

## 書目計量學

### Lecture 05 -- 洛特卡經驗法則

陳光華

國立台灣大學圖書資訊學系  
國立台灣師範大學社會教育學系  
khchen@ntu.edu.tw

## 大綱

洛特卡與學術生產率

洛特卡經驗法則的形成

洛特卡經驗法則的基本原理

洛特卡經驗法則的發展

洛特卡經驗法則的應用與評估

## 洛特卡與學術生產力

- 洛氏經驗法則的創始人-洛特卡
  - 美國的著名學者，擅長於統計分析
  - 發表《科學生產力的頻率分布》等著名論文，創立聞名的“洛特卡經驗法則”。
- 科學生產力概念
  - 衡量一門學科的發展
    - 這門學科中所發表的文獻
    - 研究者的著述規律及其與科學文獻之間的數量關係
  - 學術生產力之意
    - 研究者在學術上所表現出的能力和工作效果，通常用其生產的學術文獻之數量來衡量。
    - 衡量學術生產能力的一項定量指標。

## 洛特卡經驗法則的形成

- 學術生產率的統計研究
  - 計量研究對象
    - 選用化學和物理學兩大學科內的科學家及其著作
      - 《化學文摘》、《物理史料》
  - 統計數據
    - 與論文數量相對應的實際著者的統計數據
    - 著者頻率，即相應的著者數問著者總數的百分比
- 洛氏經驗法則的確立
  - 在科學領域裡，論文的作者頻率與論文數量之間存在著一定的關係
    - 學術生產率的倒平方律

## 洛特卡經驗法則的基本原理

- 洛氏經驗法則的基本內容
- 洛氏經驗法則的圖象描述

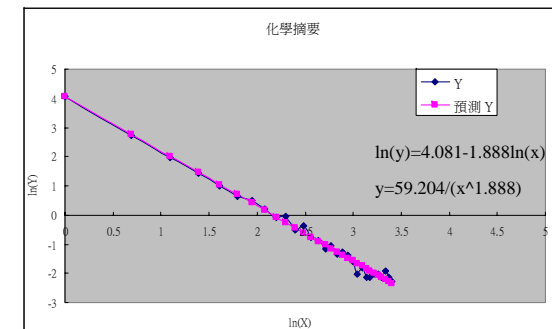
## 分析數據 (續)

著作數量(x)	化學著者數量	物理著者數量	化學著者數量百分比(x)	物理著者數量百分比(x)	著作ln(x)	著者ln(y)	著者ln(z)
17	18	3	0.26	0.23	2.833213344	-1.33207281	-1.48236183
18	19	1	0.28	0.08	2.890371758	-1.27800559	-2.58097412
19	17	0	0.25	0.00	2.944438979	-1.38923122	
20	14	0	0.21	0.00	2.995732274	-1.58338724	
21	9	1	0.13	0.08	3.044522438	-2.02521999	
22	11	3	0.16	0.23	3.091042453	-1.82454929	
23	8	0	0.12	0.00	3.135494216	-2.14300302	
24	8	3	0.12	0.23	3.17805383	-2.14300302	
25	9	2	0.13	0.15	3.218875825	-2.02521999	
26	9	0	0.13	0.00	3.258096538	-2.02521999	
27	8	1	0.12	0.08	3.295836866	-2.14300302	
28	10	0	0.15	0.00	3.33220451	-1.91985947	
29	8	0	0.12	0.00	3.36729583	-2.14300302	
30	7	1	0.10	0.08	3.401197382	-2.27653442	
合計	6820	1321	100.00	100.00			

