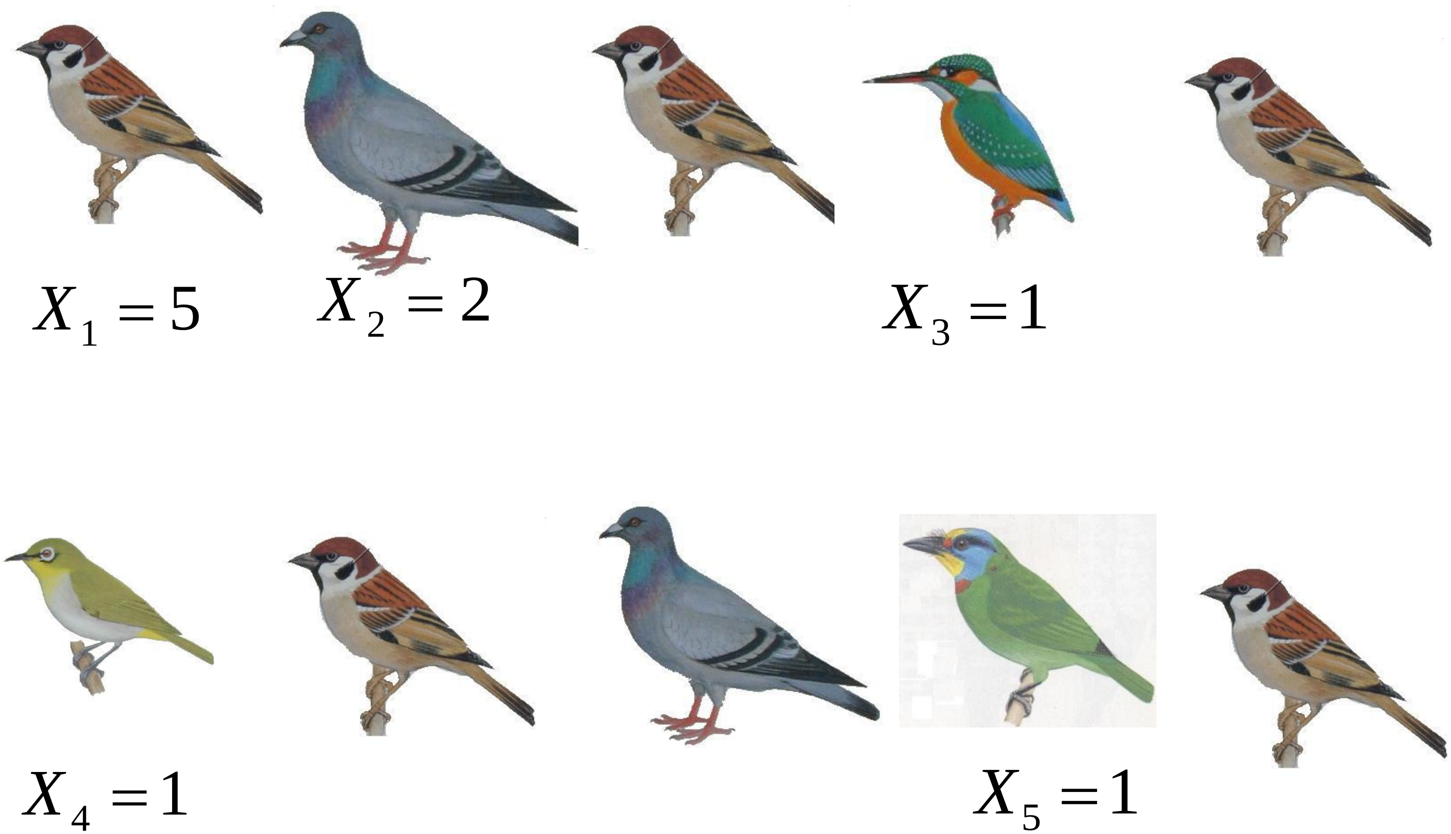


這只是未來之先兆；將來一切之縮影

二戰時解碼的統計工作

- 圖靈戰時解碼的統計工作自己沒有出版
- 他同意由他的統計助理 I.J. （Jack） Good 出版
- Good （1953） and Good and Toulmin （1956） 出版在 Biometrika 的兩篇論文是主要的工作
- 此兩篇中以類似生物調查的術語或作者的文字庫的術語而非密碼術語出版

A simple example: n =10



$f_1 = 3$

$f_2 = 1$

$f_3 = 0$

$f_4 = 0$

$f_5 = 1$

X_i：第i個物種在樣本中出現次數

樣本比例（資料中的比例)和母體(或整體)真正比例差多少?





