

運輸需求模式軟體系統

交通部運輸研究所

中華民國七十六年六月

交通部運輸研究所出版品摘要表

出版品名稱 中文：運輸需求模式軟體系統 外文：			
行政機關出版品統一編號 09109760066		運輸研究所出版品編號 76-33-526	
本所計畫：沈 鏡 光 主 持 人 研究人員：		受委託單位： 計畫主持人： 研究人員：	
研究方式： <input checked="" type="checkbox"/> 自行辦理 — 主辦單位： <input type="checkbox"/> 委託辦理 — 受委託單位： 地 址： 聯絡電話：			研究期間 自 74年7月 至 76年6月
關鍵詞：需求模式、選擇模式、產生模式、重力模式、Logit 模式、Tobit 模式、Poisson 模式			
摘 要：本報告說明製作一個需求模式軟體系統中已完成部份的理論背景、計算方式。			
出版日期	頁數	工本費	本 出 版 品 取 得 方 式
76年7月	120	410	<input checked="" type="checkbox"/> 洽本所免費贈閱 <input type="checkbox"/> 洽本所訂購 <input type="checkbox"/> 其他 ()
管制等級 本出版品： <input type="checkbox"/> 機密 <input type="checkbox"/> 解密日期為 年 月 日 <input type="checkbox"/> 承辦單位視情況通知資料組解密 <input checked="" type="checkbox"/> 一般			本 表： <input type="checkbox"/> 機密 <input type="checkbox"/> 解密日期為 年 月 日 <input type="checkbox"/> 承辦單位視情況通知資料組解密 <input checked="" type="checkbox"/> 一般
備 註：			

目 錄

第一章 系統概要

1.1 前言.....	1
1.2 計畫內容.....	2
1.3 擬議中的模式系統.....	4

第二章 資料庫概要

2.1 前言.....	7
2.2 資料庫結構.....	7
2.3 資料轉換.....	11

第三章 選擇模式

3.1 前言.....	13
3.2 Newton 方法.....	14
3.3 偏好關係.....	17
3.4 基本選擇理論.....	20
3.5 Logit 模式.....	28
3.6 具有無差異結構的 Logit 模式.....	32
3.7 常態機率選擇模式 (probit 或 normit).....	36
3.8 消去選擇模式.....	39
3.9 範例.....	43
3.10 計算方式.....	49
3.11 整合性 Logit 模式.....	57

第四章 產生模式

4.1	前言	69
4.2	Tobin 模式	70
4.2.1	估計技巧	71
4.2.2	計算方式	73
4.2.3	計算係數微分	74
4.2.4	對變異數的微分	77
4.2.5	解題程序	79
4.3	Poisson 模式	83
4.3.1	估計技巧	84
4.3.2	計算方式	84
4.3.3	計算微分	85

第五章 分配模式

5.1	前言	87
5.2	雙比例平衡法	87
5.2.1	Furness 方法	89
5.2.2	一些理論上的重要結果	91
5.3	重力模式	96
5.3.1	計算方式	97
5.4	廣義重力模式	103
5.4.1	Gray-Sen 的理論	104
5.4.2	偏差	105
5.4.3	異質性	107
5.4.4	較小的 t_{ij}	109

5.4.5	估計方法.....	109
5.4.6	計算方法.....	110
第六章	結論.....	113

第一章 系統概要

1.1 前言

經歷許多年工作經驗與若干大、小型的規劃工作，我們發現運量預測總是（或多或少）不可或缺的一環，但是却也是最為困難的工作之一，因為要了解研究區域的旅行行為、了解所收集來資料的特性、依據這些大前提與預測時所可能獲取的預測資料、規劃目的與用途，而選擇一套合用、且合理的模式。但是，若無廣泛而且立即可用的計算機軟體，則在運量模型之建立與預測上就是一個相當嚴重的瓶頸；但是前已述及，模式與預測的成功並不單在有良好的軟體，而在於有精確翔實的輸入資料，以及對規劃區域、目的、用途，甚至於該區域中人們的旅行行為，以及未來所可能出現的變遷都要有相當的了解。但是，這一點却已經脫出本計畫的範圍太遠。因此，本計畫的根本並不在於如何運用需求模式來估計、預測運量，這是一個運用工具技巧上的問題，因此我們的重點在於有沒有工具可資運用？如果沒有工具，就設計出工具；如果工具不好使用，就改善工具；所以嚴格來說我們是在解決更為根本的問題。

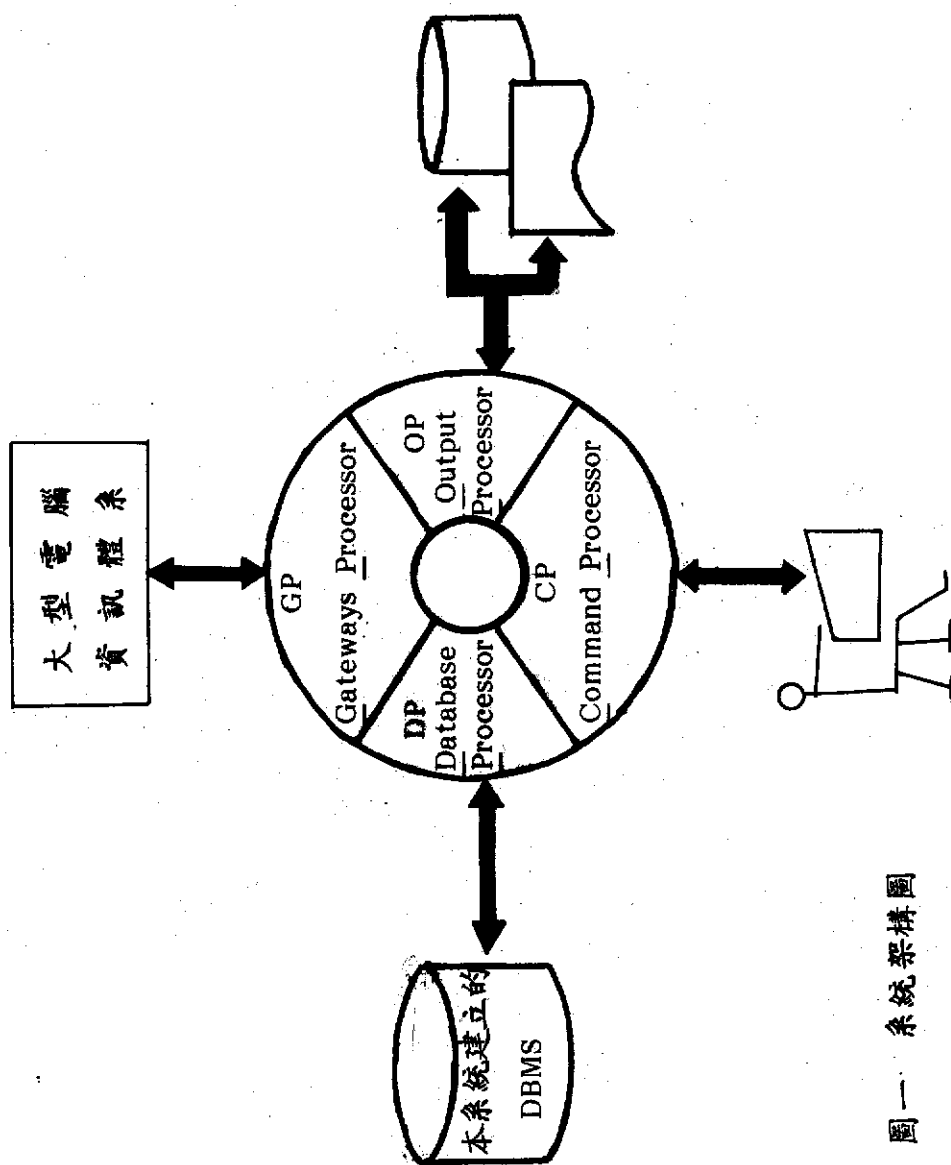
需求模式在建立、評估以及運用上其實本無定則可循，任何能夠使用的技術均可納入，但各種技術之間仍有差異，譬如，模式的立論是否與我們要解決的問題相符？模式對不良資料的容忍度有多少？對正確資訊是否足以反應出正確的行為？精確度如何？預測的技巧如何？有多少可供分析模式使用的資訊，所以設計一套軟體系統輔助模式的建立、評估、預測的工作就有此必要，此即本計畫由來。

1.2 計畫內容

在構想中，這是一個完整的體系，希望透過完整的軟體界面而降低使用人的負擔，但是因為這是一項龐大而複雜的工作，而且又是一人執行，所以採用下而上的設計，以逐步方式先行建立能夠使用的各個片段部份，再行整合成可以工作的單獨程式，最後再合併成大系統。

在概念上，本系統的架構如圖一所示，它並非完全獨立，而是透過界面程式與本所其它資料庫（如果有的話）取得連繫擷取資料；它本身也有一個獨特的資料庫管理系統（DBMS），有一個命令處理系統（CP），一個輸出管理程式（OP），而中心部份則是各個模式體系。這幾個系統的功能如下：

- (一)DBMS：以本系統所期望的型式來管理資訊。運輸資訊泰半以表格型式出現，每一「列」代表一筆資料，而「行」為該筆資料中的各個欄位。因而資料庫應有 JOIN, EXTRACT, MERGE 等動作。這一部份目前只在進行系統設計工作，程式寫作部份尚未開始。
- (二)命令處理系統（CP）：命令處理系統負責對所有來自使用人下達的命令的解析、轉譯、執行的工作，它把各道命令經分析後轉譯成內碼，交由核心部份執行。這些命令包括了檔案管理、資料轉換、估計模式、從事預測，以及其它各有關工作。目前，這一部份已測試完成 TRANSFER 的部份，但對於如何叫用或翻譯模式的規格則有待進一步修訂；主要問題是若參數太多，則使用人將無所適從，勢必像 SPSS 一樣常常因為一個小小錯誤而導致全部重新輸入。
- (三)輸出管理系統（OP）：目前尚未設計，所有輸出動作均直接由各模式完成；但在理想的情況下，模式的輸出、模式結果的儲存工作



圖一 系統架構圖

等，應統一交由此系統完成，俾能降低程式敘述的重覆部份。

(四)界面程式 (Gateways)：這一部完全沒有進行，主要原因是(1)這是與外界資料庫溝通的部份，但本所完全沒有此等資料庫；(2)由於(1)的關係，既然這些資料庫並不存在，則各個資料庫的規格為未知，所以亦無從進行。不過這一部份本來就不固定，會隨使用的系統、使用的語言、機型而異不能統一，所以應該是配合系統逐漸完成。

(五)模式系統的核心 (Kernel)：目前已部份完成若干個體模式系統，但是因為前段與後段均不甚完整，所以這些系統的輸入輸出部份全都是暫時性，為了測試程式之正確性而臨時寫作，並不完整，唯模式的計算部份則正確無誤。已完成的核心部份有 Logit，整合性 Logit，Tobit，Poisson 與線性混合有異質性模式；另外，有關重力模式部份亦已完成部份，但未能測試完成，效率可能不高。

1.3 擬議中的模式系統

預定包含在模式系統中的基礎部份有 (有星號者為未進行者)：

(一)無重覆選擇的 Logit。

(二)有重覆選擇的 Logit。

(三)整合性 Logit。

* (四)有次序性的 Logit。

* (五)層次式 (Nest) Logit。

* (六)廣義 Probit。

(七)Tobin 模式 (Tobit)。

(八)Poisson 模式。

(九)帶異質性線性模式。

* (十)聯立模式。

(±)負指數式重力模式。

(±)廣義重力模式。

第二章 資料庫概要

2.1 前言

如上一章所述，資料庫的部份目前僅具雛型，供初步測試之用。本章就討論資料庫的結構方式與大略的作業方法；此外，對資料庫中各項資料的轉換的部份亦已製作完成，所以一併在本章中概略地介紹這一部份的作業方式與觀念。

2.2 資料庫結構

運輸資料泰半以表格型狀出現，一列就是一筆觀測值，比如人口、所得、旅次產生數……等等，一行則是某個資料項目在不同樣本中的值。為了處理方便起見，每一個磁碟中都可以由本系統建立一個母表 (Mother Table 簡稱 MT)，這張表記錄了在該磁碟中的各個檔案的各項屬性 (attribute)。MT 中各一個欄位的長度固定，可以看成是一個 RECORD 結構，內容大致如下：

RECORD

file_name	: 檔案基礎名稱
number_of_col	: 檔案中行數 (變數個數)
number_of_row	: 檔案中列數 (資料筆數)
C_date	: 建立檔案的日期
C_time	: 建立檔案的時間
m_date	: 上一次修飾檔案的日期
m_time	: 上一次修飾檔案的時間
Acc_ctrl	: 使用權控制, Read-Only ?
free	: 剩餘的行數

m_value	: 遺漏值記號
used_head	: 已使用行開頭
used_tail	: 已使用行結尾
free_head	: 未使用行開頭

END

對於每一個檔案而言，都有兩個分離的部份，第一部份為控制部（延伸為CTL），第二部份為資料部（延伸為.DAT）。因此，若某檔名為TRIP，則它的兩部份為TRIP.CTL與TRIP.DAT。控制部的結構為

ARRAY [1..number_of_col] OF COLUMN

而COLUMN的結構為

RECORD

used_next	: 下一個未使用的行
used_last	: 上一個未使用的行
col_name	: 該行的名稱
:	

END

used_next 與 used_last 為一雙向串列，指出已使用的各行；而在MT中的 free_head 則是指出未曾使用或已收回的各行，示意圖如次：



MT

ABC	4	2	1	3
XYZ	5	1	5	4

1	2	3	4
0	4	0	1
4	0		2
L	M		N

ABC.CTL

XYZ.CTL

1	2	3	4	5
3	0	5	2	0
0		4		3
AA		BB		CC

2.3 資料轉換

儲存在 .DAT 中的資料可以使用算術運算，邏輯運算，加上一些內建函數 (built-in) 函數加以轉換；在寫作運算式時，可以使用每一筆資料的名稱，但若同一名稱出現在若干已打開的檔案中，則可以用檔案名稱來限定。比如說，如果 INCOME 既出現在 SOCIAL 檔，也出現在 TRIP 檔，那麼 SOCIAL.INCOME 就是 SOCIAL 檔中的 INCOME，而 TRIP.INCOME 則是 TRIP 檔的 INCOME。

轉換資料的運算以 C 語言者為準，優先順序應與 C 語言相同。在四則運算方面有 +、-、*、/ 與正負號 (+、-)；比較運算為 >、>=、<、<=、=、<>；邏輯運算有 || (或) && (且) 與 ~ (非)；還有一個三元運算?:，亦即 $a + b > c ? x - y : x + y$ 中，當 $a + b > c$ 為真時，運算式值為 $x - y$ ，不然就是 $x + y$ 。設定運算為：=。

內建函數除了三角函數，反三角函數，指數，對數 (含 \log_2 與 \log_{10}) 之外，還有一些統計學常用的函數，如 Lag 函數等等。

資料轉換 (TRANSFER) 以一特定命令完成，整個部份透過一個 LALR(1) 分析程式，將輸入轉換成四元組 (Quadruple)，再加以解釋，並且更新資料檔。舉例而言，

costime := cost / time

就把目前 (唯一) 的資料檔中的 cost 被 time 去除，而產生一筆新資料，costime。有時候，我們可能並不期望產生新資料，於是在變數前方加上一 @，就可以表示這一筆資料是臨時性的，不必在資料檔中建立出來；譬如，

@costime := cost / time

Lncostime := Ln (@costime)

這個式子把 cost 被 time 去除之後的自然對數值存入資料檔中的

LnCostime。

再看一例，

@Best := cartime ;

@Best := @Best <= bustime ? @Best : bustime

Model.Besttime := @Best <= Railtime ? @Best Railtime

就是求出 cartime , bustime , Railtime 三者中最小的一個，存到 Model 檔中叫做 Besttime 的那一行。

第三章 選擇模式

3.1 前言

本計畫是從選擇模式開始的，而這些模式也事實上已經在若干大小型計劃中使用過，但是已往的程式均寫成單獨作業的型式，所以還需要經過整合、修訂之後才能放到本系統中。本章先討論個體選擇模式的基本概念，再討論到計算方法；在基本概念方面，我們提到 Logit, Probit 與消去式選擇三種模式，但後兩者並沒有寫成可用的軟體系統。另外，在大區域、大樣本時，整合性 Logit 模式亦有其地位，因此在本章末也就此等模式做一簡介。

這些模式體系絕大多數是非線性 (Non-Linear) 的，因此系統中就要有一套非線性的估計方法；目前，我們所製作的模式均有唯一的極值，因此傳統的 Newton-Ralphson 方法就已經足夠，而不必使用 Davidon-Fletcher 等需要在下降方向求極值的方法。為方便計，下一節就先介紹 Newton 方法。

3.2 Newton 方法

若 $f(x)$ 為一實值函數，於是 Taylor 展開式謂

$$f(x + \alpha \cdot \delta) = f(x) + \alpha \left(\frac{\partial f}{\partial x} \right)' \cdot \delta + \frac{1}{2} \alpha^2 \cdot \delta' \cdot \frac{\partial^2 f}{\partial x' \partial x} \cdot \delta + \dots$$

如果忽略二次以後各項，且 x 與 δ 固定，則

$$\begin{aligned} \varphi(\alpha) &= f(x + \alpha \delta) - f(x) \\ &= \alpha \cdot \left(\frac{\partial f}{\partial x} \right)' \delta + \frac{1}{2} \alpha^2 \delta' \left(\frac{\partial^2 f}{\partial x' \partial x} \right) \delta \end{aligned}$$

我們要先選一個方向 δ ，使 $\varphi(\alpha) < 0$ ，亦即 $f(x)$ 沿 δ （至少在 x 附近）為遞降，再求一 α 使得下降幅度為最大。

當二次微分為一正定 (Positive Definite) 矩陣時，確定可以下降的方

$$\delta = - \left(\frac{\partial^2 f}{\partial x' \partial x} \right)^{-1} \left(\frac{\partial f}{\partial x} \right)$$