## 分析數據

著作數量(x)	化學著者數量	物理著者數量	化學著者數量百分比(x)	物理著者數量百分比(x)	著作ln(x)	著者ln(y)	著者ln(z)
1	3991	784	58.52	59.35	0	4.06935254	4.083434902
2	1059	204	15.53	15.44	0.693147181	2.742635781	2.737145875
3	493	127	7.23	9.61	1.098612289	1.978064609	2.263212968
4	287	50	4.21	3.79	1.386294361	1.437037651	1.331048887
5	184	33	2.70	2.50	1.609437912	0.992491193	0.915533443
6	131	28	1.92	2.12	1.791759469	0.652752758	0.751230392
7	113	19	1.66	1.44	1.945910149	0.504943254	0.363464861
8	85	19	1.25	1.44	2.079441542	0.220206692	0.363464861
9	64	6	0.94	0.45	2.197224577	-0.06356148	-0.78921465
10	65	7	0.95	0.53	2.302585093	-0.04805729	-0.63506397
11	41	6	0.60	0.45	2.397895273	-0.5088725	-0.78921465
12	47	7	0.69	0.53	2.48490665	-0.37229696	-0.63506397
13	32	4	0.47	0.30	2.564949357	-0.75670866	-1.19467976
14	28	4	0.41	0.30	2.63905733	-0.89024005	-1.19467976
15	21	5	0.31	0.38	2.708050201	-1.17792213	-0.97153621
16	24	3	0.35	0.23	2.772588722	-1.04439073	-1.48236183

## 迴歸分析



## 洛氏經驗法則的基本內容

- 揭示作者比率與文獻數量之間的關係，描述科學生產率的頻率分布規律

$$f(x) = c/x^n$$

$x: 1, 2, 3, \dots, m$

$f(x)$ : 寫了 $x$ 篇論文的作者數站所有作者數的百分比

$$f_x \cdot x^n = c$$

( $c$ 為某主題領域的特徵常數)

- 科學生產率的“倒平方律”，洛特卡統計的數據，大約 $n=2$ 。

$$\rightarrow f(x) = \frac{c}{x^2}$$

經過推導，可得 $C$ 為 $f(1)$

$y(x)$ : 寫了 $x$ 篇論文的作者數

$y(1)$ : 寫了 $1$ 篇論文的作者數

$$\rightarrow f(x) = \frac{f(1)}{x^2}$$

兩邊同乘統計的作者總數

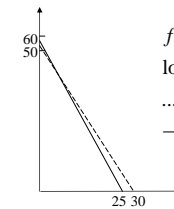
$$\rightarrow y(x) = \frac{y(1)}{x^2}$$

## 洛氏經驗法則的圖象描述

- 利用圖像描述了作者與論文之間的數量對應關係

■  $x$ 軸：作者所寫的論文數

■  $y$ 軸：寫了 $x$ 篇論文的作者比率



$$f_x \cdot x^n = c$$

$$\log f_x + n \log x = \log c$$

.....與圖中直線等價的數學表達式

→ 洛特卡定律之實質是  $\log f_x$  與  $\log x$  之間具有直線關係

## 洛氏經驗法則的文字描述

$$y(x) = \frac{y(1)}{x^2}$$

- 如果經統計得知某一學科領域內撰寫論文的作者數量，那麼就很容易計算出寫過2篇、3篇...論文的作者數量

■ 寫2篇論文的作者數量大約是寫1篇論文之作者數的 $1/4$  ( $1/2^2$ )

■ 寫3篇論文的作者數量大約是寫1篇論文之作者數的 $1/9$  ( $1/3^2$ )

■ 寫 $k$ 篇論文的作者數量大約是寫1篇論文之作者數的 $1/k^2$

## 洛特卡經驗法則的發展

- 洛氏經驗法則研究的進展
- 普賴斯經驗法則

## 洛氏經驗法則研究的進展

- 洛氏經驗法則研究的主要三個方面
  - 洛氏分布一般公式的研究
  - 洛氏經驗法則適用性的研究
  - 洛氏分布機制的研究

## 洛氏分布機制的研究

- 學科特徵
  - 理論科學領域，洛特卡經驗法則 $n=2$ 是合理的。
  - 技術科學、社會科學和人文學， $n$ 值會增大（多產學者比例降低）。
  - 規模較大和科研合作程度較高的學科， $n$ 值會變小。
- 統計條件
  - 統計研究的時間跨度和作者數量
  - 統計的時間較長，作者集合較大，會得到較客觀的應有結論。
- 研究方法
  - 對合著者和高產作者群之不同處理會影響其研究結果
  - 合著者和截刪高產作者數據是兩個必須解決的問題