5/10?　2/10?　1/10?　1/10?　1/10??

樣本中出現 *r* 次之物種其在母體中之比例大約為何? 樣本比例 = *r*/*n*

- 樣本比例加總為1, 表示沒有未看到的物種
- 如果全部種類已看完

  樣本比例~母體真正比例

如果尚有未看到的物種（表示樣本資料不完整）

- 樣本比例必定高估了母體真正比例
- 母體真正比例如何利用樣本資料來估計?

Good-Turing 頻率理論

樣本中出現 r 次 ($r = 0, 1, \ldots$) 之物種其在母體中之機率的平均為何?

$$\alpha_r = \sum_{i=1}^{S} p_i \, I(X_i = r) / f_r$$

$I(A) = 1$ 若 A 發生, 否則 $I(A) = 0$.

X_i: 第i個物種在樣本中出現次數

f_r: 資料中出現 r 次的物種數

出乎意料:Good-Turing 頻率理論

- 樣本中出現 r 次 $(r = 0, 1, \ldots)$ 之物種其在母體中之機率的平均之估計量並非 r/n, 估計量應為

$$\tilde{\alpha}_r = \frac{(r+1)}{n}\frac{f_{r+1}}{f_r} \equiv \frac{r^*}{n}$$

- 估計量應為 r^*/n, 其中

$$r^* = (r+1) f_{r+1} / f_r.$$

- 文獻上有許多理論的佐證

Good-Turing 頻率理論

- 出現一次之物種其在母體中之機率的平均並非1/*n*, 而是

$$\frac{(2f_2/f_1)}{n}$$

- 出現兩次之物種其在母體中之機率的平均並非2/*n*, 而是

$$\frac{(3f_3/f_2)}{n}$$

- 出現0次(沒有出現在樣本)之物種其在母體中之機率的平均並非 0/*n*, 而是

$$\frac{(f_1/f_0)}{n}$$

從Good-Turing 頻率理論估計未出現的物種數?

- 出現一次之物種其在母體中之機率的平均應比那些沒出現的為高

$$f_1/f_0 \leq 2f_2/f_1$$

- 可得未出現物種數之下界

$$f_0 \geq (f_1)^2 / (2 f_2)$$

Chao1 估計量

- 當出現一次之物種機率平均和沒出現的物種的平均大約相同, Chao1即為不偏 (Chao et al. 2017)
(稀有物種的機率大致相同時即可滿足)