而使下降幅度有最大的 α ，則是 $\varphi(\alpha)$ 的極小值，但是這會很花費時間，所以只求下降到足夠大即得。

下面就是一個化簡後的 Newton 程序

程序：Newton (ϵ, x, f_x)

輸入：

ϵ : 容許誤差

x : x 的一個初值，做為極小點的一個猜測。

輸出：

x : 近似的極小點

f_x : 近似極小值

begin

repeat

計算 $H := \frac{\partial^2 f}{\partial x' \partial x}$ 與 $g := \frac{\partial f}{\partial x}$;

$\delta := -H^{-1} g$;

$f_x := f(x)$;

$z := x + \delta$; $f_z := f(z)$;

if $f_z < f_x$ then

begin

repeat { 向前找更小的值 }

$s := f_z$;

$\delta := \delta + \delta$;

$z := x + \delta$; $f_z := f(z)$;

until $f_z > f_x$;

$x := x + \frac{1}{2} \delta$; $f_x := s$;

end

else

begin

repeat { 往回找更小的值 }

$\delta := \frac{1}{2} \delta$;

$z := x + \delta$; $f_z := f(z)$;

until $f_z < f_x$;

$x := z$; $f_x := f_z$;

end ;

until $\|\delta\| < \varepsilon$ AND $\|g\| < \varepsilon$
end

3.3 偏好關係

設 $A \subseteq R^n$ [1] 是一個集合，它的元素可能是無限多（目前我們不打算加上限制），而 R 是定義在 A 上的一個偏序（partial order），亦即 R 滿足：

(i) 自反律：對於每一個 A 中的元素 x 而言，則 $x R x$ ；

(ii) 傳遞律：若 $x R y$ ， $y R z$ ，則 $x R z$ 。

一般而言，偏好（preference）的理論就是一個偏序的關係。如果用 \geq 表示上述的 R ，則 $x \geq y$ 的意義就是說 x ， y 之間我們較偏好 x ，因此 $x \leq y$ 就表示偏於 y ，所以可以構作出兩個集合。

$$A_x^+ = \{ x' \in A \mid x' \geq x \},$$

$$A_x^- = \{ x' \in A \mid x' \leq x \}.$$

上面的 A_x^+ ， A_x^- 分別表示比 x 好或比 x 不好的集合，令

$$A_x^\sim = A_x^+ \cap A_x^-$$

因此 A_x^\sim 就是與 x 沒有差異（indifference）元素的集合，為了方便起見，若 $u, v \in A_x^\sim$ ，則記成 $u \approx v$ 。自然 A_x^\sim 不可能為空集合，由自反律 $x \in A_x^+$ ， $x \in A_x^-$ ，因此 $x \in A_x^\sim$ 。

因為 \geq 是一個定義在 A 集合上的關係，它只是一種集合間的對應，而無法從事量的計算，因此欲要使得 \geq 有用，則必須得經過一個量化的過程。如果把 A 想成一組方案（alternative），而由某個體從

1. R^n 表示歐基里德 n 度空間。

這一組方案中選擇其中的一個，直覺上，他必然會選擇他偏好度最高的一個，因此就此一個體而言，他對於A集合中諸元素定出了一個偏序，那麼我們是否可以找出一個指數 (index) 來，使得

$$x \geq y \Leftrightarrow U_x \geq U_y \quad (\forall x, y \in A)$$

上式中左邊的 \geq 是一個偏序，而右邊的 \geq 却是實數的大小關係（當然也是一個偏序），如果這個 U_x, U_y 存在，則量化的過程就完成了！

慶幸的是 Debreu [1] 與 Rader [2] 分別在 1954, 1963 完成了這件工作，為了解說 Debreu 的定理我們先得了解下述的定義：

定義：若 \geq 為定義在集合A上的一個偏序，則 \geq 為連續 (continuous) 的充分必要條件為對於每一個 $x \in A$ 而言， A_x^+ 與 A_x^- 均為閉集合 [3] (closed Set)。

定理 (Debreu)：若 $A \subseteq R^n$ ，且A為連結 (connect)，又 \geq 為定義在A上的一個連續偏序，則必定存在一個實值連續函數 (real

1. Debreu, G., Representation of Preference Ordering by a Numerical Function, in Decision Processes, ed. by Thrall, Coombs, and Davis, New York, Wiley, 1954, pp. 159-165.
Debreu, G., Theory of Value, New York, Wiley, 1959.
2. Rader, T., Existence of a Utility Function to Represent Preferences, Review of Economic Studies, Vol. 30(1963).
3. 如果 $x_1, x_2, \dots, x_n, \dots$ 為A中之任一序列，若 $\lim_{n \rightarrow \infty} x_n = x^*$ ，且 $x^* \in A$ ，則說A是一個閉集合。

continuous) $U : A \rightarrow R$, 使得

若 $x, y \in A$, 且 $x \geq y$, 則 $U(x) \geq U(y)$

Debreu 的定理是一個廣泛的，抽象的定理，但是，如果 A 中的元素若為有限 (finite) 個，則 Debreu 定理是派不上用場的，簡單的數學就可證明 U 函數的存在性。有了 Debreu 的定理之後我們就可以把偏好的問題轉移到一個數值的問題了；在經濟學上，這個 $U(\cdot)$ 函數常常稱為效用 (utility) 函數，不過本文為了語意上的問題不用這個名詞，我們只需了解到 U 只是 \geq 在實數上的一個表現 (representation) 那就足夠了，所以我們就把 $U(\cdot)$ 稱作 \geq 的表現函數 (representation function)。

另外，值得一提的是 $U(\cdot)$ 的存在而非唯一，比如說，若 V 為一連續的單調上升 (monotonic increasing) 函數 (即 $V : R \rightarrow R$, 且 $x \leq y$, 則 $V(x) \leq V(y)$) , 則我們極易證明 $V \circ U(\cdot)$ 也是 Debreu 定理中所需要的函數：若 $p, q \in A$, 且 $p \geq q$, 則 $U(p) \geq U(q)$, 於是 $V \circ U(p) = V(U(p)) \geq V(U(q)) = V \circ U(q)$ 。

3.4 基本選擇理論

就理論的導引來分，隨機選擇模型可以約略分成三個範疇，第一類叫做固定表現模型，由 Luce, Suppes 等人提出；第二類叫做隨機表現模型，McFadden 1973 年的論文可以看作集大成之作；第三類是由 Tversky 領導而發展出來的消去選擇模型，今分述於后。

對於某一個體 i 而言，他會把自己所面對的各個方案構成的集合 A 給上一個偏好的關係， \geq_i ；如果 A 的元素個數有限，則 \geq_i 的表現函數之存在性自無任何疑問。假設此一表現函數為 $U_i(x)$ ，此地 $x \in A$ 。首先我們假設 $U_i(x)$ 不為隨機變數，第 i 個體選擇 $x \in A$ 之機率記成 $P_i(x, A)$ ，於是固定表現模型的理論就是： $P_i(x, A)$ 為各個 $U_i(x)$ 的函數，亦即

$$P_i(x, A) = F(U_i(x), U_i(y), \dots) \quad \forall x, y \in A$$

然而在隨機選擇模型中，我們假設 $U_i(x)$ 為一隨機變數，利用偏好的理論，我們可以看出

$$\begin{aligned} P_i(x, A) &= \text{prob} [x \geq_i y, \forall y \in A, x \neq y] \\ &= \text{prob} [U_i(x) \geq U_i(y), \forall y \in A, x \neq y] \end{aligned}$$

如果 A 之元素個數有限，則 A 中之元素可以給上一個編號，設此編號為 $1, 2, 3, \dots, J$ ；再把隨機表現函數 $U_i(x)$ 分解成兩部份，即

$$U_{ij} = V_{ij} + \epsilon_{ij}$$

此地

U_{ij} ：第 i 個體的表現函數，對第 j 個方案的函數值。

V_{ij} : 第 i 個體的表現函數，對第 j 個方案的函數值的非隨機部份。

ϵ_{ij} : 第 i 個體的表現函數，對第 j 個方案函數值的隨機部份。

因此

$$\begin{aligned} P_i(x, A) &= \text{prob} [U_{ix} \geq U_{ij}, V_j \in A, j \neq x] \\ &= \text{prob} [V_{ix} + \epsilon_{ij} \geq V_{ij} + \epsilon_{ij}, V_j \in A, j \neq x] \\ &= \text{prob} [\epsilon_{ij} - \epsilon_{ij} \leq V_{ix} - V_{ij}, V_j \in A, j \neq x] \\ &= \text{prob} [\epsilon_{ij} - \epsilon_{ij} \leq W_{i,xj}, V_j \in A, j \neq x] \end{aligned}$$

上式中 $W_{i,xj} = V_{ix} - V_{ij}$ 。

如果適當地給於 ϵ_{ij} 一個機率分配 (density 或 cumulative)，則 $P_i(x, A)$ 不難自理論的觀點計算出來，這就是目前在運輸界大行其道的觀念。

至於 Tversky 的理論當另闢章節討論，此地先不多說。

前曾述及，表現函數經過一個連續單調上昇函數的轉換後仍舊是一個表現函數，因為指數函數 $\exp(\cdot)$ 為連續，且單調上昇，因此 $\exp(U) = e^U$ 亦為一表現函數，最重要的是 $\exp(U) > 0$ ，因此在以下的討論中我們不妨假設 $U > 0$ 。就固定表現模型而言，

$$P_i(x, A) = \frac{U_{ix}}{\sum_{j \in A} U_{ij}}$$

無疑是最簡單的，也是最容易接受的，就是 Luce“百分比”型的模型。從這個式子中可以看出，當 U_{ix} 增加，而其他的 U_{ij} 不變時 P_i (

$x, A)$ 增加；而 U_{ix} 不變，其他任一 U_{ij} 增加時， $P_i(x, A)$ 就會減少，Krantz 在 1964 年觀察到這個現象，他叫做完全純量性 (simple scalability)，其正確定義如下：

定義：設 A 中有 J 個元素，對於第 i 個體而言，選擇在 A 中的元素 x 之機率設為

$$P_i(x, A) = F_A(U_{i1}, U_{i2}, U_{ix}, \dots, U_{is})$$

且 F_A ，即 $P_i(x, A)$ 對 U_{ix} 而言，為一增函數，對所有 $j \neq x$ 而言為減函數，則此一選擇機率我們稱作滿足完全純量性，以下為方便計記成 SS。

接著討論由 Tversky 在 1972 年提出來的次序獨立性 (order independence) [1] 觀念。設 A, B 已知，而且對某一個體 i 而言； $P_i(x, A) \geq P_i(y, A)$ ，此地 $x, y \in A - B (\neq \emptyset)$ ；若 $z \in B$ ，我們考慮 $P_i(z, B^U\{x\})$ 與 $P_i(z, B^U\{y\})$ 的大小關係：如果不計較 x, y 與 A, B 的影響，則因 x 在 A 中之強度超過 y ，於是在 B 中亦然，因此 z 在 $B^U\{x\}$ 中的強度就小於在 $B^U\{y\}$ 中的強度，是故

$$P_i(z, B^U\{x\}) \leq P_i(z, B^U\{y\})$$

於是 Tversky 給出如次的定義

-
1. Tversky, A., Choice by Elimination, J. of Mathe. Psych., Vol. 9(1972), pp. 341-367.

定義：若 $x, y \in A - B$, $z \in B$, 且以下之機率都不為 0 或 1, 如果

$$P_i(x, A) \geq P_i(x, B) \Leftrightarrow P_i(z, B \cup \{x\}) \leq P_i(y, \{x, y\})$$

則說此一選擇機率具有次序獨立性, 簡記成 OI。

於是 Tversky 證明了下述定理。

定理 $SS \Leftrightarrow OI$.

接著, 我們討論不相干方案的獨立性 (independence from irrelevant alternative), 簡記成 IIA。設 $x, y \in A$, 且 $P_i(x, \{x, y\})$, $P_i(x, A)$, $P_i(y, A)$ 均不為 0, 則 IIA 性質就是

$$\frac{P_i(x, \{x, y\})}{P_i(y, \{x, y\})} = \frac{P_i(x, A)}{P_i(y, A)}$$

上式說明了 x, y 之間二元選擇 (binary choice) 機率的比與其所在之集合無關, 亦即是與 x, y 無關的方案 ($z \in A - \{x, y\}$) 不構成對上式的影響。下面的討論為簡便計, 把 $P_i(x, \{x, y\})$ 記成 $P_i(x, y)$ 。

首先, 我們把上式改寫成

$$P_i(y, A) = \frac{P_i(y, x)}{P_i(x, y)} \cdot P_i(x, A)$$

則

$$1 = \sum_{y \in A} P_i(y, A) = \left(\sum_{y \in A} \frac{P_i(y, x)}{P_i(x, y)} \right) \cdot P_i(x, A),$$

故得

$$P_i(x, A) = \frac{1}{\sum_{y \in A} \frac{P_i(y, x)}{P_i(x, y)}}$$

令

$$V_{yx} = \frac{P_i(y, x)}{P_i(x, y)}$$

我們可以輕易算得

$$V_{yx} = \frac{V_{yp}}{V_{xp}} \quad p \in A - \{x, y\}$$

若 A 之元素有限，且把 V_{ab} 當成未知數，則對於 $a, b \in A$ 而言，上面型式的式子構成一個降級 (degenerate) 的二次聯立方程式，因此知道：對於每一 $p \in A$ 而言，都存在一個數 U_p ，使得當 $x, y \in A$ 時

$$V_{xy} = \frac{U_x}{U_y}$$

最後我們就得到了

$$P_i(x, A) = \frac{U_{ix}}{\sum_{y \in A} U_{iy}}$$

上式中 U_{ix} 中的 i 表示個體的編號，這就是有名的 Luce 模型，很顯然的，Luce 模型滿足 SS 條件，亦即

$$IIA \Leftrightarrow SS \Leftrightarrow OI$$

事實上 IIA 的條件要比 OI 來得強，IIA 要求的是機率比在不同的方案集合中為常數，而 OI (亦即 SS) 只是很單純地指出 $P_i(x, A)$, $P_i(y, A)$, $P_i(z, B \cup \{x\})$, $P_i(z, B \cup \{y\})$ 之間的大小次序而已！

Luce 提出 IIA 特性自然有他獨到之處，由前述之 Luce 模型，若 $A \subseteq B$ ，則

$$P_i(x, A) = \frac{U_{ix}}{\sum_{j \in A} U_{ij}}, \quad P_i(x, B) = \frac{U_{ix}}{\sum_{j \in B} U_{ij}}$$

因此我們只需研究在 A 中的選擇特性就可以了，至於在 B 中的特性可以自 A 中的性質直接引導出來；又因為 $IIA \Leftrightarrow OI$ ，我們還可以知道， x 在 A 中的優勢在 B 中也有決定的影響。在應用上 IIA 性質無疑是相當有用的假設，但是並非所有的選擇行為都滿足 IIA 特性，假設我們要購置一張唱片，不過得在三張已知的唱片中加以選擇，若第一張為莫札特的朱彼得交響曲，第二、三張均為貝多芬的命運交響曲，不過指揮與樂團不同；如果我們對於莫札特與貝多芬的作品同樣喜好，則選擇莫札特與貝多芬的機率各為 0.5，因此這三張唱片選擇機率各

為 0.5, 0.25, 0.25。如果加入 IIA 特性，則因為對於作者沒有偏愛，所以再因為對於樂團與指揮亦無執著，故

$$\frac{P(\text{莫札特})}{P(\text{貝多芬-1})} = \frac{P(\text{莫札特, 貝多芬-1})}{P(\text{貝多芬-1, 莫札特})} = 1,$$

$$\frac{P(\text{貝多芬-1})}{P(\text{貝多芬-2})} = \frac{P(\text{貝多芬-1, 貝多芬-2})}{P(\text{貝多芬-2, 貝多芬-1})} = 1$$

由上面兩個式子得到

$$P(\text{莫札特}) = P(\text{貝多芬-1}) = P(\text{貝多芬-2}) = \frac{1}{3},$$

顯然地 IIA 性質導致了一個矛盾，與事實不符的結果！

我們再考慮另一例子，假有兩家百貨公司 A 與 B，它們同時提供了兩種不同的用品，為方便計，令 A_1, A_2 表示 A 公司的第一，第二種產品， B_1, B_2 表示 B 公司的第一，二種產品，如果我們對公司沒有成見，而對產品亦復沒有偏好，令 $\Sigma = \{A_1, A_2, B_1, B_2\}$ ，則得

$$P(A_1, \Sigma) = P(A_2, \Sigma) = P(B_1, \Sigma) = P(B_2, \Sigma) = \frac{1}{4}$$

我們考慮 $\Delta = \{A_1, B_1, B_2\}$ (其他的情況相同)，因為對公司沒有成見，則得

$$P(A_1, \Delta) = \frac{1}{2}$$

$$P(B_1, \Delta) + P(B_2, \Delta) = \frac{1}{2}$$

但因對貨品沒有偏愛，故得

$$P(B_1, \Delta) = P(B_2, \Delta) = \frac{1}{4}$$

如果OI成立，利用OI的定義，我們卻得

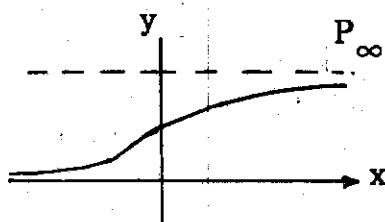
$$P(A_1, \Delta) = P(B_1, \Delta) = P(B_2, \Delta) = \frac{1}{3}$$

又是矛盾，因此IIA性質也無法解釋這個現象。故我們得到一個結論；不管IIA，還是SS或OI雖然有其優點，但是它的優點却還是其缺點，因為很多直觀上正確的現象却無法解釋。