## 洛氏經驗法則的研究

- 洛氏分布一般公式的研究
  - 如何確定  $n$  和  $c$  的數值便成了洛氏分布數據研究的重要任務和關鍵步驟
    - 洛氏經驗法則侷限於指數 ( $n$ ) 等於2的情形
    - Vlachy的發現
      - $n = 2$ 為一特例。 $n$ 的取值要受到資料量的多寡、學科的性質和發展程度等因素之影響。
    - 不同的 $n$ 值將產生一個顯著不同的常數 $c$ ，而且 $n$ 值的較小變化（特別是在 $n < 2$ 時）就會引起 $c$ 值的明顯變化。
- 洛氏經驗法則適用性的研究
  - 一般認為，在一定統計條件下，洛氏經驗法則在大多數學科領域是適用的，能夠描述科學文獻作者分布規律和科學家著述的行為模式。

## 普賴斯經驗法則

- 學者總人數開平方，所得到的人數撰寫了全部科學論文的50%。

$sx$  is total number of papers,

$x_{\max}$  is the maximum for  $x$

$$\frac{1}{2} sx(1, x_{\max}) = sx(1, m) = sx(m+1, x_{\max})$$

- 寫 $m$ 篇以上論文的學者

$sy$  is the total number of authors

$$(sy(1, x_{\max}))^{1/2} = sy(m+1, x_{\max})$$

## 普賴斯經驗法則之數學推導

$$sx(1, m) = \sum_{i=1}^m i * y(i) = \sum_{i=1}^m i * \frac{C}{i^n}$$

Assume  $n = 2$

$$sx(1, m) = \sum_{i=1}^m \frac{C}{i} = C * \sum_{i=1}^m \frac{1}{i} \cong C * (\ln(m) + 0.577)$$

$$sx(1, x_{\max}) \cong C * (\ln(x_{\max}) + 0.577)$$

$$sx(1, x_{\max}) / 2 = sx(1, m)$$

$$2 \ln(m) = \ln(x_{\max}) - 0.577$$

$$\frac{1}{2} (\ln(x_{\max}) + 0.577) = \ln(m) + 0.577$$

$$\ln\left(\frac{x_{\max}}{m^2}\right) = 0.577$$

$$\frac{x_{\max}}{m^2} = e^{0.577} = 1.781 \Rightarrow m^2 = 0.562 * x_{\max}$$

$$m = 0.749 \sqrt{x_{\max}} = 0.749 (x_{\max})^{\frac{1}{2}}$$

## 普賴斯經驗法則之重要推論

- 發表了  $0.749(\sqrt{x_{\max}})$  篇以上論文的學者們所發表的論文總數等於全部論文總數的一半
- 傑出學者中最低產那位學者所發表的論文數，等於最高產學者發表論文數的平方根之 0.749 倍
- 傑出學者佔全體學者之比例為

$$0.812 / \sqrt{x_{\max}}$$

## 洛特卡經驗法則的應用與評估

- 洛氏經驗法則的應用
- 洛氏經驗法則的侷限及原因
- 洛氏經驗法則適用性研究與鑑定

## 洛氏經驗法則的基本應用

- 科研人員的著述狀況和發展論文的數量
  - 測定具有不同數量論文的作者之分布規律
  - 揭示科學工作者生產科學文獻的能力
- 分析每一篇論文的作者數量
  - 闡明完成一篇科學論文要“占用”多少作者

## 洛氏經驗法則的具體應用

- 在資訊科學、圖書館學方面
  - 測定各學科的“平均文獻作者數”，合理編制著者索引、規劃檢索刊物體系之參考價值和指導意義。
- 在預測科學方面
  - 預測文獻數量的增長速度和文獻流的動向
  - 預測科學家數量的增長和科學發展的規模
- 在科學學和人才學方面
  - 科學家的活動規律、研究人材的著述特徵，便於科學學的理论研究和科學史的探討。