出現0次(沒有出現)之物種其在母體中之機率的平均

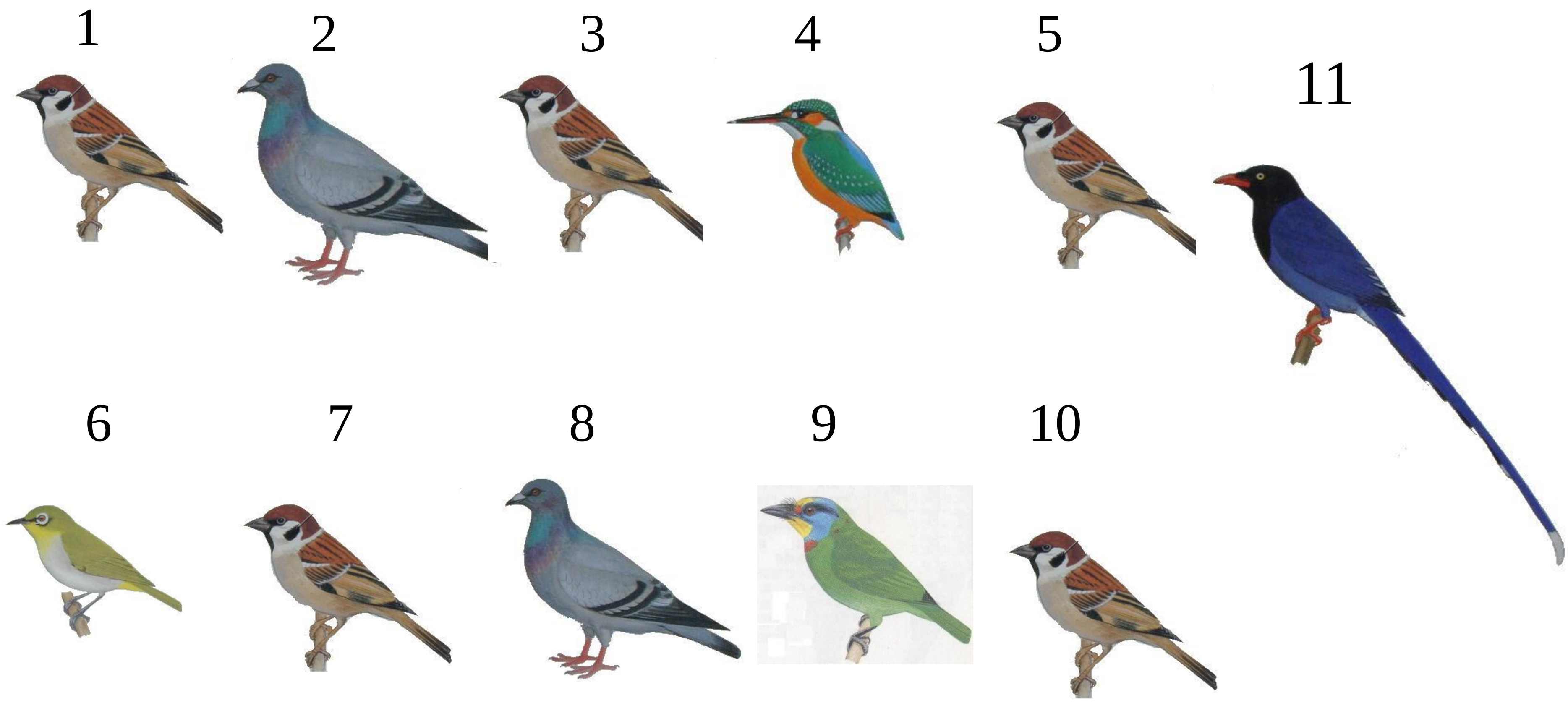
$$\tilde{\alpha}_0 = \frac{(f_1 / f_0)}{n} \quad \Rightarrow \quad est(\alpha_0 f_0) = f_1 / n$$

$$\alpha_0 = \sum_{i=1}^{S} p_i I(X_i = 0) / f_0$$

$$\Rightarrow \quad \alpha_0 f_0 = \sum_{i=1}^{S} p_i I(X_i = 0)$$

樣本中沒出現之物種其在母體中機率的總和之估計量即為資料中出現一次的比例

直觀就可了解(r = 0)



樣本涵蓋 或 樣本完整度
Sample coverage

$$C = \sum_{i=1}^{S} p_i I[X_i > 0]$$

樣本中出現之物種其在母體中機率的總和

C 估計量即為

$$\hat{C} = 1 - est(\alpha_0 f_0) = 1 - f_1 / n$$

如何標準化樣本?

- 傳統生態學的研究中，一般都是採用標準化樣本數(但樣本數是自己主觀選取的)