3.5 Logit 模式

這是一個最常見到與用到，而且也被討論得最多的模型，在進入正題前我們得先討論一些有名的事實。1845 年，Verhulst 在研究人口成長的時候，他發現這條曲線是這樣的（見圖）

$$y = \frac{P_{\infty}}{1 + e^{a-bx}}$$



上式中 P_{∞} ， a ， $b > 0$ ， P_{∞} 就是人口成長的極限。Verhulst 就把這條曲線（毫無理由的）叫做推理曲線（logistic curve，這個譯名是隨台大葉樹藩教授的文章而來），如果我們不考慮人口的概念，則

$$y = \frac{1}{1 + e^{a-bx}}$$

於是它就構成一個機率分配函數。

接著是 Gnedenko 的一個定理。在次序統計學（order statistics）中，我們自某一機率分配中抽出 n 個樣本，並且把它以某一大小順序（不一定是數的大小）排列出來，設為 $X_{(1)}$ ， $X_{(2)}$ ，……， $X_{(n)}$ ，我們問 $n \rightarrow \infty$ 時 $X_{(n)}$ 的分配是什麼？也就是說一組樣本中最大的那一個樣本的分配是什麼？Gnedenko 證明了這樣的一個定理：

定理 (Gnedanko)

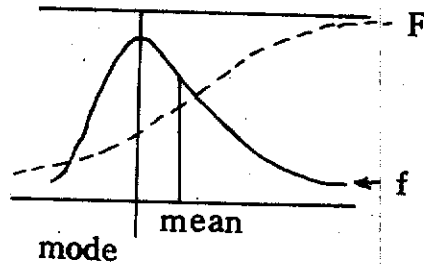
若 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 為某一機率分配中之 n 個隨機樣本，又 $n \rightarrow \infty$ 時 $X_{(n)}$ 的機率分配存在，則此一機率分配必為下列三者之一（與原始的機率分配無關）：

$$\begin{aligned} \Lambda_1(x) &= \exp(-\exp(-x)) & -\infty < x < \infty \\ \Lambda_2(x) &= \begin{cases} \exp(-(-x)^\alpha) & x \leq 0, \alpha > 0 \\ 0 & x > 0 \end{cases} \\ \Lambda_3(x) &= \begin{cases} 0 & x > 0 \\ \exp(-x^{-\alpha}) & x \leq 0, \alpha > 0 \end{cases} \end{aligned}$$

Gnedanko 的定理給出一個啓示，亦即是在 n 個樣本中之最大者 $X_{(n)}$ 之分配趨於三種型式相似的分配，然而，在選擇理論中我們常選各表現函數值中之最大者，因此在原理上二者是相吻合的，所以我們可以給予 ϵ_j 一個 Gnedanko 定理中的分配，但是那一個？一般而言（可以與迴歸理論之殘差項加以比較），我們比較喜歡採取第一類的標準型，叫做標準 Weibull 分配，即

$$\begin{aligned} \text{prob} [\epsilon \leq \omega] &= e^{-e^{-\omega}} \\ &= \exp[-\exp(-\omega)] \end{aligned}$$

它的平均數是 $0.577 \dots$ (Euler 常數)，變異數為 $\frac{\pi^2}{6}$ ，最重要的是眾數 (mode) 為零 (見下圖)，也就是說大部份參與選擇的個體沒有表現上的誤差。



設

$$F(y) = \exp[-\exp(-y)]$$

$$f(y) = F'(y) = \exp(-y) \cdot \exp[-\exp(-y)]$$

最後代入

$$P_i(x, A) = \int_{-\infty}^{\infty} f(\epsilon_x) \cdot \prod_{\substack{j \in A \\ j \neq x}} F(U_x - U_j + \epsilon_x) d\epsilon_x$$

$$= \int_{-\infty}^{\infty} \exp(-\epsilon_x) \cdot \exp(-\exp(-\epsilon_x))$$

$$\cdot \prod_{\substack{j \in A \\ j \neq x}} \exp(-\exp(-(U_x - U_j + \epsilon_x))) d\epsilon_x$$

$$= \int_{-\infty}^{\infty} \exp(-\epsilon_x) \cdot \exp[-\exp(-\epsilon_x)]$$

$$\cdot \left(1 + \sum_{\substack{j \in A \\ j \neq x}} \exp(-(U_x - U_j)) \right) d\epsilon_x$$

上式的積分可以用 $z = -e^{-\epsilon_x}$ 作變數轉換計算出來，它的結果是

$$P_i(x, A) = \frac{e^{U_x}}{\sum_{j \in A} e^{U_j}}$$

上式就是我們熟知的 Logit 模型了。

從上式，我們可以看出，它的式子與 Luce 模型相同，因此它也具有 IIA 特性，所以也連帶地，也滿足了 SS 與 OI 特性，所以雖然 Logit 模型簡單，却有一項（即 IIA）不太容易令人接受的特性。另外，若 A 中之元素只有兩個，亦即是二元性的選擇時，則

$$P_i(x, y) = \frac{1}{1 + e^{U_y - U_x}}, \quad P_i(y, x) = \frac{1}{1 + e^{U_x - U_y}}$$

這正好是一個推理曲線的機率分配。

3.6 具有無差異結構的Logit模式

在討論 IIA 特性的時候我們看出兩個特性非常近似的方案常會造成 IIA, SS 或 OI 不成立, 因此前述的 Logit 模型也就不成立了; 為了補救這一點缺陷, 我們可以引入無差異的結構。在實用上 $x \approx y$ 時 $U_x = U_y$ 的可能性是很小的, 因此可以比照 Zeeman 所引入的寬容空間 (tolerance space) 的觀點, 我們定義

$$x \approx y \Leftrightarrow |U_x - U_y| < \delta_{xy} \quad \forall x, y \in A$$

此地 δ_{xy} 是一個正的常數, 因此當 $x, y \in A$ 時, 兩者間的關係只有三種 $x > y$, $x \approx y$ 與 $x < y$ 。

若表現函數依舊是 $U_x + \epsilon_x$, 則

$$\begin{aligned} (1) \quad P_{i,xy} &= \text{prob} [x \approx y] \\ &= \text{prob} [| (U_x + \epsilon_x) - (U_y + \epsilon_y) | < \delta_{xy}] \\ &= \text{prob} [U_x - U_y - \delta_{xy} \leq \epsilon_y - \epsilon_x \leq U_x - U_y + \delta_{xy}] \\ &= \text{prob} [\epsilon_y - \epsilon_x \leq U_x - U_y + \delta_{xy}] - \text{prob} [\epsilon_x - \epsilon_y \leq U_y \\ &\quad - U_x + \delta_{xy}] \\ &= \frac{e^{U_x}}{e^{U_x} + e^{U_y - \delta_{xy}}} - \frac{e^{U_x}}{e^{U_x} + e^{U_y + \delta_{xy}}} \end{aligned}$$

這就是 x, y 之間無差異的機率。

利用上節的技巧, 我們計算出有差異結構的機率, 即

$$(2) \quad P_{i,x} = \text{prob} [x > y, \forall y \in A - \{x\}]$$

$$= \text{prob} [U_x + \epsilon_x > U_y + \epsilon_y + \delta_{xy}, V_y \in A - \{x\}]$$

上式是說， x 之表現函數值至少要比 y 之表現函數值大過 δ_{xy} 才能顯出 x, y 之間的差異，因此接上式

$$= \text{prob} [\epsilon_y - \epsilon_x < U_x - U_y - \delta_{xy}, V_y \in A - \{x\}]$$

$$= \frac{e^{U_x}}{e^{U_x} + \sum_{\substack{y \in A \\ y \neq x}} e^{U_y + \delta_{xy}}}$$

因為 $\delta_{xy} \geq 0$ ，所以我們不妨比較一下在 Logit 模型中的 $P_i(x, A)$ 與 $P_{i,xy}$ ；很明顯的

$$P_i(x, A) \geq P_{i,xy},$$

這是因為 $P_i(x, A)$ 之中含有無差異機率緣故。

接著再回頭研究一下 $P_{i,xy}$ 。我們再定出一個參數 θ_{xy} ， $0 \leq \theta_{xy} \leq 1$ ，亦即是一個比率，使得 $\theta_{xy} P_{i,xy}$ ， $(1 - \theta_{xy}) P_{i,xy}$ 這兩部份分別是當 x, y 無差異時選擇 x 或 y 的機率，因此很顯然地

$$(3) \quad P_i(x, A) = P_{i,x} + \sum_{\substack{y \in A \\ y \neq x}} \theta_{xy} P_{i,xy}$$

$$= \frac{e^{U_x}}{e^{U_x} + \sum_{\substack{y \in A \\ y \neq x}} e^{U_y + \delta_{xy}}} + \sum_{\substack{y \in A \\ y \neq x}} \theta_{xy} \cdot$$

$$\left[\frac{e^{U_x}}{e^{U_x} + e^{U_y - \delta_{xy}}} - \frac{e^{U_x}}{e^{U_x} + e^{U_y + \delta_{xy}}} \right]$$

上式中 $\theta_{qp} = 1 - \theta_{pq}$ 。這就是具有無差異結構的 Logit 模型。

很明顯地，(3)式不再滿足 IIA 性質，同時，因為各個 δ_{xy} 值的差異也會使得(3)式不滿足 SS (亦即 OI) 的條件，因此我們可以把(3)式看做是 Logit 模式的一個推廣。

不過，這一個新的模型的定義式是相當複雜的(也許，這可以看做一個消去了強假設條件—IIA 性質—所必須付出的代價)，首先，我們看到新引入了兩組參數 δ_{xy} 與 θ_{xy} ，把它寫成矩陣就是(注意，只當 $x \neq y$ 時 δ_{xy} 才有意義，而且 $\theta_{yx} = 1 - \theta_{xy}$)：

$$\Delta = \begin{bmatrix} 0 & \delta_{12} & \delta_{13} & \cdots & \delta_{1m} \\ & 0 & \delta_{23} & \cdots & \delta_{2m} \\ & & 0 & \cdots & \vdots \\ & & & \ddots & \delta_{m-1,m} \\ & & & & 0 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 1 & \theta_{11} & \theta_{12} & \cdots & \theta_{1m} \\ 1-\theta_{11} & 1 & \theta_{23} & \cdots & \theta_{2m} \\ 1-\theta_{12} & 1-\theta_{23} & 1 & \cdots & \theta_{3m} \\ \vdots & \vdots & & \ddots & \vdots \\ \vdots & & & & \theta_{m-1,m} \\ 1-\theta_{1m} & 1-\theta_{2m} & \cdots & \cdots & 1 \end{bmatrix}$$

上式是假設A中有 m 個元素寫成的， \triangle 中因為對稱（假設）的緣故，一共多了 $\frac{1}{2}m(m-1)$ 個元素，而 θ 亦然，於是就一共增加了 $m(m-1)$ 個參數，這對於 m 很大的時候是需要相當長的時間來估計的。不但如此，這些參數還有限制條件，即 $\delta_{xy} \geq 0$ ，而且 $1 \geq \theta_{xy} \geq 0$ ，因此又增加了估計手續上的複雜性；然而，技術上却是沒問題的。

最後，也是最重要的，不管 θ_{xy} 的值是什麼，當所有 δ_{xy} 都是零時(3)就退化成基本的推理機率模式，因此，我們可以用(3)式當做正確模式，估計之後再用顯著性檢定來測驗A中諸元素之間是否有無差異性，或者是(3)式根本就退化成基本的推理機率模式。

3.7 常態機率選擇模式(Probit或Normit)

接著，我們討論常態分配的情形。比照迴歸理論， ϵ_j 假設為常態分配是相等合理的，所以我們設 $E(\epsilon_j) = 0$ ($E(\cdot)$ 表示數学期望值)。則

$$P_i(x, A) = \int_{-\infty}^{W_{xi}} \cdots \int_{-\infty}^{W_{xm}} \phi(r_1, r_2, \cdots, r_m; 0; \Omega) dr_1 dr_2 \cdots dr_m$$

上式中 $\phi(\cdot, \cdot, \cdot)$ 為多元常態分配之密度函數， m 表示 A 集合元素個數， 0 表示期望值，而 $\Omega = [\omega_{ij}]$ 為其共變異矩陣 (covariance matrix)，則

$$\omega_{ij} = E(\epsilon_i \epsilon_j) + E(\epsilon_x^2) - E(\epsilon_i \epsilon_x) - E(\epsilon_j \epsilon_x)$$

又在積分中不含有 r_x 項。利用轉換公式可以把這個常態分配標準化；為了討論方便計，設 $x = 1$ ，則

$$P_i(1, A) = \int_{-\infty}^{W_{12}} \cdots \int_{-\infty}^{W_{1m}} \phi(r; 0; \Omega) dr_2 dr_3 \cdots dr_m$$

$$\omega_{ij} = E(\epsilon_i \epsilon_j) + E(\epsilon_1^2) - E(\epsilon_i \epsilon_1) - E(\epsilon_j \epsilon_1);$$

因此令

$$\lambda_{2j} = \omega_{2j} \quad \epsilon_2 = r_2$$

又當 $k=2, 3, \dots, m$ 時

$$\lambda_{ik} = \omega_{ik} - \sum_{j=2}^{i-1} \frac{\lambda_{ji} \lambda_{jk}}{\lambda_{jj}}$$

$$\epsilon_i = r_i \sum_{j=2}^{i-1} \frac{\lambda_{ji} r_j}{\lambda_{jj}}$$

於是諸 t_i 就變成為平均值為 0，變異數 λ_{ii} 的常態分配，最後把 t_i 標準化就得到了

$$(1) \quad P_i(1, A) = \int_{-\infty}^{W_{12}} \phi \left[\frac{t_2}{\sqrt{\lambda_{22}}} \right] \int_{-\infty}^{W_{13} - \frac{\lambda_{23}}{\lambda_{22}} t_2} \phi \left[\frac{t_3}{\sqrt{\lambda_{33}}} \right] \\ \dots \int_{-\infty}^{W_{1m} - \sum_{j=2}^{m-1} \frac{\lambda_{jm}}{\lambda_{jj}} t_j} \phi \left[\frac{t_m}{\sqrt{\lambda_{mm}}} \right] dt_m \dots dt_2$$

這就是所謂的常態機率分配模型。

若 A 中只有兩個元素，設為 1, 2，則

$$(2) \quad P_i(1, A) = \int_{-\infty}^{W_{12}} \phi \left[\frac{t_2}{\sqrt{\lambda_{22}}} \right] dt_2,$$

這個無限積分 (improper integral) 在統計上叫做誤差 (error) 函數，利用電子計算機可以計算到任意有效位數 (指在計算機能記憶的範圍內)，所以，(2) 式雖然沒有一個如同 Logit 模型一樣有一個簡單的數學式，但我們仍然認為是可以接受的。

再反觀 (1) 式，它不再具有單純的誤差函數的型式，不過就數值分

析 (numerical analysis) 的觀點而言，這個積分可以使用 Clark 近似值，所以並不困難，因此 常態 機率選擇模型雖然沒有一個明顯的數學表示，但是在計算上是沒有困難的，雖然精確度稍遜。

3.8 消去選擇模式

前述討論的模型是正面地處理問題，但 Tversky 幾經研究之後發現正面的技巧不若反面處理來得合理，舉例而言，我們進餐廳吃東西，如果不吃魚，則所有魚類的菜就不必考慮了，其次，如果不吃辣，則餘下的菜餚中辣的又被刪去；第三，考慮到價錢問題，多過兩佰元的不必考慮。因此，餘下可供選擇的菜餚就是那些不辣的價錢低於兩佰元，又不是魚的菜。如果繼續下去，就會剩下一些我們需要的菜了，這個過程 Tversky 叫做消去選擇模型。

當然，在每一次消去的過程中被我們刪去的方案自然不止一個，設所有的方案集合為 Δ ，而 $A \subseteq \Delta$ ，我們想要研究在 A 中選出 x ($x \in A$) 的機率，亦即 $P(x, A)$ 。反過來說，如果刪去 A 中的某些方案，則無異於是選擇了這些方案在 A 中的補集合 (complement set)，所以我們可以考慮這個機率， $Q_A(B)$ ，它表示在 A 中選擇 B ($B \subseteq A$) 的機率。設在 B 中選擇 x 的機率寫成 $P(x, B)$ ，用機率的知識得到

$$(1) \quad P(x, A) = \sum_{B \subseteq A} Q_A(B) \cdot P(x, B)$$

事實上，上式的 $Q_A(B)$ 不一定可以觀察到或利用數學算出，因此 Tversky 的下一步工作就在於如何在(1)中消去 $Q_A(B)$ 這一個不可知的機率項。

$$\text{假設 } B, C \subseteq A, \text{ 則 } P(B, A) = \sum_{y \in B} P(y, A), \text{ 利用 IIA}$$

特性，我們可以算出下面的等式

$$\frac{P(B, A)}{P(C, A)} = \frac{\sum_{y \in B} P(y, A)}{\sum_{z \in C} P(z, A)} = \frac{\sum_{y \in B} P(y, T)}{\sum_{z \in C} P(z, T)},$$

因為 $P(B, A) = Q_A(B)$, Tversky 把這個觀念加以擴張，亦即假設 $Q_A(B)$, $Q_A(C)$ 滿足

$$(2) \quad \frac{Q_A(B)}{Q_A(C)} = \frac{\sum_{B_j \cap A = B} Q_T(B_j)}{\sum_{C_i \cap A = C} Q_T(C_i)}.$$

注意上式，如果規定 B_j 都是單個元素的集合，則(2)就退化成 IIA 特性了，最重要的是(2)中右邊分式機率的和不完全包含了單個元素集合的選擇，亦即是說，各個方案之間的選擇性是彼此相關的。利用(2)式，Tversky 把(1)的機率轉化成一個表現函數的型式：

定理 (Tversky) :

若選擇模型如(1)，則(2)成立之充分必要條件為存在一個定義在 T 上的表現函數使得

$$(3) \quad P(x, A) = \frac{\sum_{B_j} U(B_j) P(x, A \cap B_j)}{\sum_{A_k \cap A \neq \emptyset} U(A_k)}$$

這個定理有幾點是值得討論的：

(i) U 是定義在整個 T 上的，所以對於每一個 $B \subseteq T$ 而言， $U(B)$

都有意義，但是在 Luce 模型中 $U(\cdot)$ 只定義在每一個方案上，因為只能談論 $U(x)$ ，所以 Tversky 模型較 Luce 模型來得廣泛。