## 洛氏經驗法則適用性研究與鑑定

- 科爾提出使用K-S檢驗法 (Kolmogorov-Smirnov)
  - 彌補直觀統計帶來的侷限，合理地評價洛氏經驗法則對其學科的適用性
  - 步驟
    - 計算K-S值
      - $\frac{1.63}{\sqrt{y}}$  (y為成對樣本數)
    - 找出最大偏差值D
      - $D = \text{Max} |F_n(x) - S_n(x)|$
      - $F_n(x)$ 為累積作者頻率的理論值； $S_n(x)$ 為累積作者頻率的觀察值
    - 比較D值K-S值之大小
      - $D < K$ -S值，則抽樣分布符合洛氏經驗法則
      - $D > K$ -S值，不符合洛氏經驗法則

## 洛氏經驗法則的侷限及原因

- 與研究的學科領域有關
- 經驗法則本身的侷限
  - 洛特卡的倒平方經驗法則只是對兩個學科領域抽樣的簡單推廣而建立起來的一個通式，並非一精確的統計分析
- 統計數據的侷限
  - 研究時間跨度夠長，作者數目大才能夠近似地表示為倒平方律。

## 洛氏經驗法則適用性研究與鑑定(續)

- 科爾提出的K-S檢驗法 (Kolmogorov-Smirnov) (續前頁)
  - 對洛特卡統計的二項抽樣數據進行鑑定
    - 《物理史料》統計的數據符合洛氏經驗法則
    - 《化學文摘》數據亦符合洛氏經驗法則
  - K-S檢驗法屬於非參數檢驗，其臨界值取決於觀測數據範圍的大小
    - 樣本愈大，檢驗的效果愈明顯
  - K-S檢驗法是現行的最有效之檢驗法

## K-S檢驗法實例

著作數量 (x)	化學著者數量	觀察值(%)	觀察值(累積%)	推估值(%)	推估值(累積%)	Diff	MaxDiff	K-S Value
1	3991	58.51906	58.51906	59.20400	59.20400	0.00685	0.01704	0.297
2	1059	15.52786	74.04692	15.99582	75.19982	0.01153		
3	493	7.22874	81.27566	7.43954	82.63936	0.01364		
4	287	4.20821	85.48387	4.32177	86.96113	0.01477		
5	184	2.69795	88.18182	2.83593	89.79706	0.01615		
6	131	1.92082	90.10264	2.01003	91.80708	0.01704	K-S=1.63/sqrt(#PairedSample)	
7	113	1.65689	91.75953	1.50247	93.30956	0.01550		
8	85	1.24633	93.00587	1.16766	94.47722	0.01471		
9	64	0.93842	93.94428	0.93485	95.41207	0.01468		
10	65	0.95308	94.89736	0.76622	96.17828	0.01281		
11	41	0.60117	95.49853	0.64003	96.81831	0.01320		
12	47	0.68915	96.18768	0.54307	97.36138	0.01174		
13	32	0.46921	96.65689	0.46690	97.82829	0.01171		
14	28	0.41056	97.06745	0.40594	98.23423	0.01167		
15	21	0.30792	97.37537	0.35636	98.59059	0.01215		
16	24	0.35191	97.72727	0.31548	98.90607	0.01179		

## 結語與討論

## K-S檢驗法實例 (續)

著作數量 (x)	化學著者數量	觀察值(%)	觀察值(累積%)	推估值(%)	推估值(累積%)	Diff	MaxDiff	K-S Value
17	18	0.26393	97.99120	0.28136	99.18743	0.01196		
18	19	0.27859	98.26979	0.25258	99.44001	0.01170		
19	17	0.24927	98.51906	0.22807	99.66808	0.01149		
20	14	0.20528	98.72434	0.20702	99.87509	0.01151		
21	9	0.13196	98.85630	0.18880	100.06389	0.01208		
22	11	0.16129	99.01760	0.17292	100.23682	0.01219		
23	8	0.11730	99.13490	0.15900	100.39582	0.01261		
24	8	0.11730	99.25220	0.14673	100.54255	0.01290		
25	9	0.13196	99.38416	0.13584	100.67839	0.01294		
26	9	0.13196	99.51613	0.12615	100.80454	0.01288		
27	8	0.11730	99.63343	0.11747	100.92201	0.01289		
28	10	0.14663	99.78006	0.10968	101.03169	0.01252		
29	8	0.11730	99.89736	0.10265	101.13434	0.01237		
30	7	0.10264	100.00000	0.09628	101.23062	0.01231		
	6820							