- 相同的樣本數對一個物種繁多的生物群落可能只觀察到一小部份的物種，但對一個物種稀少的生物群落可能已看光所有物種

- 2012提出將樣本涵蓋用於標準化做為比較物種數之基礎; 2014 推廣至物種多樣性（樣本涵蓋是資料估計之客觀的樣本完整程度）

標準化以比較物種多樣性

- 物種數並無考慮物種豐富度
- 且統計上估計困難
- 多年來生態界對如何量化物種多樣性的諸多爭辯目前已有共識:利用有效物種數曲線

Hill指標族 (order q)有效物種數

$$^{q}D = \left(\sum_{i=1}^{S} p_i^q \right)^{1/(1-q)}$$

MacArthur 1965,
Hill 1973

- $q = 0$, ^{0}D =物種數
- q =1, ^{1}D =熵指標指數:有效豐富種

$$^{1}D = \exp\left(-\sum_{i=1}^{S} p_i \log p_i \right)$$

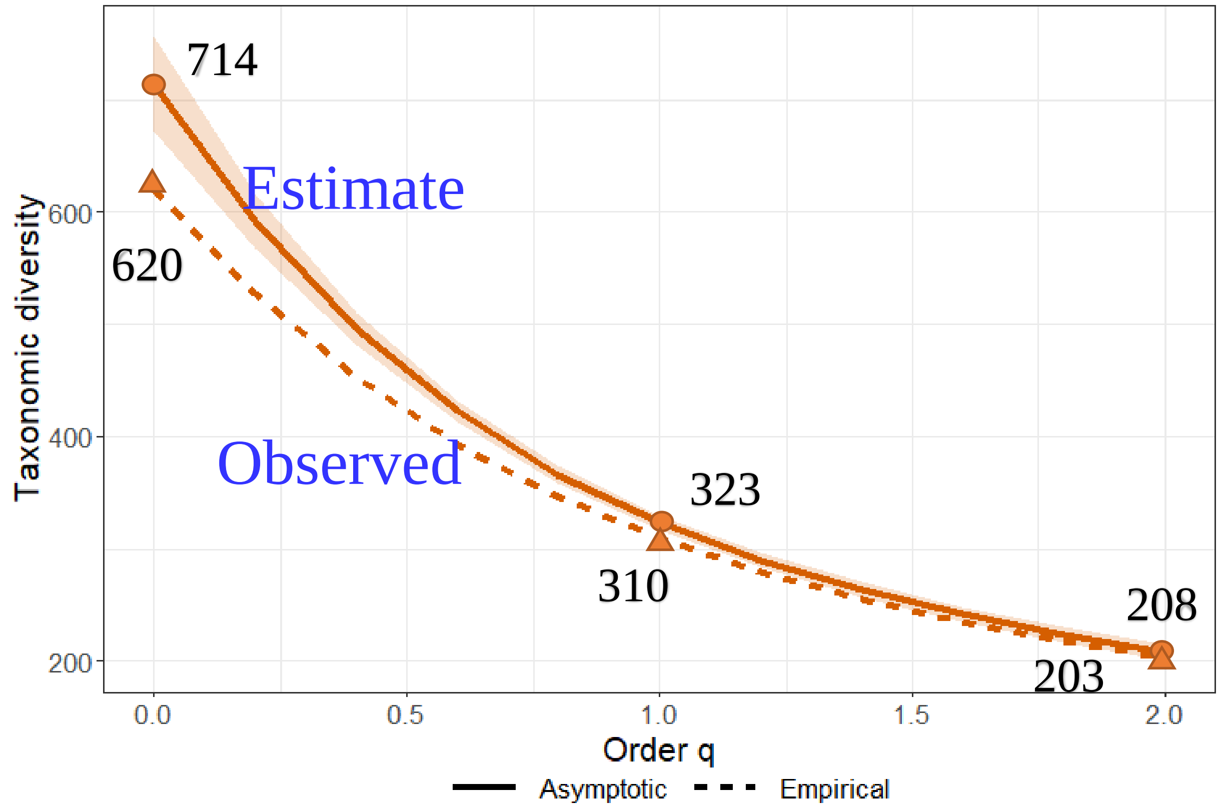
- $q = 2$, ^{2}D = Simpson 倒數 = 1/重複指標:

$$^{2}D = 1/\sum_{i=1}^{S} p_i^2$$

有效優勢種

利用統計模型所得有效物種數曲線估計:

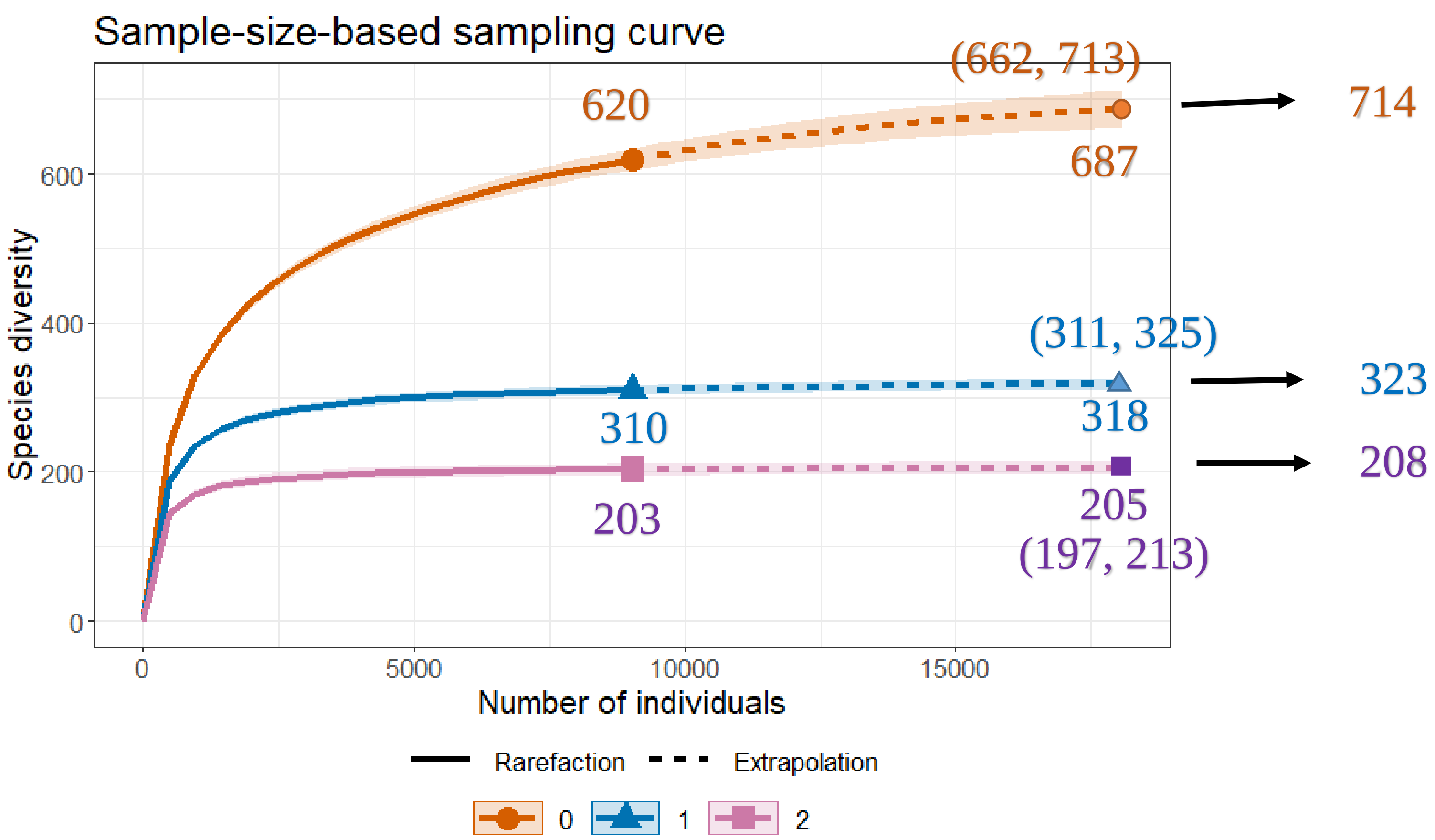
假若在馬來亞時間夠久,大約有多少種蝴蝶呢？
ANS: 至少 714種, 95% 信賴區間--[680, 770]



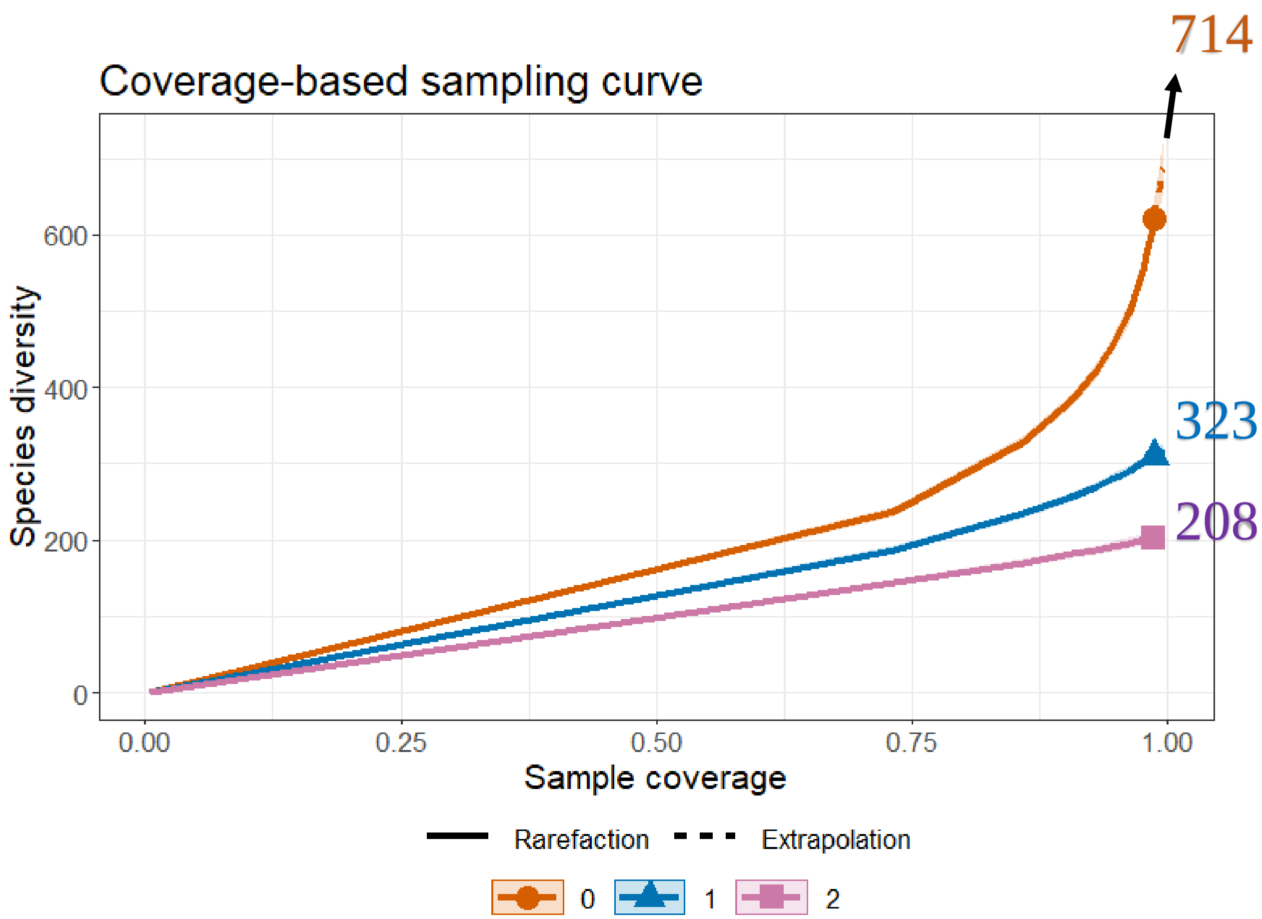
物種數為下界時必須用稀釋與外插曲線
(多樣性累積曲線之估計)