(ii) 在 A 中選擇 x 的機率由在 A 中各包含了 x 的子集中選擇 x 之機率與各個表現函數所決定。

(iii) 因為 B_j 可以分成兩種情形，即 $B_j \subseteq A$ 或 $B_j \not\subseteq A$ ；就 A_k 而言也可以分成 $A_k \cap A \neq \emptyset$ ， A 與 $A \subseteq A_k$ 兩項，因此(3)式可以再寫成

$$P(x, A) = \frac{\sum_{A \not\subseteq B_j} U(B_j) \cdot P(x, A \cap B_j) + \sum_{A \subseteq B_j} U(B_j) \cdot P(x, A)}{\sum_{\substack{A_k \cap A \\ \neq \emptyset, A}} U(A_k) + \sum_{A \subseteq A_k} U(A_k)}$$

注意上式的分子包含了 $P(x, A)$ ，因此移項合併得到

$$(4) \quad P(x, A) = \frac{\sum_{A \not\subseteq B_j} U(B_j) P(x, A \cap B_j)}{\sum_{\substack{A_k \cap A \\ \neq \emptyset, A}} U(A_k)}$$

(iv) 如 (i) 中所述，若 $U(\cdot)$ 只定義在單一方案的集合上時亦即 $U(B_j) \neq 0$ 的充分必要條件為 B_j 中只有一個元素，則

$$P(x, A) = \frac{U(x)}{\sum_{y \in A} U(y)}$$

這就是有名的 Luce 模型。

(v) 令

$$T_{xy} = \{ B_j \mid x \in B_j, y \notin B_j \}$$

$$T_{yx} = \{ A_k \mid y \in A_k, x \notin A_k \}$$

如果 $A = \{ x, y \}$ ，則得 Restle 的二元選擇模型：

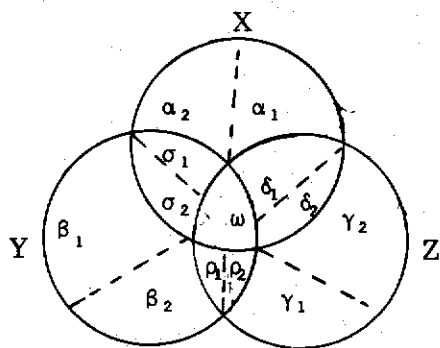
$$P(x, y) = \frac{\sum_{B_j \in T_{xy}} U(B_j)}{\sum_{B_j \in T_{xy}} U(B_j) + \sum_{A_k \in T_{yx}} U(A_k)}$$

從上面的五點討論來看，Tversky 選擇模型具有比較大的彈性，也具有比較廣泛的內含，Tversky 叫做消去選擇模型 (Elimination by Aspect)，我們就簡稱作 E B A 模型。

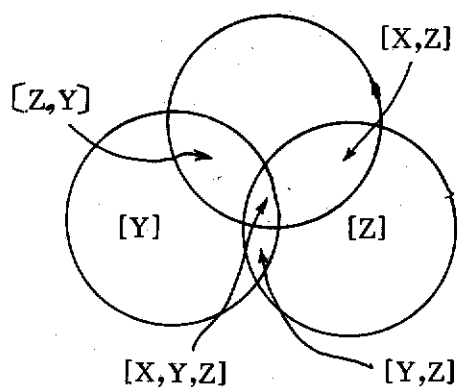
3.9 範例

我們考慮一個實際的問題。假設走進餐廳吃飯，則第一個問題就是點菜，通常菜單上五花八門，應該點那些？通常我們採取這個步驟：第一，兩佰元以上的不要，於是留下的菜都是兩佰元以下的，第二，如果不愛吃魚，則凡是魚類都不吃，因此餘下的是低於兩佰元，又不是魚類的菜；第三，不吃辣的，因此把辣的菜又全部刪掉了，……如此類推就得到最後我們所要的菜了！回顧這個過程，所有菜單上的菜都是我們所面對的方案，然後再把這些方案歸類成若干組，比如說，價錢大於兩佰元，魚類，辣的，……等等，然後把不能接受（或不適合）的方案刪去，如此以往，留下（亦即適合）的方案愈來愈少一直到只剩下一個（或一組）為止，則此一個（或一組）就是我們所選擇的方案了！

參看下圖，設 X, Y, Z 把十三個方案 $\alpha_i, \beta_i, \gamma_i, \sigma_i, \delta_i, \rho_i, \omega$ ($i = 1, 2$) 分成三組，亦即



(a)



(b)

$$X = \{ \alpha_1, \alpha_2, \sigma_1, \sigma_2, \delta_1, \delta_2, \omega \}$$

$$Y = \{ \beta_1, \beta_2, \sigma_1, \sigma_2, \rho_1, \rho_2, \omega \}$$

$$Z = \{ \tau_1, \tau_2, \delta_1, \delta_2, \rho_1, \rho_2, \omega \}$$

接著我們再引入幾個符號

$$[X] = \{ \text{只滿足} X \text{之條件，而不滿足} Y, Z \text{的方案} \} =$$

$$\{ \alpha_1, \alpha_2 \}$$

$$[X, Y] = \{ \text{同時滿足} X, Y \text{條件，而不滿足} Z \text{的方案} \}$$

$$= \{ \sigma_1, \sigma_2 \}$$

$$[X, Y, Z] = \{ \text{同時滿足} X, Y, Z \text{之條件的方案} \} = \{ \omega \}$$

其餘類推。因此所有的方案就被分割成七個相互分離 (disjoint) 的集合 (見上圖(b))，即 $[X]$ ， $[Y]$ ， $[Z]$ ， $[X, Y]$ ， $[X, Z]$ ， $[Y, Z]$ ，與 $[X, Y, Z]$ 。如果依前述消去的觀念來看，則不論刪掉的是 X ， Y ，或 Z 都會連帶地刪掉 $[X, Y, Z]$ ，因此 ω 這個方案在消去式的選擇方式中是沒有意義的 (注意，這個 ω 不一定存在)。

其次，我們研究一下如何選到 X 的問題。因為 X 中已經分成四部份 $[X]$ ， $[X, Y]$ ， $[X, Z]$ ， $[X, Y, Z]$ ，如上述 $[X, Y, Z]$ 可以不必討論，因此只需研究 $[X]$ ， $[X, Y]$ 與 $[X, Z]$ 即可。設對於每一個方案而言都有一個表現函數 U ，令

$$\Delta = \sum_{i=1}^2 [U(\alpha_i) + U(\beta_i) + U(\tau_i) + U(\delta_i) + U(\sigma_i) + U(\rho_i)] ,$$

則

(a) 選 $[X]$ 的情形，此時相當於在 X, Y, Z 中把滿足 Y, Z 特性的方案全部刪去，如果我們承認 IIA，則在所有方案中選擇到 $[X]$ 。

的機率為

$$\frac{U(\alpha_1) + U(\alpha_2)}{\Delta}$$

(b)考慮 $[X, Y]$ 的情形。這相當於在 X, Y, Z 中刪去 Z ，然後在滿足 X, Y 條件諸方案中以 X 觀點所作的選擇。因為 $[X, Y] = \{\sigma_1, \sigma_2\}$ ，則選擇到 $[X, Y]$ 的機率為

$$\frac{U(\sigma_1) + U(\sigma_2)}{\Delta}$$

如果在 $[X, Y]$ 中選擇 X 的機率記成 $\text{prob}(X, \{X, Y\})$ ，則在所有方案中選擇的機率就是

$$\frac{U(\sigma_1) + U(\sigma_2)}{\Delta} \cdot \text{prob}(X, \{X, Y\}) ;$$

(c)考慮 $[X, Z]$ 的情形，如(b)中所述，其機率為

$$\frac{U(\delta_1) + U(\delta_2)}{\Delta} \cdot \text{prob}(X, \{X, Z\})$$

綜合(a)，(b)，(c)得到 X, Y, Z 中以消去的觀念而選擇到 X 的機率為

$$\text{prob}(X, \{X, Y, Z\}) = \frac{U(\alpha_1) + U(\alpha_2) + [U(\sigma_1) + U(\sigma_2)] \cdot \sum_{i=1}^2 [U(\alpha_i) + U(\beta_i) + U(\gamma_i)] + \dots}{\dots}$$

$$\frac{\text{prob}(X, \{X, Y\}) + [U(\delta_1) + U(\delta_2)] \text{prob}(X, \{X, Z\})}{U(\delta_1) + U(\delta_{\sigma_1}) + U(\rho_1)}$$

令 $U(X) = U(\alpha_1) + U(\alpha_2)$, $U(X, Y) = U([X, Y]) = U(\sigma_1) + U(\sigma_2)$,; 且為了區別起見, 把 $\text{prob}(X, \{X, Y\})$, $\text{prob}(X, \{X, Y, Z\})$ 記成 $P^*(X; Y)$, $P^*(X; Y, Z)$, 因此上式就變成

$$(1) \quad P^*(X; Y, Z) = \frac{U(X) + U(X, Y) \cdot P^*(X; Y) + U(X, Z)}{U(X) + U(Y) + U(Z) + U(X, Y) + U(X, Z)} \\ \cdot \frac{P^*(X; Z)}{+ U(Y, Z)}。$$

同理, 我們可以計算 $P^*(X; Y)$, $P^*(X; Z)$ 如次:

$$P^*(X; Y) = P^*(X, \{X, Y\}) = \frac{U(X) + U(X, Z)}{U(X) + U(Y) + U(X, Z) + U(Y, Z)}$$

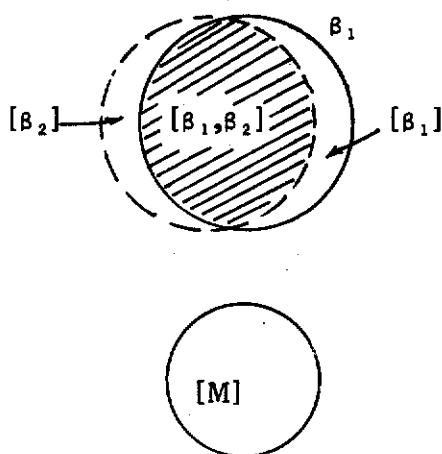
$$P^*(X; Z) = P^*(X, \{X, Z\}) = \frac{U(X) + U(X, Y)}{U(X) + U(Z) + U(X, Y) + U(Y, Z)}$$

這就是最簡單而由 Tversky 在 1972 年所導出的消去選擇模型 (Elimination by Aspect), 以下簡稱爲 EBA 模型。

從上面計算 $P^*(X; Y, Z)$ 的式子來看, 它不但用到了 $U(X)$, $U(X, Y)$ 等項, 最重要的, 它還與 $P^*(X; Y)$, $P^*(X; Z)$ 有關, 也就是說全面的選擇機率是與其局部的選擇有關的 (雖然我們使用了 IIA 性質), 因此我們說, 這是個遞歸 (recursive) 的定義式。

在結束之前, 我們再看看前述的唱片問題。設貝多芬的兩張

唱片為 B_1, B_2 則可設 $U(B_1) = U(B_2) = \alpha$; 又莫札特的唱片令為 M , 設 $U(M) = \beta$; 再設 $U(B_1, B_2) = \gamma$ 。因為 $M, B_1 ; M, B_2$ 完全相異, 故 $[M, B_1], [M, B_2]$ 為空集合 (見下圖) , 於是可令 $U(M, B_1) = U(M, B_2) = 0$ 。



首先, 我們可以證明 $\alpha + \gamma = \beta$ 。利用

$$\begin{aligned} \frac{1}{2} = P^*(B_1; M) &= \frac{U(B_1) + U(B_1, B_2)}{U(B_1) + U(M) + U(B_1, B_2) + U(M, B_2)} \\ &= \frac{\alpha + \gamma}{\alpha + \gamma + \beta} \end{aligned}$$

就得到了 $\alpha + \gamma = \beta$ 。接著, 我們可以算出

$$P^*(B_1; B_2, M) = \frac{\alpha + \frac{1}{2}\gamma}{2\alpha + \beta + \gamma} = P^*(B_2; B_1, M)$$

但是

$$P^*(M; B_1, B_2) = \frac{\beta}{2\alpha + \beta + \gamma} = \frac{\alpha + \gamma}{2\alpha + \beta + \gamma}$$

$$\geq P^*(B_1; B_2, M) = P^*(B_2; B_1, M)$$

因為 $\alpha \neq 0$ ， $\beta \neq 0$ ，故上面等式成立的充要條件是 $\gamma = 0$ ，亦即 $U(B_1, B_2) = 0$ ；換句話說， $[B_1, B_2] = \emptyset$ 。所以導出 $P^*(M; B_1, B_2) = P^*(B_1; B_2, M) = P^*(B_2; B_1, M) = \frac{1}{3}$ 的條件是 B_1 與 B_2 完全相異（如果 EBA 為真）。一般而言，選擇 M 的機率會大過選擇 B_1 或 B_2 的機率（後二者相等），但若 $[B_1], [B_2] \rightarrow \emptyset$ ，亦即 $[B_1, B_2]$ 就是 B_1 或 B_2 的全部，也就是說 B_1, B_2 完全相同，則得 $\alpha \rightarrow 0$ 所以就得到

$$P^*(M; B_1, B_2) = \frac{1}{2}$$

$$P^*(B_1; B_2, M) = P^*(B_2; B_1, M) = \frac{1}{4}$$

因此 IIA 性質就這樣地被完全克服（至少在這個問題上面），雖然在導出這個 EBA 模型時我們還得藉 IIA 之力。

3.10 計算方式

目前系統中只製作完成 Logit 部份，所以僅就此部份討論其計算方式。對於參與選擇者 t 而言，假定他選了 i ，則機率為

$$P_{ti} = \frac{\exp(X_{ti} \beta)}{\sum_k \exp(X_{tk} \beta)}$$
$$= \frac{1}{1 + \sum_{k \neq i} \exp(X_{tk} - X_{ti}) \beta}$$

如果令 $X_{tki} = X_{tk} - X_{ti}$ ，則

$$P_{ti} = \frac{1}{1 + \sum_{k \neq i} \exp(X_{tki} \beta)}$$

對於其它 j ($\neq i$) 而言， t 選擇 j 的機率為：

$$P_{tj} = \frac{\exp(X_{tj} \beta)}{\sum_k \exp(X_{tk} \beta)} \quad (\text{分子分母同除 } \exp(X_{ti} \beta))$$
$$= \frac{\exp(X_{tji} \beta)}{1 + \sum_{k \neq i} \exp(X_{tki} \beta)}$$

這樣在計算時不但減少一次求 $\exp(\quad)$ 的機會，而且因為 X_{tki} 為兩個值之差，因而令 $\exp(\quad)$ 溢位的機會亦較少。

因此，概似函數 $L(\beta)$ 為

$$L(\beta) = \prod_t P_{ti} = \prod_t \frac{1}{1 + \sum_{k \neq i} \exp(X_{tki} \beta)}$$

求 $L(\beta)$ 之極大，相當於求 $-\ln(L(\beta))$ 之極小，故令

$$\begin{aligned} \ell(\beta) &= -\ln(L(\beta)) \\ &= \sum_t \ln \left[1 + \sum_{k \neq i} \exp(X_{tki} \beta) \right] \end{aligned}$$

求 $\ell(\beta)$ 之一階微分得到

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \sum_t \frac{\sum_{s \neq i} \exp(X_{tsi} \beta) \cdot X_{tsi}}{1 + \sum_{k \neq i} \exp(X_{tki} \beta)} \\ &= \sum_t \sum_{s \neq i} \frac{\exp(X_{tsi} \beta) \cdot X_{tsi}}{1 + \sum_{k \neq i} \exp(X_{tki} \beta)} \\ &= \sum_t \sum_{s \neq i} P_{ts} \cdot X_{tsi} \end{aligned}$$

另外，因為

$$\begin{aligned} \frac{\partial P_{tk}}{\partial \beta} &= \frac{\partial}{\partial \beta} \left(\frac{\exp(X_{tki} \beta)}{1 + \sum_{s \neq i} \exp(X_{tsi} \beta)} \right) \\ &= P_{tk} \cdot X_{tki} - P_{tk} \left(\sum_{s \neq i} P_{ts} \cdot X_{tsi} \right) \end{aligned}$$

所以求二階分得到

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \beta' \partial \beta} &= \sum_t \sum_{k \neq i} (P_{tk} X'_{tki} - P_{tk} (\sum_{s \neq i} P_{ts} \cdot X'_{tsi})) \cdot X_{tki} \\ &= \sum_t \sum_{k \neq i} P_{tk} X'_{tki} \cdot X_{tki} - \sum_t (\sum_{k \neq i} P_{tk} \cdot X_{tki}) (\sum_{s \neq i} P_{ts} \cdot X_{tsi})\end{aligned}$$

因為式子冗長，所以在計算時先把它們分解，令

$$T_t = 1 + \sum_{s \neq i} \exp(X_{tsi} \beta)$$

$$V_t = \sum_{s \neq i} \exp(X_{tsi} \beta) \cdot X_{tsi}$$

$$W_t = \sum \exp(X_{tsi}) \cdot X'_{tsi} \cdot X_{tsi}$$

於是

$$\ell(\beta) = \sum_t \text{Ln}(T_t)$$

$$\frac{\partial \ell}{\partial \beta} = \sum_t \frac{V_t}{T_t}$$

$$\frac{\partial^2 \ell}{\partial \beta' \partial \beta} = \sum_t \frac{W_t - (\frac{V_t}{T_t})' \cdot V_t}{T_t}$$

這樣在計算時就會比較簡單明瞭。

因為 $\exp(\)$ 函數極易遭致溢位，所以為防止這個現象發生，我

們要加入若干測試。當在某個 t 時發現 $X_{t,i} \beta \rightarrow \infty$ ，則 $P_{t,r} \rightarrow 0$ ($r \neq s$)，而 $P_{t,s} \rightarrow 1$ ；不但如此， $\ell(\beta) \rightarrow \infty$ ，在求極小值時，已不在下降方向。反之，若有 $X_{t,i} \beta \rightarrow -\infty$ ，則 T_t 、 V_t 與 W_t 中均可不考慮這一項。