iNEXT 套件 (interpolation-extrapolation) Hsieh et al. (2016) 標準化樣本數之稀釋與外插方法

假若再去馬來亞多看兩年會再多看到幾種沒看過的呢??
ANS: 67, 95% 信賴區間--[42, 93]



"樣本涵蓋或完整程度"標準化樣本之稀釋與外插方法 iNEXT套件output



推廣至生物多樣性其它層次

- 物種(分類)多樣性指標：僅考慮物種數目與物種豐富度或均勻的程度
- 系統演化多樣性指標：除了考量物種數目及豐富度外，並加入物種**分類或演化親緣的關係**
- 功能性(生態系)指標：除考量物種數目及豐富度外，並加入物種在**生態系中的特質**，以了解生態系的穩定程度
- 食物鏈或網路多樣性

台灣墾丁森林樣區物種多樣性 Only species richness and abundances involved (DBH ≤ 2cm)黃心柿和鐵色最多

- 2008 census (68 species)
- 2013 census (58 species)

		2008	2013	2008	2013
Species code	Species	Raw abundance		Relative abundance	
DIOSMA	*Diospyros maritima*	6178	2386	0.63047	0.49750
DRYPLI	*Drypetes littoralis*	789	595	0.08052	0.12406
AGLAFO	*Aglaia formosana*	594	431	0.06062	0.08987
CRYPCO	*Cryptocarya concinna*	428	308	0.04368	0.06422
BEILER	*Beilschmiedia erythrophloia*	205	138	0.02092	0.02877
DIOSPH	*Diospyros philippensis*	172	167	0.01214	0.02023
CHAMMA	*Champereia manillana*	130	58	0.01327	0.01209
GONOCA	*Gonocaryum calleryanum*	119	78	0.01214	0.01626
DIOSER	*Diospyros eriantha*	119	97	0.01755	0.03482
GLYCCI	*Glycosmis citrifolia*	103	86	0.01051	0.01793
DENDME	*Dendrocnide meyeniana*	82	18	0.00837	0.00375
DRACAN	*Dracaena angustifolia*	71	25	0.00725	0.00521
BOEHWA	*Boehmeria wattersii*	68	37	0.00694	0.00771
PALAFO	*Palaquium formosanum*	61	34	0.00623	0.00709
MELAMU	*Melanolepis multiglandulosa*	60	19	0.00612	0.00396
CROTCA	*Croton cascarilloides*	53	41	0.00541	0.00855
LIODFO	*Liodendron formosanum*	48	32	0.00490	0.00667
ARDISI	*Ardisia sieboldii*	47	41	0.00480	0.00855
MALLPH	*Mallotus philippensis*	40	22	0.00408	0.00459
ANTIPE	*Antidesma pentandrum var. barbatum*	31	18	0.00316	0.00375