綜合起來，我們假設輸入的資料檔中對 t 而言有如下的欄位：

C_t ：被選擇的方案

A_t ：可供選擇的方案數目（含 C_t 一項）

A_{tj} ：表示可供選擇的各個方案（含 C_t ）

X_{tj} ：對應著 A_{tj} 的解釋變數

其次，我們要把這些資料轉換成 $X_{t,i}$ 的型式。

程序：Transform(N)

輸入：

N ：觀測值個數

輸出：

無，但經轉換後的資料在工作檔 WORK 中。

計算程序：

begin

open (work) ;

read (input, C_t , A_t , A_{tj} , X_{tj}) ; { $j=1, 2, \dots$ }

write (work, C_t , A_t-1 , $X_{t1}-X_{ct}$, $X_{t2}-X_{ct}$, ...) ;

close (work) ;

end

其次是計算 $\ell(\beta)$ ， $\frac{\partial \ell}{\partial \beta}$ 與 $\frac{\partial^2 \ell}{\partial \beta' \partial \beta}$ 的程式。

程序：Logit-1 (N, β, ℓ, G, H)

輸入：

N : 觀測值個數；

β : 係數向量。

輸出：

ℓ : 概似函數值；

G : 梯度 (Gradient)，即 $\frac{\partial \ell}{\partial \beta}$ ；

H : 二階微分矩陣 (對稱)。

計算程序：

begin

rewind (WORK)；

$\ell := 0$ ； $G := 0$ ； $H := 0$ ；

for $t := 1$ to N do

begin

$V := 0$ ； $W := 0$ ； { V 與 W 分別對應於 V_t 與 W_t }

read ($i, A, X_{t1i}, X_{t2i}, \dots, X_{tAi}$)；

$T_t := 0$ ；

for $S := 1$ to A do

begin

$xb := X_{tSi} \beta$

if $xb \rightarrow \infty$ then

begin

$G := G + \sum X_{tSi}$ ；

goto 200；

end

```

else if  $xb \rightarrow -\infty$  then
    goto 100
else
     $\text{expxb} := \exp(xb)$ ;
     $T_t := T_t + \text{expxb}$ ;
     $V := V + \text{expxb} \cdot X_{t,i}$ ;
     $W := W + \text{expxb} \cdot X'_{t,i} \cdot X_{t,i}$ ;
100 : end ;
     $\ell := \ell + \ln(T_t)$ ;
     $G := G + V / T_t$ ;
     $H := H + (W - (V / T_t)' \cdot V_t) / T_t$ ;
200 : end ;
end ;

```

如果有重覆選擇出現。令 C_t 為 t 之選擇集合，其方案為 C_1, C_2, \dots, C_m ； t 選擇 C_k 方案有 S_{tk} 次，令 $N_t = \sum S_{tk}$ 。利用多項分配，若 t 選擇 C_k 之機率為 P_{tk} ，則 t 選擇 C_1 有 S_1 次， C_2 有 S_2 次， \dots ， C_m 有 S_m 次的機率為

$$\frac{N_t!}{S_{t1}! S_{t2}! \dots S_{tm}!} P_{t1}^{S_{t1}} P_{t2}^{S_{t2}} \dots P_{tm}^{S_{tm}}$$

所以概似函數為

$$L(\beta) = \prod_t \left\{ \frac{N_t!}{\prod_k S_{tk}!} \prod_k P_{tk}^{S_{tk}} \right\}$$

取對數後有

$$L(\beta) = \sum_t \left\{ \ln \left(\frac{N_t!}{\prod_k S_{tk}!} \right) + \sum_k S_{tk} \cdot \ln(P_{tk}) \right\}$$

其中第一項為常數，並不影響求極值工作，故可以省略。因此，令

$$\ell(\beta) = \sum_t \sum_k S_{tk} \cdot \ln(P_{tk})$$

對 β 求微分，得到

$$\begin{aligned} \frac{\partial \ell}{\partial \beta} &= \sum_t \sum_k S_{tk} (X_{tk} - \sum_q P_{tq} X_{tq}) \\ &= \sum_t \sum_k S_{tk} X_{tk} - \sum_t \sum_k S_{tk} \left(\sum_q P_{tq} X_{tq} \right) \\ &= \sum_t \sum_k S_{tk} X_{tk} - \sum_t \left(\sum_q P_{tq} X_{tq} \right) \sum_k S_{tk} \\ &= \sum_t \sum_k S_{tk} X_{tk} - \sum_t \sum_q N_t P_{tq} X_{tq} \\ &= \sum_t \sum_k (S_{tk} - N_t P_{tk}) X_{tk} \\ &= - \sum_t \sum_k (N_t P_{tk} - S_{tk}) \cdot X_{tk} \end{aligned}$$

二階微分為

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \beta' \partial \beta} &= - \left\{ \sum_t \sum_q N_t P_{tq} X'_{tq} X_{tq} - \sum_t \sum_k N_t P_{tk} \left(\sum_q P_{tq} X'_{tq} \right. \right. \\ &\quad \left. \left. \cdot X_{tq} \right) \right\} \end{aligned}$$

計算的細節我們不在此地討論，因為與 Logit-1 大同小異。不過值得一提的是，有重覆選擇的情況下，計算成本是非常昂貴的，因為每一個人的每一個方案的機率都要算到。

3.11 整合性Logit 模式

有一個可以化簡有重覆選擇的估計方法，特別是，重覆次數愈多，則愈好使用，這就是整合性Logit 模式。

假設某人對K項方案作選擇，而且每一項方案對他而言都可以選取；如果他一共對這些方案選擇了T次，其中對第j個方案選擇了 S_j 次，於是

$$T = \sum_{j=1}^k s_j$$

再令

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{bmatrix}, F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{bmatrix} = \begin{bmatrix} s_1 / T \\ s_2 / T \\ \vdots \\ s_k / T \end{bmatrix} = S / T$$

因此S是為選擇次數，F為選擇頻率（frequency），目前我們假設 s_i （ $1 \leq j \leq k$ ）都不是零，所以 f_i 亦然。

如果每次選擇都相互獨立，也就是說這一次的選擇與上一次和下一次無關……等等， s_j 應該是形成一個多項分配（Multinomial Distribution），它的密度函數（Density Function）為

$$M(S | T, \theta_1, \theta_2, \dots, \theta_k) = \frac{T!}{\prod_{j=1}^k (s_j!)} \prod_{j=1}^k \theta_j^{s_j}$$

上式中， θ_p 表示某人在選擇時，會選中第 p 個方案的機率；在目前我們先假設它是個已知數。

由多項分配的特性，我們知道 S 的平均數（或期望值）與變異數為

$$E[S] = \begin{bmatrix} T\theta_1 \\ T\theta_2 \\ \vdots \\ T\theta_k \end{bmatrix}, \text{VAR}[S] = \begin{bmatrix} T\theta_1(1-\theta_1) & -T\theta_1\theta_2 & \cdots & -T\theta_1\theta_k \\ -T\theta_2\theta_1 & T\theta_2(1-\theta_2) & \cdots & -T\theta_2\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ -T\theta_k\theta_1 & -T\theta_k\theta_2 & \cdots & T\theta_k(1-\theta_k) \end{bmatrix}$$

令

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{bmatrix}, \Sigma = \begin{bmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 & \cdots & -\theta_1\theta_k \\ -\theta_2\theta_1 & \theta_2(1-\theta_2) & \cdots & -\theta_2\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ -\theta_k\theta_1 & -\theta_k\theta_2 & \cdots & \theta_k(1-\theta_k) \end{bmatrix}$$

於是

$$E[S] = T\theta, \text{VAR}[S] = T\Sigma$$

如果某人一再重覆地選擇，而使得 $T \rightarrow \infty$ （亦即 T 很大的時候），由機率理論中的極限定理，我們有

$$(S - T\theta)(T\Sigma)^{-\frac{1}{2}} \xrightarrow{\mathcal{L}} Z$$

此地 Z 是一個滿足標準常態分配的隨機變數；又 \mathcal{L} 表示當 $T \rightarrow \infty$ 時，左式的隨機變數的機率分配趨近右邊隨機變數的機率分配。

把上式左邊加以變化得到

$$\sqrt{T} (F - \theta) \xrightarrow{f} Z$$

於是

$$\sqrt{T} (F - \theta) \xrightarrow{f} Z \sim N(0, \frac{1}{T} \Sigma)$$

也就是說，當 T 充分大的時候， F ——表示某個人的選擇頻率，趨近一個以 θ 為平均數，共變異矩陣為 Σ / T 的常態分配；這是個非常有用的結果。

接著，再看一個函數 g 如下：

$$g : \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_k \end{bmatrix} \rightarrow \begin{bmatrix} u_2 \\ \text{Ln} \frac{u_2}{u_1} \\ u_3 \\ \text{Ln} \frac{u_3}{u_1} \\ \vdots \\ u_k \\ \text{Ln} \frac{u_k}{u_1} \end{bmatrix}$$

此地 U_i ($1 \leq i \leq k$) 均大於零；如果用 g 來轉換 F ，於是就由斯陸斯基定理 (Slutsky Theorem)，我們得到 (當 $T \rightarrow \infty$ 時)：

$$\sqrt{T} [g(F) - g(\theta)] \xrightarrow{f} \frac{\partial g}{\partial F} \bigg|_{\theta} \cdot Z$$

上式中， $Z \sim N(0, \frac{1}{T} \Sigma)$ 。由 g 的定義，求它對 F 的偏微分，即

是

$$\frac{\partial g}{\partial F} = \begin{bmatrix} -\frac{1}{f_1} & \frac{1}{f_2} & 0 & 0 & \cdots & 0 \\ -\frac{1}{f_1} & 0 & \frac{1}{f_3} & 0 & \cdots & 0 \\ -\frac{1}{f_1} & 0 & 0 & \frac{1}{f_4} & \cdots & 0 \\ \vdots & & & & \ddots & \\ -\frac{1}{f_1} & 0 & 0 & 0 & \cdots & \frac{1}{f_k} \end{bmatrix}$$

因此以 θ 代入上式，令為 J ，得到

$$J = \frac{\partial g}{\partial F} \Big|_{\theta} = \begin{bmatrix} -\frac{1}{\theta_1} & \frac{1}{\theta_2} & 0 & 0 & \cdots & 0 \\ -\frac{1}{\theta_1} & 0 & \frac{1}{\theta_3} & 0 & \cdots & 0 \\ -\frac{1}{\theta_1} & 0 & 0 & \frac{1}{\theta_4} & \cdots & 0 \\ \vdots & & & & \ddots & \\ -\frac{1}{\theta_1} & 0 & 0 & 0 & \cdots & \frac{1}{\theta_k} \end{bmatrix}$$

因為 $Z \sim N(0, \frac{1}{T} \Sigma)$ ，因此我們不妨求 $J \cdot Z$ 的分配；由多項常態分配的定理知，

$$E[JZ] = J \cdot E[Z] = J \cdot 0 = (\text{零矩陣})$$

$$\text{VAR}[JZ] = J \cdot \text{VAR}[Z] \cdot J' = J \cdot \left(\frac{1}{T} \Sigma\right) J'$$

利用 J 與 Σ 的值 (它們只包含了 θ_1) 可以算出

$$V = J \cdot \Sigma \cdot J' = \begin{bmatrix} \frac{1}{\theta_1} + \frac{1}{\theta_2} & \frac{1}{\theta_1} & \frac{1}{\theta_1} \cdots \frac{1}{\theta_1} \\ \frac{1}{\theta_1} & \frac{1}{\theta_1} + \frac{1}{\theta_3} & \frac{1}{\theta_1} \cdots \frac{1}{\theta_1} \\ \vdots & & \\ \frac{1}{\theta_1} & \frac{1}{\theta_1} & \frac{1}{\theta_1} \cdots \frac{1}{\theta_1} + \frac{1}{\theta_k} \end{bmatrix}$$

所以

$$J \cdot Z \sim N\left(0, \frac{1}{T} V\right)$$

這就是我們所需要的結果；也就是說，當某人對 K 個方案重複地並且獨立地選擇次數 $T \rightarrow \infty$ (即很大) 時，於是選擇頻率與實際被選中機率之間有如下的關係

$$\begin{bmatrix} \text{Ln} \frac{f_2}{f_1} \\ \text{Ln} \frac{f_3}{f_1} \\ \vdots \\ \text{Ln} \frac{f_k}{f_1} \end{bmatrix} - \begin{bmatrix} \text{Ln} \frac{\theta_2}{\theta_1} \\ \text{Ln} \frac{\theta_3}{\theta_1} \\ \vdots \\ \text{Ln} \frac{\theta_k}{\theta_1} \end{bmatrix} \xrightarrow{\mathcal{L}} J \cdot Z \sim N\left(0, \frac{1}{T} V\right)$$

如果有 $i = 1, 2, \dots, N$ 個人參與選擇，其中第 i 個人的選擇次數，選擇頻率分別記成 S_i, F_i ，則

$$S_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{ik} \end{bmatrix}, F_i = \begin{bmatrix} f_{i1} \\ f_{i2} \\ \vdots \\ f_{ik} \end{bmatrix} = \begin{bmatrix} s_{i1}/T_i \\ s_{i2}/T_i \\ \vdots \\ s_{ik}/T_i \end{bmatrix} = S_i/T_i, \theta_i = \begin{bmatrix} \theta_{i1} \\ \theta_{i2} \\ \vdots \\ \theta_{ik} \end{bmatrix}$$

此地的 $T_i = \sum_{j=1}^k s_{ij}$ 。再令第 i 個人的對數機率比之差為 ϵ_i ，即

$$\epsilon_i = \begin{bmatrix} \ln \frac{f_{i2}}{f_{i1}} & - & \ln \frac{\theta_{i2}}{\theta_{i1}} \\ \ln \frac{f_{i3}}{f_{i1}} & - & \ln \frac{\theta_{i3}}{\theta_{i1}} \\ \vdots & & \vdots \\ \ln \frac{f_{ik}}{f_{i1}} & - & \ln \frac{\theta_{ik}}{\theta_{i1}} \end{bmatrix}$$

由前節得如， $\epsilon_i \sim N(0, \frac{1}{T} V_i)$ ，此地

$$V_i = \begin{bmatrix} \frac{1}{\theta_{i1}} + \frac{1}{\theta_{i2}} & \frac{1}{\theta_{i1}} & \dots & \frac{1}{\theta_{i1}} \\ \frac{1}{\theta_{i1}} & \frac{1}{\theta_{i1}} + \frac{1}{\theta_{i3}} & \dots & \frac{1}{\theta_{i1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\theta_{i1}} & \dots & \dots & \frac{1}{\theta_{i1}} + \frac{1}{\theta_{ik}} \end{bmatrix}$$

我們把所有的 ϵ_i ($1 \leq i \leq N$) 收集在一起，成為一個更大的向量，即 $N(K-1)$ 列，1 行：

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

如果任意兩個人的選擇均相互獨立，也就是說當 $i \neq i'$ 時，第 i 個人對某個方案的選擇與第 i' 個人的選擇無關，因此 $\epsilon_i, \epsilon_{i'}$ 之間就沒有相關性，亦即共變異矩陣為零，基於此， ϵ 的共變異矩陣 V^* 如下：

$$V^* = \begin{bmatrix} \frac{V_1}{T_1} & & & 0 \\ & \frac{V_2}{T_2} & & \\ & & \ddots & \\ 0 & & & \frac{V_N}{T_N} \end{bmatrix}$$

所以 ϵ 就是

$$\epsilon \sim N(0, V^*)$$

當然， s_{ij} (或 f_{ij}) 是一個觀察而來的值 (假設它們均大於零)，而 θ_{ij} 在前述的說明中卻又一直當成已知數來看；然而實際上，這個「真正」的選擇機率卻是個未知數，所以我們的重點在於為 θ_{ij} 立出個模式。關於 θ_{ij} 的模式中，最簡單的莫過於 Logit 模式，顧名思義，正是前節所討論的結果；因此，假設

$$\theta_{ij} = \frac{e^{X_{ij}\beta}}{\sum_{t=1}^k e^{X_{it}\beta}}$$

此地 X_{it} 表示第 i 個人有關第 t 個方案的屬性資料（列向量），而 β 是一個待估計的係數（行向量）。如果我們採用這個模式，於是

$$\begin{aligned} \ln \frac{\theta_{ij}}{\theta_{i1}} &= \ln \left[\left(\frac{e^{X_{ij}\beta}}{\sum_{t=1}^k e^{X_{it}\beta}} \right) / \left(\frac{e^{X_{i1}\beta}}{\sum_{t=1}^k e^{X_{it}\beta}} \right) \right] \\ &= \ln \frac{e^{X_{ij}\beta}}{e^{X_{i1}\beta}} = \ln e^{(X_{ij} - X_{i1})\beta} \\ &= (X_{ij} - X_{i1})\beta \end{aligned}$$

此地 $j = 2, 3, \dots, K$ 。如果令

$$y_{ij} = \ln \frac{s_{ij}}{s_{i1}}, Y_1 = \begin{bmatrix} y_{i2} \\ y_{i3} \\ \vdots \\ y_{ik} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \quad (2 \leq j \leq k)$$

$$Z_{ij} = X_{ij} - X_{i1}, Z_1 = \begin{bmatrix} Z_{i2} \\ Z_{i3} \\ \vdots \\ Z_{ik} \end{bmatrix}, Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix} \quad (2 \leq j \leq k)$$

那麼利用上述 ϵ, ϵ_i 的定義，在 T_i 很大的時候，我們得到

$$\epsilon = Y - Z\beta \sim N(0, V^*)$$

這不過是一個常見的迴歸方程式！然而這個方程式却不能用

$$\beta = (Z'Z)^{-1} (Z'Y)$$

來估計 β 的值，其道理在於 ϵ 並不滿足一般迴歸理論（簡稱 OLS）的同質性（Homoscedasticity），亦即 $E[\epsilon'\epsilon] \neq \sigma^2 I$ ；事實上卻是 $E[\epsilon'\epsilon] = V^*$ ，所以 $\hat{\beta}$ 的求法就得使用廣義迴歸理論（Generalized Least Square，簡稱 GLS），亦即

$$\beta^* = (Z'V^{*-1}Z)^{-1} (Z'V^{*-1}Y)$$

使用 GLS 的好處是在於 β^* 的變異數要比 $\hat{\beta}$ 來得小，亦即可以取得有效性（Efficiency），從而估計值也就有了較高的精確度。