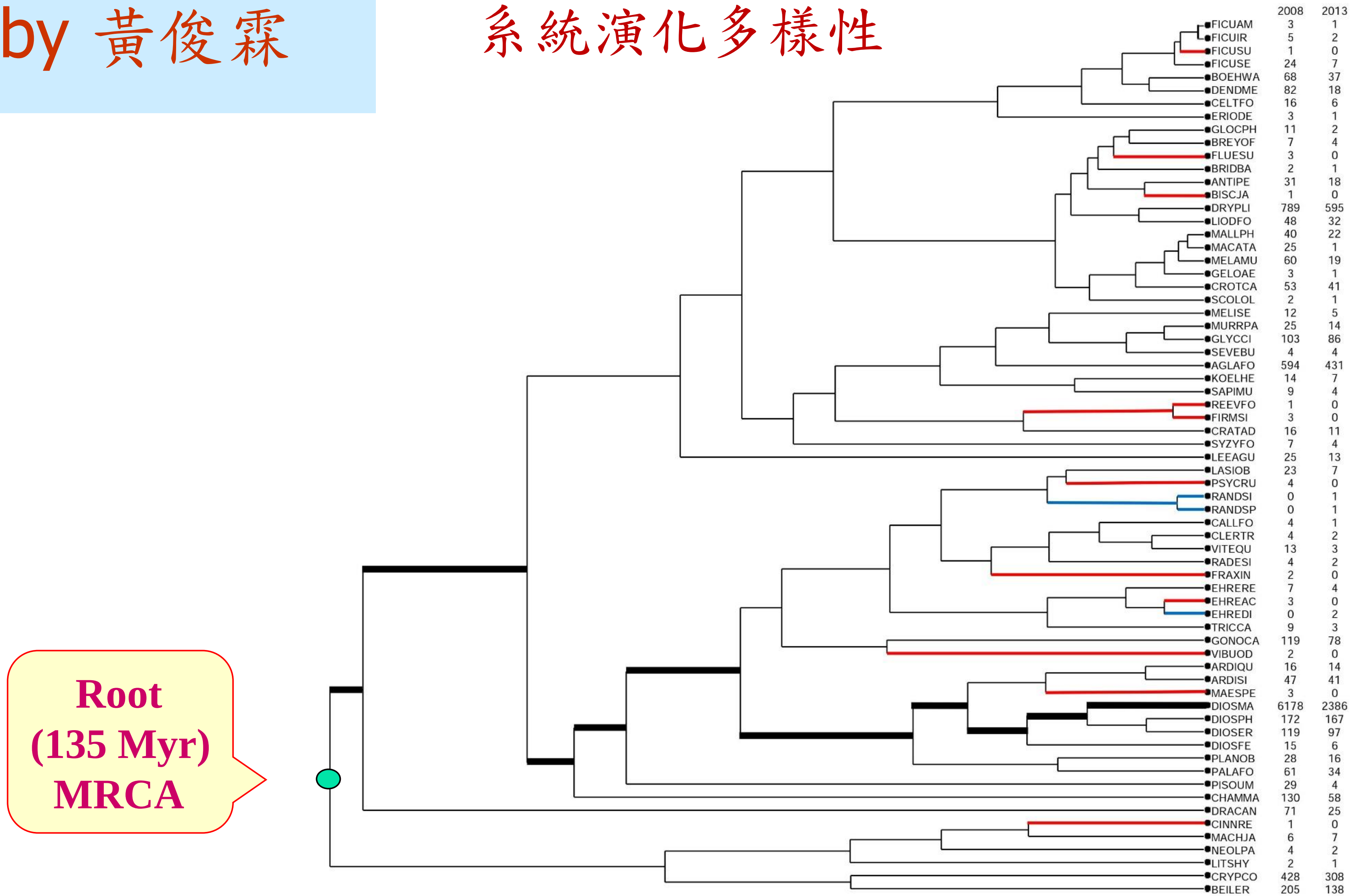
PISOUM	*Pisonia umbellifera*	29	4	0.00296	0.00083
PLANOB	*Planchonella obovata*	28	16	0.00286	0.00334
MURRPA	*Murraya paniculata*	25	14	0.00255	0.00292
MACATA	*Macaranga tanarius*	25	1	0.00255	0.00021
LEEAGU	*Leea guineensis*	25	13	0.00255	0.00271
FICUSE	*Ficus septica*	24	7	0.00245	0.00146
LASIOB	*Lasianthus obliquinervis*	23	7	0.00235	0.00146
CRATAD	*Crateva adansonii subsp. formosensis*	16	11	0.00163	0.00229
ARDIQU	*Ardisia quinquegona*	16	14	0.00163	0.00292
CELTFO	*Celtis formosana*	16	6	0.00163	0.00125
DIOSFE	*Diospyros ferrea*	15	6	0.00153	0.00125
KOELHE	*Koelreuteria henryi*	14	7	0.00143	0.00146
VITEQU	*Vitex quinata*	13	3	0.00133	0.00063
MELISE	*Melicope semecarpifolia*	12	5	0.00122	0.00104
GLOCPH	*Glochidion philippicum*	11	2	0.00112	0.00042
SAPIMU	*Sapindus mukorossii*	9	4	0.00092	0.00083
TRICCA	*Trichodesma calycosum*	9	3	0.00092	0.00063
BREYOF	*Breynia officinalis*	7	4	0.00071	0.00083
SYZYFO	*Syzygium formosanum*	7	4	0.00071	0.00083
EHRERE	*Ehretia resinosa*	7	4	0.00071	0.00083
MACHJA	*Machilus japonica var. kusanoi*	6	7	0.00061	0.00146
FICUIR	*Ficus irisana*	5	2	0.00051	0.00042
RADESI	*Radermachera sinica*	4	2	0.00041	0.00042
CLERTR	*Clerodendrum trichotomum*	4	2	0.00041	0.00042
SEVEBU	*Severinia buxifolia*	4	4	0.00041	0.00083