然而，求解 β^* 也並非是沒有困難的。Z, Y 當然是已知的觀測值， V^* 中的 T_i 亦然，但是 V^* 中還有 θ_{ij} ，這些 θ_{ij} 卻又是 β 的函數，所以嚴格來說， V^* 不可以看做是個已知的矩陣；當然，從 $\epsilon = Y - Z\beta \sim N(0, V^*)$ 可以用最大概似法來解，但是當 K 與 N 很大時，從計算成本的觀點來看是很不經濟的；因此，我們不妨用兩段式的 OLS 來克服這個困難，也就是說，在第一段的 OLS 估計中，我們求出 β ，再利用 $\hat{\beta}$ 而求出 $\hat{\theta}_{ij}$ ，做為 θ_{ij} 的估計值，因此就得到了 V^* 的一個估計值 \hat{V}^* ，接著以 \hat{V}^* 代入 GLS 從而計算出 β^* 。從計算成本來看，這不過是兩次 OLS 的手續而已，是相當經濟的，而且當 T_i ($1 \leq i \leq N$) 很大時， β^* 將不會與最大概似法所得到的結果有什麼差異。

在計算 $\hat{\beta}$ ，即第一次迴歸時，計算的工作沒有什麼特殊的地方；不過在第二次求 β^* 時卻需要矩陣 \hat{V}^* ，這一個矩陣列數與行數都相當大，因此把它保存在計算機記憶體中是件相當不合算的事，不過整

件工作卻可以利用 \hat{V}^* 的特殊性質而加以化簡。因為

$$V^* = \begin{bmatrix} \frac{\hat{V}_1^*}{T_1} & & 0 \\ & \frac{\hat{V}_2^*}{T_2} & \\ 0 & & \frac{\hat{V}_N^*}{T_N} \end{bmatrix}$$

所以

$$(V^*)^{-1} = \begin{bmatrix} T_1 (\hat{V}_1^*)^{-1} & & 0 \\ & T_2 (\hat{V}_2^*)^{-1} & \\ 0 & & T_N (\hat{V}_N^*)^{-1} \end{bmatrix}$$

如果把 Z 與 Y 也相對地分解成 N 個參與選者的組合，則

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

於是我們有

$$\begin{aligned}\beta^* &= [Z' (\hat{V}^*)^{-1} Z]^{-1} [Z' (\hat{V}^*)^{-1} Y] \\ &= \left[\sum_{i=1}^N Z_i \left(\frac{\hat{V}_i^*}{T_i} \right)^{-1} Z_i \right]^{-1} \left[\sum_{i=1}^N Z_i \left(\frac{\hat{V}_i^*}{T_i} \right)^{-1} Y_i \right]\end{aligned}$$

因為 $\sum Z_i \left(\frac{\hat{V}_i^*}{T_i} \right)^{-1} Z_i$ 與 $\sum Z_i \left(\frac{\hat{V}_i^*}{T_i} \right)^{-1} Y_i$ 所用的記憶體極為有限（只與係數的個數有關，而與參與選擇的人數， N ，無關），因此也就大量地節省了記憶體。

另外一個問題是，我們假設了 $s_{ij} \neq 0$ ，從而 $f_{ij} \neq 0$ ，因此才順利地使用自然對數，但是實際上 s_{ij} 是有可能為零的，在處理這類問題時，我們習慣上是加上 $\frac{1}{2}$ ，亦即是 $s_{ij} + \frac{1}{2}$ 來處理；加上 $\frac{1}{2}$ 的理由我們會在後面討論廣義重力模式再行介紹。

第四章 產生模式

4.1 前言

旅次產生模式是第二項主題，在我們的研究計畫中仍然以個體模式為主要課題，基本上，這一類型的技巧包含有 Tobin 模式（習稱 Tobit），Poisson 模式；另外，用迴歸模式來從事旅次產生的研究是常見的，但此等模式易發生不滿足均質的條件，從而修正異質性也是一件重要工作。另一項重要論題常發生在分目的別旅次產生模式中；如果第 i 區（或第 i 個家庭）第 p 個目的別的旅次產生模式為

$$Y_{ip} = X_{ip}B_p + \epsilon_{ip} \quad (p = 1, 2, \dots)$$

於是當某個目的別的旅次增加，則有可能使另一個目的別的旅次降低。舉例而言，如果 p 為購物， q 為娛樂，則一個家庭的總旅次產生數因為相當固定，因而 Y_{ip} 增加，自然就會使 Y_{iq} 也因之而異；換言之，在各目的別模式中的應變數之間並非完全獨立。這一型模式以計量經濟學上的聯立模型為最佳技巧。

因為處理異質性與聯立模型的軟體與文獻很多，故本章僅討論 Tobin 與 Poisson 模式。

4.2 Tobin 模式

在傳統旅次產生模式中，迴歸分析是最常見到的技巧之一；在迴歸分析中，不論是個體模型——以家庭為旅次產生單位，或是以交通分區為單位，我們都會有如下的型式：

$$T_i = X_i \beta + \varepsilon_i$$

此地 X_i 為第 i 個單位的解釋變數， T_i 為我們觀察到的旅次產生量，這個量毫無疑問地，應該不會小於零；然而，用來解釋 T_i 這個不可能觀察到負值的 $X_i \beta$ 却顯然地有小於零的可能，因此 ε_i 這個殘差項就不能滿足一般迴歸方程式的要求，所以尋常的迴歸模型在此就不能成立。事實上，這種情況在以交通分區為單位的旅次產生模式中不太可能發生，因為一個分區的（縱使再分成目的別）旅次產生量總是會有若干，但若以家庭為單位的模型，這個狀況出現的可能性就非常大了。

如果把 $X_i \beta + \varepsilon_i$ （或是更廣義地寫成 $U_i = f(X_i, \beta) + \varepsilon_i$ ）視為一個效用函數，用這個效用函數來解釋旅次產生的量 T_i ，於是當 $X_i \beta + \varepsilon_i \leq 0$ ，則表示效用過低，因此沒有產生旅次的價值，從而我們的觀察值 T_i 就是 0。但是，我們能夠觀察到的却只有第一種情況，而對第二種情況而言， $T_i = 0$ 就隱含了效用過低的因素了，因此我們的模式為

$$T_i = X_i \beta + \varepsilon_i \quad \text{若右式} > 0$$

$$T_i = 0 \quad \text{若右式} \leq 0$$

這就是 Tobin 模式，通常就叫做 Tobit。

不過， $X_i \beta$ 是個連續函數，而我們所能觀察到的 T_i 卻是一個離

散量 (0, 1, 2, 3, …)，因此 $X_i \beta$ 需整合過後才能得到離散值。譬如，

$$X_i \beta \in (0, a_1) \rightarrow T_i = 0$$

$$X_i \beta \in (a_1, a_2) \rightarrow T_i = 1$$

⋮

⋮

$$X_i \beta \in (a_n, a_{n+1}) \rightarrow T_i = n$$

$$X_i \beta \in (a_{n+1}, \infty) \rightarrow T_i = n + 1$$

$a_1, a_2, \dots, a_n, a_{n+1}$ 可能為已知 (譬如 $a_i = i$)，但亦可能為待定的未知數，而需要在估計模式時一併求出，但因為此等模式較為困難，所以目前暫時不考慮這種類型。另外，因為旅次產生模式以線性的可能性較大，因此我們只考慮

$$T_i = X_i \beta + \epsilon_i \quad \text{右式} > 0$$

$$T_i = 0 \quad \text{右式} \leq 0$$

這一個特例，而不考慮把 $X_i \beta$ 改成更廣義的 $f(X_i, \beta)$ 型式。

4.2.1 估計技巧

在估計模式時我們仍然只考慮 ϵ_i 滿足相同的常態分配， $\epsilon_i \sim N(0, \sigma^2)$ ，而 σ^2 為未知，此假設相當於說明模式無異質性 (Heteroscedasticity)；但若有，也不難用後面將要討論的方法做局部的解決。

假設所有的觀測值分成兩組，一組中 T_i 為 0，令這一組的觀測值為 C_0 ，而另一組者為 C_1 ，因此在 C_1 中的觀測值 T_i 全部大於 0。 C_1 中的各個觀測值其實就是尋常迴歸模式，因為 $X_i \beta$ 可以直接解釋 T_i ，因此 $\epsilon_i = T_i - X_i \beta$ 為常態分配。為方便起見，令

$$f(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$$

$$F(x, \sigma^2) = \int_{-\infty}^x f(t, \sigma^2) dt$$

亦即 $f(x, \sigma^2)$ 為常態分配 (平均數為 0, 變異數為 σ^2) 的密度函數, 而 $F(x, \sigma^2)$ 為 $f(x, \sigma^2)$ 自 $-\infty$ 的到 x 的累計機率。

因為當 $i \in C_1$, 則 $\varepsilon_i = T_i - X_i \beta \sim N(0, \sigma^2)$, 所以這些 ε_i 的概似函數為

$$\prod_{i \in C_1} f(\varepsilon_i, \sigma^2) = \prod_{i \in C_1} f(T_i - X_i \beta, \sigma^2)$$

再回頭看 C_0 中的觀察值, 因為每一個觀察值 T_i 都是零, 這表示右式 $X_i \beta + \varepsilon_i$ 小於零, 亦即 $X_i \beta + \varepsilon_i \leq 0$, 因此 $\varepsilon_i \leq -X_i \beta$ 。因為 $\varepsilon_i \sim N(0, \sigma^2)$, 所以 $\varepsilon_i \leq -X_i \beta$ 的機率為

$$\begin{aligned} p(\varepsilon_i | \varepsilon_i \leq -X_i \beta) &= F(-X_i \beta, \sigma^2) \\ &= 1 - F(X_i \beta, \sigma^2) \end{aligned}$$

於是這些 ε_i 的概似函數為

$$\prod_{i \in C_0} p(\varepsilon_i | \varepsilon_i \leq -X_i \beta) = \prod_{i \in C_0} (1 - F(X_i \beta, \sigma^2))$$

於是就可以立出一個 $C_0 \cup C_1$ 的概似函數如下:

$$L(\beta, \sigma^2) = \prod_{i \in C_0} [1 - F(X_i \beta, \sigma^2)] \cdot \prod_{i \in C_1} f(T_i - X_i \beta, \sigma^2)$$

求 $L(\beta, \sigma^2)$ 的極大值 (為唯一, 但不在此地證明), 就可以得到 Tobit 模式的係數 β , 與變異數 σ^2 。

4.2.2 計算方式

在求 $L(\beta, \sigma^2)$ 的極大值時, 通常都會使用 $\ln(L(\beta, \sigma^2))$ 函數。令 C_0 中有 M 個元素, C_1 中有 N 個元素, 則

$$\begin{aligned} \ell(\beta, \sigma^2) &= \ln(L(\beta, \sigma^2)) \\ &= \ln \left\{ \prod_{i \in C_0} [1 - F(X_i \beta, \sigma^2)] \cdot \prod_{i \in C_1} f(T_i - X_i \beta, \sigma^2) \right\} \\ &= \sum_{i \in C_0} \ln [1 - F(X_i \beta, \sigma^2)] + \sum_{i \in C_1} \ln f(T_i - X_i \beta, \sigma^2) \end{aligned}$$

先化簡右邊的第二式

$$\begin{aligned} \ln f(T_i - X_i \beta, \sigma^2) &= \ln \left\{ \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp \left(-\frac{1}{2} \left(\frac{T_i - X_i \beta}{\sigma} \right)^2 \right) \right\} \\ &= -\ln(\sqrt{2\pi}\sigma) - \frac{1}{2} \frac{(T_i - X_i \beta)^2}{\sigma^2} \\ &= -\ln(\sqrt{2\pi}) - \ln(\sqrt{\sigma^2}) - \frac{1}{2} \frac{(T_i - X_i \beta)^2}{\sigma^2} \\ &= C - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{(T_i - X_i \beta)^2}{\sigma^2} \end{aligned}$$

把這一項相加

$$\sum_{i \in C_1} \ln f(T_i - X_i \beta, \sigma^2) = -N \cdot C - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i \in C_1} \frac{(T_i - X_i \beta)^2}{\sigma^2}$$

因為 N, C 為常數，對求極大與極小值卻沒有影響，所以刪除它，因此最後的概似函數為

$$\begin{aligned} \ell(\beta, \sigma^2) = & \sum_{i \in C_0} L_n [1 - F(X_i \beta, \sigma^2)] - \frac{N}{2} \ln \sigma^2 - \\ & \frac{1}{2} \sum_{i \in C_1} \frac{(T_i - X_i \beta)^2}{\sigma^2} \end{aligned}$$

4.2.3 計算係數微分

非線性最佳化程序通常需要一階或二階微分，但 Tobin 模式却更為複雜，茲先計算函數微分如次。

為方便起見，令

$$f(x, \sigma^2) = \frac{1}{2\pi\sigma} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]$$

$$F(x, \sigma^2) = \int_{-\infty}^x f(t, \sigma^2) dt$$

所以

$$\frac{\partial f}{\partial \beta} = \frac{\partial f(X_i \beta, \sigma^2)}{\partial \beta} = -f(X_i \beta, \sigma^2) \cdot \frac{X_i \beta}{\sigma^2} \cdot X_i$$

$$\frac{\partial F}{\partial \beta} = \frac{\partial F(X_i \beta, \sigma^2)}{\partial \beta} = f(X_i \beta, \sigma^2) \cdot X_i$$

因此求 $\ell(\beta, \sigma^2)$ 第一部份的微分

$$\frac{\partial}{\partial \beta} \text{Ln} [1 - F(X_i \beta, \sigma^2)] = - \frac{f(X_i \beta, \sigma^2)}{1 - F(X_i \beta, \sigma^2)} \cdot X_i$$

第二部份的微分為

$$\frac{\partial}{\partial \beta} \text{Ln} f(T_i - X_i \beta, \sigma^2) = \frac{T_i - X_i \beta}{\sigma^2} \cdot X_i$$

所以

$$(1) \quad \frac{\partial \ell}{\partial \beta} = - \sum_{i \in C_0} \frac{f(X_i \beta, \sigma^2)}{1 - F(X_i \beta, \sigma^2)} \cdot X_i + \sum_{i \in C_1} \frac{T_i - X_i \beta}{\sigma^2} \cdot X_i$$

其次計算二階微分如下。因為

$$\begin{aligned} \frac{\partial}{\partial \beta} \left[\frac{f(X_i \beta, \sigma^2)}{1 - F(X_i \beta, \sigma^2)} \right] &= \frac{1}{1 - F(X_i \beta, \sigma^2)} \cdot \frac{\partial f(X_i \beta, \sigma^2)}{\partial \beta} \\ &\quad + \frac{f(X_i \beta, \sigma^2)}{[1 - F(X_i \beta, \sigma^2)]^2} \cdot \frac{\partial F(X_i \beta, \sigma^2)}{\partial \beta} \\ &= - \frac{f(X_i \beta, \sigma^2)}{1 - F(X_i \beta, \sigma^2)} \cdot \frac{X_i \beta}{\sigma^2} \cdot X_i \\ &\quad + \frac{f^2(X_i \beta, \sigma^2)}{[1 - F(X_i \beta, \sigma^2)]^2} \cdot X_i \end{aligned}$$

因此

$$(2) \quad \frac{\partial^2 \ell}{\partial \beta \partial \beta'} = \sum_{i \in C_0} \frac{f(X_i \beta, \sigma^2)}{1 - F(X_i \beta, \sigma^2)} \cdot \frac{X_i \beta}{\sigma^2} \cdot X_i' X_i$$

$$= \sum_{i \in C_0} \frac{f(X_i, \beta, \sigma^2)^2}{[1 - F(X_i, \beta, \sigma^2)]^2} X_i' X_i - \frac{1}{\sigma^2} \sum_{i \in C_1} X_i' X_i$$

上面(1)式與(2)式都是在已知 σ^2 的情況下計算的，但其實 σ^2 為未知，所以解法應分成兩段，第一段為先估計 σ^2 的值 $\hat{\sigma}_1^2$ ，再用 Newton 方法解出 $\hat{\beta}_1$ ，用 $\hat{\beta}_1$ 再估計 $\hat{\sigma}_2^2$ ，用 $\hat{\sigma}_2^2$ 估計 $\hat{\beta}_2$ ，……，亦即用 $\hat{\beta}_{i-1}$ 估計出 $\hat{\sigma}_i^2$ ，把 $\hat{\sigma}_i^2$ 代入上面的式子再估計 $\hat{\beta}_i$ 。最後的結果為 $\hat{\beta}_i$ 與 $\hat{\sigma}_i^2$ 的極限值，

$$\hat{\beta}^* = \lim_{i \rightarrow \infty} \hat{\beta}_i$$

$$\hat{\sigma}^{2*} = \lim_{i \rightarrow \infty} \hat{\sigma}_i^2$$

因為最大概似法為求 ℓ 的極大值，故可令

$$\bar{\ell}(\beta, \sigma^2) = -\ell(\beta, \sigma^2)$$

而改求 $\bar{\ell}(\beta, \sigma^2)$ 的極小值，從而一階與二階微分為

$$\frac{\partial \bar{\ell}}{\partial \beta} = - \frac{\partial \ell}{\partial \beta}$$

$$\frac{\partial^2 \bar{\ell}}{\partial \beta \partial \beta'} = - \frac{\partial^2 \ell}{\partial \beta \partial \beta'}$$

所以 Newton 方法的搜尋方向為

$$\begin{aligned}\delta &= - \left[\frac{\partial^2 \bar{\ell}}{\partial \beta \partial \beta'} \right]^{-1} \cdot \frac{\partial \bar{\ell}}{\partial \beta} \\ &= - \left[\frac{\partial^2 \ell}{\partial \beta \partial \beta'} \right]^{-1} \cdot \frac{\partial \ell}{\partial \beta}\end{aligned}$$

4.2.4 對變異數的微分

計算概似函數 $\ell(\beta, \sigma^2)$ 對 σ^2 的微分並不困難。因為

$$F(X_i \beta, \sigma^2) = \int_{-\infty}^{X_i \beta} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{t}{\sigma}\right)^2 dt$$