PSYCRU	*Psychotria rubra*	4	0	0.00041	0.00000
NEOLPA	*Neolitsea parvigemma*	4	2	0.00041	0.00042
CALLFO	*Callicarpa formosana*	4	1	0.00041	0.00021
EHREAC	*Ehretia acuminata*	3	0	0.00031	0.00000
FICUAM	*Ficus ampelas*	3	1	0.00031	0.00021
GELOAE	*Gelonium aequoreum*	3	1	0.00031	0.00021
ERIODE	*Eriobotrya deflexa*	3	1	0.00031	0.00021
FIRMSI	*Firmiana simplex*	3	0	0.00031	0.00000
MAESPE	*Maesa perlaria var formosana*	3	0	0.00031	0.00000
FLUESU	*Flueggea suffruticosa*	3	0	0.00031	0.00000
LITSHY	*Litsea hypophaea*	2	1	0.00020	0.00021
FRAXIN	*Fraxinus insularis*	2	0	0.00020	0.00000
SCOLOL	*Scolopia oldhamii*	2	1	0.00020	0.00021
BRIDBA	*Bridelia balansae*	2	1	0.00020	0.00021
VIBUOD	*Viburnum odoratissimum*	2	0	0.00020	0.00000
BISCJA	*Bischofia javanica*	1	0	0.00010	0.00000
CINNRE	*Cinnamomum reticulatum*	1	0	0.00010	0.00000
FICUSU	*Ficus superba var. japonica*	1	0	0.00010	0.00000
REEVFO	*Reevesia formosana*	1	0	0.00010	0.00000
EHREDI	*Ehretia dicksonii*	0	2	0.00000	0.00042
RANDSI	*Randia sinensis*	0	1	0.00000	0.00021
RANDSP	*Randia spinosa*	0	1	0.00000	0.00021

Phylogenetic diversity: adding evolutionary history

by 黃俊霖

系統演化多樣性



Functional diversity 功能性多樣性: based on species traits (特質)

和葉子有關的測量值

1. Thickness (葉厚度)
2. LMA (比葉重) (leaf mass per area, 單位葉面積的葉片乾重)
3. LDMC (葉乾物質含量) (leaf dry-matter content, 反映植物利用資源的能力)(%)

結論

- 幾乎所有的生物多樣性定量及評估都是根據抽樣資料，因此統計抽樣、統計模型及推論扮演重要的分析角色

- 目前生物體之間交互作用(食物鏈或網路多樣性),生態系之功能與永續穩定,以及如何比較不同時間與空間的生物多樣性資料都需統計方法

- 借用王羲之<蘭亭集序>之語:

仰觀宇宙之大，俯察品類之盛。所以遊目騁懷，足以極視聽之娛，信可樂也。


感謝聆聽
並祝福大家