令 $z = t/\sigma$ 從事變數代換，

$$F(X_i \beta, \sigma^2) = \int_{-\infty}^{X_i \beta / \sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) dz$$

所以

$$\begin{aligned}\frac{\partial F(X_i \beta, \sigma^2)}{\partial \sigma^2} &= \frac{\partial F}{\partial (X_i \beta / \sigma)} \cdot \frac{\partial (X_i \beta / \sigma)}{\partial \sigma^2} \\ &= \left\{ \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} (X_i \beta / \sigma)^2\right) \cdot X_i \beta \right\} \\ &\quad \cdot \frac{-1}{2 \sigma^3}\end{aligned}$$

$$\begin{aligned}
&= \frac{-1}{2\sigma^2} \cdot \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}(X_i\beta/\sigma)^2\right) \right. \\
&\quad \left. \cdot X_i\beta \right\} \\
&= -\frac{f(X_i\beta, \sigma^2)}{2\sigma^2} \cdot X_i\beta
\end{aligned}$$

因此求 $l(\beta, \sigma^2)$ 對 σ^2 偏微分得到

$$\begin{aligned}
\frac{\partial l}{\partial \sigma^2} &= \sum_{i \in C_0} \frac{1}{1 - F(X_i\beta, \sigma^2)} \cdot \left(-\frac{\partial F(X_i\beta, \sigma^2)}{\partial \sigma^2} \right) \\
&\quad - \frac{N}{2\sigma^2} - \frac{1}{2} \sum_{i \in C_1} (T_i - X_i\beta)^2 \cdot \frac{\partial}{\partial \sigma^2} \left(\frac{1}{\sigma^2} \right) \\
&= \sum_{i \in C_0} \frac{f(X_i\beta, \sigma^2)}{1 - F(X_i\beta, \sigma^2)} \cdot \frac{1}{2\sigma^2} \cdot X_i\beta - \frac{N}{2} \frac{1}{\sigma^2} \\
&\quad + \frac{1}{2\sigma^4} \sum_{i \in C_1} (T_i - X_i\beta)^2
\end{aligned}$$

因為在極值時偏微分為零，因此令上式為0，可以解出 β 與 σ^2 的關係如下，

$$\hat{\sigma}^2 = \frac{\sum_{i \in C_1} (T_i - X_i\beta)^2}{N - \sum_{i \in C_0} \frac{f(X_i\beta, \sigma^2)}{1 - F(X_i\beta, \sigma^2)} \cdot X_i\beta}$$

換言之， $l(\beta, \sigma^2)$ 可以分成兩個問題：已知 σ^2 解 β ，已知 β 解 σ^2

，所以這是個可分離 (Separable) 的題目。

4.2.5 解題程序

在討論如何解出 β 與 σ^2 之前，我們先要有兩個函數，第一個算出概似函數以及一階、二階微分，第二個算出已經 β 時的 σ^2 估計值。

函數： $\ell(\beta, \sigma^2, M, N, G, H)$

參數：

β : 輸入的係數；

σ^2 : 輸入的變異數；

M : $T_i = 0$ 的個數；

N : $T_i > 0$ 的個數；

G : 輸出的一階微分向量；

H : 輸出的二階微分矩陣；

計算程序：

begin { 計算概似函數，一階與二階微分 }

$\ell := -N * \text{Ln}(\sigma^2) / 2$; $G := 0$; $H := 0$;

for $i := 1$ to $M+N$ do

begin

readin (X_i, T_i) ;

$xb := X_i \cdot \beta$;

$\epsilon := T_i - xb$;

$f := f(xb, \sigma^2)$; { $f()$ 為常態分配密度函數 }

$F := 1 - F(xb, \sigma^2)$; { $F()$ 為常態分配函數 }

if $T_i \leq 0$ then

begin

```

         $\ell$  :=  $\ell + \text{Ln}(F)$  ;
         $\epsilon$  :=  $f / F$  ;
         $G$  :=  $G - t * X_i$  ;
         $H$  :=  $H + (t * x_b / \sigma^2 - t * t) * X'_i \cdot X_i$  ;
    end
else
    begin
         $\ell$  :=  $\ell - \epsilon * \epsilon / (2 * \sigma^2)$  ;
         $G$  :=  $G + (\epsilon / \sigma^2) * X_i$  ;
         $H$  :=  $H - (X_i \cdot X_i) / \sigma^2$  ;
    end
end
end

 $\ell$  :=  $\ell - \ell$  ;
end

```

計算新 σ^2 的程序較為簡單，只不過是依上節的式子計算而已。

函數：Sigma (β , σ^2 , M, N)

參數：

β : 輸入的係數；
 σ^2 : 輸入的舊變異數；
M : $T_i = 0$ 的個數；
N : $T_i > 0$ 的個數；

計算程序：

```

begin { 計算新變異數 }
     $S_1$  := 0 ;
     $S_2$  := 0 ;

```

```

for i := 1 to M+N do
  begin
    readin ( Xi , Ti ) ;
    xb := Xi ·  $\beta$  ;
    if Ti > 0 then
      begin
         $\epsilon$  := Ti - xb ;
        S1 := S1 +  $\epsilon$  *  $\epsilon$  ;
      end
    else
      begin
        f := f ( xb ,  $\sigma^2$  ) ;
        F := 1 - F ( xb ,  $\sigma^2$  ) ;
        S2 := S2 - ( f / F ) * xb ;
      end
    end
  end
  Sigma := S1 / S2 ;
end

```

最後是 Tobin 模式的核心部份，我們用一個 Newton-Ral-
phson 方法來解它：

程序：tobin (β , σ^2 , M , N , LIMIT , connt , G , H)

參數：

β : 模式的係數 (輸出) ;
 σ^2 : 模式的變異數 (輸出) ;
M : T_i = 0 的個數 ;
N : T_i > 0 的個數 ;

LIMIT : 最大反覆次數;

count : 實際反覆次數(輸出);

G : 在收斂時的一階微分, $\|G\| = 0$;

H : 在收斂時的二階微分的反矩陣, $-H$ 可視為係數之共變異矩陣看待。

計算程序:

begin

OLS (X, T, β , σ^2); { 由一般迴歸算出 β , σ^2 做初值 }

count := 1;

done := FALSE;

while NOT done do

begin

Newton (β , σ^2 , M, N, G, H); { 牛頓法 }

$\bar{\sigma}^2$:= Sigma (β , σ^2 , M, N); { 新的 σ^2 }

if $|(\bar{\sigma}^2 - \sigma^2) / \sigma^2| < 0.00001$

then done := TRUE

else σ^2 := $\bar{\sigma}^2$

end

end

4.3 Poisson 模式

另外一個有用的模式為Poisson模式。如果在某一時間區間中產生的旅次彼此無關，於是旅次產生總數用Poisson分配來解釋也是相當合理的；Poisson分配為

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!}$$

此地 n 為旅次產生數， $P(n)$ 為在平均產生數為 λ 之下，會產生 n 個旅次的機率。因此，Poisson旅次產生模式就在於解釋平均旅次產生數 λ ，而非總數；S.Lerman建議兩種型式，即 $\lambda = X\beta$ ， $\lambda = \exp(X\beta)$ ，此地 X 為觀察值， β 待定係數，所以第 i 戶產生 n_i 個旅次的機率就可以寫成

$$P(n_i) = \frac{(X_i\beta)^{n_i} \exp(-X_i\beta)}{n_i!}$$

或者是

$$\begin{aligned} P(n_i) &= \frac{[\exp(X_i\beta)]^{n_i} \exp(-X_i\beta)}{n_i!} \\ &= \frac{\exp((n_i-1)X_i\beta)}{n_i!} \end{aligned}$$

第一種型式有一個缺點，因為 $X_i\beta$ 可能小於零，與一般迴歸技巧的問題相同；第二個型式固然可以克服這一個缺陷，但是旅次產生呈指數型的方式並非常見，所以下面仍以 $X\beta$ 型為主討論。

Poisson模式是一個機率分配模式，我們所估計出來的是機率分配中的平均數，所以要得到總數，就要使用其它的整合技巧(Aggregation)。

4.3.1 估計技巧

估計 Poisson 模式在計算上要比 Tobin 模式來得簡單。因為 i 的平均旅次產生數為 $X_i\beta$ ，觀察到的實際旅次產生數為 n_i 因此 i 會產生 n_i 個旅次的機率為

$$\frac{(X_i\beta)^{n_i} \exp(-X_i\beta)}{n_i!}$$

所以，對於所有 i 的概似函數為這些機率為連乘積，

$$L(\beta) = \prod_i \frac{(X_i\beta)^{n_i} \exp(-X_i\beta)}{n_i!}$$

求 $L(\beta)$ 的極大值就可以解出 β 的估計值，再用 $X_i\beta$ 就可以求得 i 的平均旅次產生數，所有 $X_i\beta$ 的和就是平均旅次產生總量。

4.3.2 計算方式

求 $L(\beta)$ 的極大值不如求 $\ln(L(\beta))$ 的極大值來得方便，所以把 $L(\beta)$ 取對數，得到的新函數 $\ell(\beta)$ 如下：

$$\begin{aligned}\ell(\beta) &= \ln(L(\beta)) \\ &= \sum_i \{ \ln(X_i\beta)^{n_i} + \ln \exp(-X_i\beta) - \ln(n_i!) \} \\ &= \sum_i n_i \ln(X_i\beta) - \sum_i X_i\beta - \sum_i \ln(n_i!)\end{aligned}$$

上式中最後一項為常數，所以對極值沒有影響，所以我們在計算時的 $\ell(\beta)$ 函數為

$$\ell(\beta) = \sum_i (n_i \ln(X_i\beta) - X_i\beta)$$

4.3.3 計算微分

$\ell(\beta)$ 的型式單純，所以一階與二階微分的計算都極為簡單，首先對 β 微分

$$\begin{aligned}\frac{\partial \ell}{\partial \beta} &= \sum_i \left\{ n_i \frac{1}{X_i \beta} \frac{\partial}{\partial \beta} (X_i \beta) - X_i \right\} \\ &= \sum_i \left(\frac{n_i}{X_i \beta} - 1 \right) X_i\end{aligned}$$

接著是二階微分

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = - \sum_i \left(\frac{n_i}{(X_i \beta)^2} X_i' - X_i \right)$$

由二階微分可以看出，它是負定矩陣 (Negative Definite Matrix)，所以一階微分方向永遠指向極大值。

程序： $\ell(N, \beta, G, H)$

輸入：

N : 觀測值個數；

β : 輸入係數；

G : 輸出一階微分；

H : 輸出二階微分；

函數值： 概似函數值。

計算程序：

begin

$\ell := 0$; $G := 0$; $H := 0$;

for $i := \ell$ to N do

begin

readin (X_i, n_i) ;

$xb := X_i \beta$;

$\ell := (n_i * \text{Ln}(xb) - xb) + \ell$;

$G := G + (n_i / xb - 1) * X_i$;

$H := H - n_i / (xb * xb) * X_i' * X_i$;

end

$\ell := -\ell$;

end

第五章 分配模式

5.1 前言

分配模式是接續在產生模式之後的研究，其實，一些選擇模式（如整合性 Logit 模式）都可以看成為分配模式，但是分配模式在需求模式架構下依然以重力模式為其骨幹。幾乎所有重力模式類似應用中，均不能脫離雙比例平衡法，故本章亦將雙比例平衡法納入討論。雙比例平衡法常被稱為 Furness 法。

古典重力模式以 A.G. Wilson 的極大熵 (Maximum Entropy) 模式最為有名，但極大熵理論需要使用 Sterling 漸近式，而不如最近資訊增益來得直接，故本章中就以後者來解說雙比例平衡法與 Wilson 重力模式。

Wilson 重力模式中的阻力係數型如 $\exp(-\beta C_{ij})$ ，此地 C_{ij} 為 i 到 j 的成本；這種型式的使用範圍相當有限，因此我們介紹一個更為廣泛，把阻力係數定成為 $\exp(Z_{ij}\beta)$ 的廣義重力模式；此地 Z_{ij} 為一向量，可以包含有時間、成本、產生、吸引等因素，這一類型的重力模式在 BPR 等知名度甚高的系統中都曾經用過，但是效力却不如本章介紹的 Gary-Sen 模式來得大。

5.2 雙比例平衡法

雙比例平衡法傳統上都叫做 Furness 法，它的起源在於解決下面的問題：已知矩陣 $[t_{ij}]$ ， $[o_i]$ ， $[d_j]$ ， $1 \leq i \leq M$ ， $1 \leq j \leq N$ ，是否有另一矩陣 $[t^*_{ij}]$ 存在，滿足：

$$t_{ij}^* = a_i b_j t_{ij}$$

$$\sum_j t_{ij}^* = o_i$$

$$\sum_i t_{ij}^* = d_j$$

如果有的話， $[t_{ij}^*]$ 是否為唯一。

今 $T = \sum_{ij} t_{ij}$ ， $T^* = \sum_i o_i = \sum_j d_j$ ；則 $[t_{ij}/T]$ 與 $[t_{ij}^*/T^*]$ 可視為兩個機率分配看待， t_{ij}/T 表示在調整前自 i 到 j 所觀察到的機率，稱為事前機率分配 (Prior Distribution)；而 t_{ij}^*/T^* 為調整後自 i 到 j 的機率，未知，稱為事後機率 (Posterior Distribution)。從事前機率分配因為引入新資訊而導致事後機率發生，其中所引入的資訊以資訊增益 (Information Gain) 度量：

$$\frac{t_{ij}^*}{T^*} \ln \left(\left(t_{ij}^*/T^* \right) / \left(t_{ij}/T \right) \right)$$

總資訊增益為

$$\sum \frac{t_{ij}^*}{T^*} \ln \left(\frac{t_{ij}^*}{T^*} / \frac{t_{ij}}{T} \right)$$

但依雙比例平衡法的假定是“目前與未來相似”的前提，則所導入的總增益應為最小，而且 t_{ij}^* 滿足列和與行和的等式條件，因此可以引入兩組拉氏乘數 (Lagrange Multiplier) λ_i, μ_j ($1 \leq i \leq M, 1 \leq j \leq N$)，而立出一式如下

$$\begin{aligned} \varphi(t_{ij}^*) = & \sum_{ij} \frac{t_{ij}^*}{T^*} \ln \left(\frac{t_{ij}^*}{T^*} / \frac{t_{ij}}{T} \right) - \sum_i \lambda_i \left(\sum_j t_{ij}^* - o_i \right) \\ & - \sum_j \mu_j \left(\sum_i t_{ij}^* - d_j \right) \end{aligned}$$

求此式極值即可得到 t_{ij}^* 。首先對 t_{ij}^* 微分得到

$$\frac{\partial \varphi}{\partial t_{ij}^*} = \frac{1}{T^*} \ln \left(\frac{t_{ij}^*}{T^*} / \frac{t_{ij}^*}{T} \right) + \frac{1}{T^*} - \lambda_i - \mu_j$$

令上式微分為零，解出 t_{ij}^* 為

$$t_{ij}^* = t_{ij} \left(\frac{T^*}{T} \right) e^{-1 + \lambda_i T^* + \mu_j T^*}$$

分別對 λ_i ， μ_j 求微分，令其為零可得

$$\sum_j t_{ij}^* = o_i, \quad \sum_i t_{ij}^* = d_j$$

換言之，Furness 法的式子為資訊增益極小的解。

令 a_i 與 b_j 如下，

$$a_i = \left(\frac{T^*}{T} \right) e^{-1 + \lambda_i T^*}$$

$$b_j = e^{\mu_j T^*}$$

於是 t_{ij}^* 可改寫成標準型

$$t_{ij}^* = a_i b_j t_{ij}$$

但 a_i 與 b_j 的解並非唯一。假設 k 是一個常數，則

$$a'_i = k a_i$$

$$b'_j = b_j / k$$

也滿足 $t_{ij}^* = a'_i b'_j t_{ij}$ ，因此也是一解；在實用上常令某一 a_i (比如 a_1) 為 1，而求得 a_i 與 b_j 的唯一解。

5.2.1 Furness 方法

Furness 觀察到要解這個問題並非一步可得，而需要反覆計算。

令 $T^0 = [t_{ij}]$ ，則

$$\begin{aligned}
 t_{ij}^{(1)} &= \frac{o_i}{\sum_k t_{ik}} t_{ij} \\
 &\vdots \\
 t_{ij}^{(n)} &= \frac{d_i}{\sum_k t_{ik}^{(n-1)}} t_{ij}^{(n-1)} \\
 t_{ij}^{(n+1)} &= \frac{o_i}{\sum_k t_{ik}^{(n)}} t_{ij}^{(n)}
 \end{aligned}$$

因此有連串的矩陣 $T^0, T^2, \dots, T^{n+1}, \dots$ 。在 n 為偶數時，第 j 行的和必為 d_j ，但第 i 列的和未必是 o_i ；經由 $n+1$ （奇數）次修正，第 j 行和的和未必為 d_j ，但第 i 列的和一定是 o_i ，因此平衡行和、平衡列和，交錯的工作一直到兩者均平衡為止。因為這種計算方法，故稱之為雙比例平衡法（Bi-Proportional Balance Method）。

上述的技巧是直接平衡矩陣，沒有 a_i 與 b_j 的值，但經過修飾之後仍然可以得出 a_i 與 b_j 的值。考慮上面 $t_{ij}^{(n)}, t_{ij}^{(n+1)}$ 等式，令 $d_j^{(n-1)} = \sum_k t_{kj}^{(n-1)}$ ， $o_i^{(n)} = \sum_k t_{ik}^{(n)}$ ，於是

$$\begin{aligned}
 t_{ij}^{(n)} &= \frac{d_i}{d_j^{(n-1)}} \cdot t_{ij}^{(n-1)} \\
 t_{ij}^{(n+1)} &= \frac{o_i}{o_i^{(n)}} \cdot t_{ij}^{(n)} = \frac{o_i}{o_i^{(n)}} \left(\frac{d_i}{d_j^{(n-1)}} \cdot t_{ij}^{(n-1)} \right) \\
 &= \left(\frac{o_i}{o_i^{(n)}} \right) \left(\frac{d_i}{d_j^{(n-1)}} \right) \cdot t_{ij}^{(n-1)}
 \end{aligned}$$

如果連續代入 $t_{ij}^{(n-1)}$ ，上式可以改寫成

$$t_{ij}^{(n+1)} = \left(\frac{o_i}{o_i^{(n)}} \cdot \frac{o_i}{o_i^{(n-1)}} \cdots \frac{o_i}{o_i^{(1)}} \right) \left(\frac{d_i}{d_j^{(n-1)}} \cdot \frac{d_i}{d_j^{(n-3)}} \cdots \frac{d_i}{d_j^{(1)}} \right) \cdot t_{ij}$$

如果令 $a_i^{2^n}, b_j^{2^{n-1}}$ 為

$$a_i^{2^n} = \frac{o_i}{o_i^{2^n}} \cdot \frac{o_i}{o_i^{2^{n-2}}} \cdot \dots \cdot \frac{o_i}{o_i^2}$$

$$b_j^{2^{n-1}} = \frac{d_i}{d_j^{2^{n-1}}} \cdot \frac{d_i}{d_j^{2^{n-3}}} \cdot \dots \cdot \frac{d_j}{d_j^1}$$

於是

$$t_{ij}^{2^{n+1}} = a_i^{2^n} \cdot b_j^{2^{n-1}} \cdot t_{ij}$$

因此，當 $n \rightarrow \infty$ 而使 $t_{ij}^{2^{n+1}} \rightarrow t_{ij}^*$ 時， $a_i^{2^n} \rightarrow a_i^*$ ， $b_j^{2^{n-1}} \rightarrow b_j^*$ 就有

$$t_{ij}^* = a_i^* b_j^* t_{ij}$$

這就是標準型。

5.2.2 一些理論上的重要結果

Furness 方法看似簡單，但在實用上會有許多潛在的困難。下面的討論先假定每一個 t_{ij} 均為正 ($t_{ij} > 0$)，令

$$A = \begin{bmatrix} a_1 & & 0 \\ & a_2 & \\ 0 & & \ddots \\ & & & a_m \end{bmatrix} \quad B = \begin{bmatrix} b_1 & & 0 \\ & b_2 & \\ 0 & & \ddots \\ & & & b_n \end{bmatrix}$$

於是 $T^* = [t_{ij}^*]$ ， $T = [t_{ij}]$ 可以寫成

$$T^* = A \cdot T \cdot B$$

有許多人都已經證明在 $T > 0$ 的條件之下， T^* ， A 與 B 為唯一有在，因而 Furness 可以安全作業。但在實用上， T 中的元素未必全部大於零；譬如說，對角線的區內旅次常常為 0，而且如果在調查時無法觀察到 $i \rightarrow j$ 的旅次時， $t_{ij} = 0$ ，因此就違背了上述 $T > 0$ 的條件，從而 T^* ， A 與 B 不一定存在，甚至於往往不會收斂；而且收斂時也不

一定會有正確的結果。舉例而言，若

$$T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, O = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, D = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

於是在平衡完列之後，

$$T^1 = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}$$

再接著平衡行，得到

$$T^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}$$

往後的工作就一直在 T^1 與 T^2 之間交替出現，永遠不會收斂，這就是 T 中有零的第一個反例。

再考慮

$$T = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, O = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, D = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

下面就是各個步驟

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \xrightarrow{\text{列}} \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \xrightarrow{\text{行}} \begin{bmatrix} 0 & \frac{3}{2} \\ 2 & \frac{1}{2} \end{bmatrix} \xrightarrow{\text{列}} \begin{bmatrix} 0 & \frac{2}{5} \\ \frac{8}{5} & \frac{2}{5} \end{bmatrix} \xrightarrow{\text{行}} \begin{bmatrix} 0 & \frac{5}{3} \\ 2 & \frac{1}{3} \end{bmatrix}$$

$$\begin{matrix} * \\ \rightarrow \end{matrix} \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

這個 T 是很自然的，因在對角綫上的區內旅次往往因為收集不到而有 0 的量，上面的計算過程就把對角綫上非零的量變成零。

有許多文獻建議，如果把 T 中的零用一個很小的數去取代，因而使 $T > 0$ ，於是 Furness 方法就一定收斂，而且結果有唯一。但是，這個概念是錯的，考慮

$$T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, O = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, D = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

假設 ϵ 是一個很小的數，而把 T 改寫成 $T(\epsilon)$ 如下，

$$T(\epsilon) = \begin{bmatrix} \epsilon & \epsilon & 1 \\ \epsilon & \epsilon & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

我們可以證明 $T(\epsilon)$ 可以在 Furness 方法處理之下收斂，再令 $\epsilon \rightarrow 0$ ，我們得到

$$\lim_{\epsilon \rightarrow 0} (T^*(\epsilon)) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

我們仍然會令會把原來不為零的對角綫元素變成 0，而把原來為零的元素變成不是零。但若把原來為 0 的地方也用 0 填滿，就得到

$$\begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

這就根本不是正確的結果。

以上所述，說明 T 中元素不全為零時所可能發生的問題，因此對於使用上必須要特別小心。

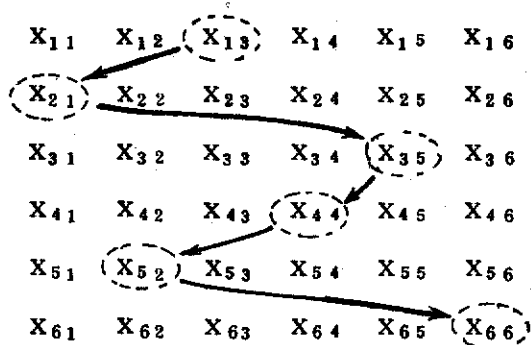
Sinkhorn 曾經就這個問題提出一個結果，讓我們可以確定在什麼情況下，當 T 中元素不全為正（即 $T \geq 0$ ）時，用 Furness 法算出來 T^* 是否合理。

令 σ 是一個置換（Permutation）， $\sigma(i)$ 表示經過置換後第 i 個位置的值；比如說

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 5 & 4 & 2 & 6 \end{pmatrix}$$

則 $\sigma(1)=3$ ， $\sigma(2)=1$ ， $\sigma(3)=5$ ， $\sigma(4)=4$ ， $\sigma(5)=2$ ， $\sigma(6)=6$

。對於一個方陣 $X = [x_{ij}]$ 而言， $x_{1\sigma(1)}$ ， $x_{2\sigma(2)}$ ， \dots ， $x_{n\sigma(n)}$ 叫做一條對角線，以上面例子而言，對角線就是下圖有虛線的六個數：



而 $\sigma(1)=1$, $\sigma(2)=2$, \dots , $\sigma(n)=n$ 就是主對角綫，亦即自左上角到右下角的 n 個值。

Sinkhorn 的結果是這樣的：

- (1) 如果 T 中的每一個正的元素都在某一條對角綫上，則 $T^* = A \cdot T \cdot B$ 不但存在，而且唯一。
- (2) 如果 T 中有一條對角綫，其上每一個元素都為正數，則 Furness 方法會收斂，但不能把收斂的結果表示成 $T^* = A \cdot T \cdot B$ 的型式。

回顧上面的例子，

$$T_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad T_2 = \begin{bmatrix} 0 & 1 \\ & \\ 1 & 1 \end{bmatrix}$$

T_1 完全沒有正對角綫，所以 T_1^* 不存在； T_2 有一條正對角綫（右上到左下），但右下角元素不在任何正對角綫上，所以 T_2 會收斂，但無法表示成 $T_2^* = A \cdot T \cdot B$ 的型式。

所以，在 Furness 法中應該有一個段落，檢查是否有一條正對角綫，避免永不收斂；至於 $A \cdot T \cdot B$ 的型式並非必需，所以使用傳統的形式計算即可。

5.3 重力模式

傳統重力模式並沒有一套完整的理論架構，只是一些實驗性的公式，而且產生、吸引的條件並非一定能夠滿足，或是不一定收斂。

A.G. Wilson 曾經提出一個修正的重力模式，把旅行成本考慮在內；令 $C = [c_{ij}]$ 表示 i 到 j 的旅行成本，而 C^* 表示系統中的總旅行成本，Wilson 建議

$$t_{ij}^* = a_i b_j \exp(-\beta c_{ij})$$

$$\sum_j t_{ij}^* = o_i$$

$$\sum_i t_{ij}^* = d_j$$

$$\sum_{i,j} c_{ij} t_{ij}^* = C^*$$

因此我們需要解出的除了 a_i, b_j 之外還有一個 β 。

Wilson 重力模式也可以使用資訊增益理論得來，與前述相同，利用拉氏乘數法則可得

$$\begin{aligned} \varphi = \sum_{i,j} \frac{t_{ij}^*}{T^*} \ln \left(\frac{t_{ij}^*}{T^*} / \frac{t_{ij}}{T} \right) - \sum_i \lambda_i \left(\sum_j t_{ij}^* - o_i \right) \\ - \sum_j \mu_j \left(\sum_i t_{ij}^* - d_j \right) \\ - h \left(\sum_{i,j} c_{ij} t_{ij}^* - C^* \right) \end{aligned}$$

對 t_{ij}^* 求微分得到

$$\frac{\partial \varphi}{\partial t_{ij}^*} = \frac{1}{T^*} \ln \left(\frac{t_{ij}^*}{T^*} / \frac{t_{ij}}{T} \right) + \frac{1}{T^*} - \lambda_i - \mu_j - h c_{ij}$$

令上式為零，解出 t_{ij}^* 如下，

$$t_{ij}^* = t_{ij} \left(\frac{T^*}{T} \right) \exp(-1 + \lambda_i T^* + \mu_j T^* + h c_{ij})$$

但是，重力模式的做法却與 Furness 法不同，它的想法是把 t_{ij}^* 用 $a_i b_j \exp(-\beta c_{ij})$ 表現出來，因而 t_{ij} 為 1。因此，令

$$a_i = \left(\frac{T^*}{T} \right) \exp(-1 + \lambda_i T^*)$$

$$b_j = \exp(\mu_j T^*)$$

$$\beta = -h$$

所以很明顯地，

$$t_{ij}^* = a_i b_j \exp(-\beta c_{ij})$$

5.3.1 計算方式

本節討論如何解出 a_i, b_j 與 β 的技巧。對於任何一個 β 而言，可以利用 $[c_{ij}]$ 而算出一個新矩陣 $[\exp(-\beta c_{ij})]$ ，再以這個矩陣為主，利用 Furness 法加以平衡而得到 \bar{a}_i, \bar{b}_j 與 \bar{t}_{ij} ：

$$\bar{t}_{ij} = \bar{a}_i \bar{b}_j \exp(-\bar{\beta} c_{ij})$$

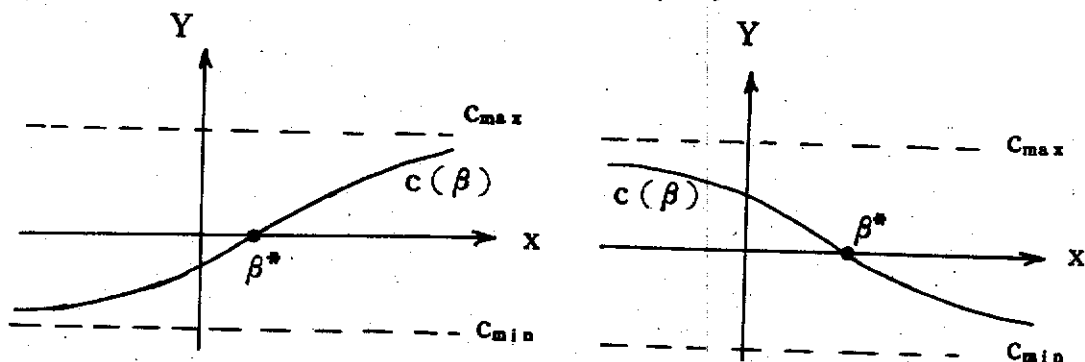
$$\sum_j \bar{t}_{ij} = o_i$$

$$\sum_i \bar{t}_{ij} = d_j$$

但是 $\sum c_{ij} \bar{t}_{ij}$ 却不一定是 C^* ；換言之， $\sum c_{ij} t_{ij} - C^* \neq 0$ 。因為 $\sum c_{ij} \bar{t}_{ij} - C^*$ 的值受到我們對 β 的假設，因此我們可以立出一個函數

$$c(\beta) = \sum_{i,j} c_{ij} \bar{t}_{ij} - C^*$$

A. W. Evans 曾經證明過 $c(\beta)$ 有一個極大值 c_{\max} ，有一個極小值 c_{\min} ，而且 $c(\beta)$ 為一個可以微分的遞增或遞降函數；換言之， $c'(\beta) \neq 0$ 。更重要的是 $c_{\max} \geq 0 \geq c_{\min}$ ，因此 $c(\beta)$ 一定有解，換句話說，唯一存在有一個 β^* 使得 $c(\beta^*) = 0$ 。



因為 $c(\beta)$ 中只有一個未知數 β ，雖然計算函數值不很容易，所以解出 β^* 其實是就是一個尋根的工作。雖然 $c'(\beta)$ 存在，但計算相當繁雜，計算 $c(\beta)$ 與 $c'(\beta)$ 的值所花費的時間是非常可觀的；另外，在我們的工作中發現，有許多實際工作中模式在根附近的斜率非常大，因而使 $c'(\beta)$ 趨於無窮，也使我们放棄使用 Newton 方法；甚至於也不使用割綫法。所以，能夠使用、而且一定安全的方法就是二分法了，雖然速度會比較慢。

二分法的動作分成兩個階段，第一階段是找出兩個點 a 與 b ，使得 $c(a)$ 與 $c(b)$ 異號，於是在 $[a, b]$ 之間就必有一個根；

第二階段每次取區間中點來測試，然後將區間減半，直到找到根或區間長度已小於誤差範圍為止。一般而言，尋根時的誤差範圍不容過小，以免造成時間的浪費，通常 0.01 到 0.0001 就已經很夠用。

尋根第一階段自 $\beta = 0$ 起，求 $c(0)$ 的值。在 $\beta = 0$ 時， $\exp(-0 \cdot c_{ij}) = 1$ ，因此矩陣為 $[o_i d_j / k]$ ， $k = \sum o_i = \sum b_j$ ，因此

$$c(0) = \sum_{i,j} \frac{o_i d_j}{k} \cdot c_{ij} - C^*$$

對於其它的 β 而言，先計算出矩陣 $[\exp(-\beta c_{ij})]$ ，再用 $[o_i]$ 與 $[d_j]$ 以 Furness 法平衡而導出 $[t_{ij}]$ 。最後計算 $c(\beta)$ 如下：

$$c(\beta) = \sum_{i,j} \tilde{t}_{ij} \cdot c_{ij} - C^*$$

程序：bound (a, fa, b, fb, beta, found, ϵ)

參數：a : 左邊端點。

fa : a 的函數值。

b : 右邊端點。

fb : b 的函數值。

beta : 當 found 為 TRUE 時，表示 $c(\beta) = 0$ 的根；

found: TURE 表示已解出根，FALSE 表示未解出根，而根在 $[a, b]$ 之間；

ϵ : 容許誤差，當 $|c(\beta)| < \epsilon$ 時， β 就認為是一個根。

計算程序：

begin

$a := 0 ; fa := \sum \left(\frac{a_i b_j}{k} \right) * c_{ij} - C^* ;$

$b := 1 ; fb := c(b) ;$

$\Delta := 1 ;$

found:=FALSE;

if $fa * fb > 0$ then

begin

if $fa > 0$ then

if $fa > fb$ then

begin

$a := b ; fa := fb ;$

end

else $\Delta := -1$

else

if $fa > fb$ then

$\Delta := -1$

else

begin

$a := b ; fa := fb ;$

end ;

$\Delta := \Delta + \Delta ; b := a + \Delta ; fb := c(b) ;$

while $fa * fb > 0$ do

begin

$\Delta := \Delta + \Delta ; a := b ; fa := fb ;$

$b := a + \Delta ; fb := c(b) ;$

~ 100 ~

```

        end ;
    end ;
    if  fa=0  then
        begin
            beta := a ; found := TRUE ;
        end ;
    if  fb = 0  then
        begin
            beta := b ; found := TRUE ;
        end ;
    end ;
end ;

```

程序 : bisect (a, fa, b, fb, ϵ , beta)

參數 :

a , fa : 一邊端點及函數值。
 b , fb : 另一邊端點及函數值。
 ϵ : 容許誤差。
 beta : c (b) 的根

計算程序 :

```

begin
    c : ( a + b ) / 2 ; { 中點 }
    fc := ( b )
    while NOT ( | fc | <  $\epsilon$  ) do
        begin
            if fa * fc < 0 then

```

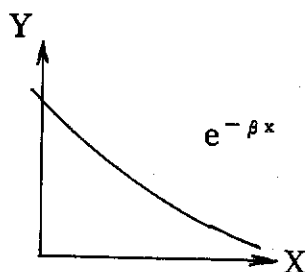
```

begin
    b := c ;
    fb := fc ;
end
else
begin
    a := c ;
    fa := fc ;
end
c := (a + b) / 2 ;
fc := c ( b ) ;
end ;
end ;

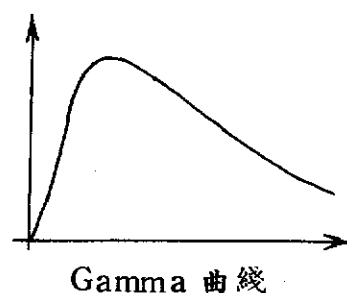
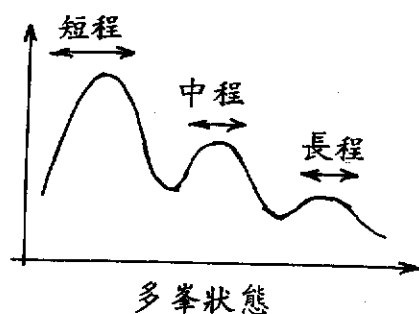
```

5.4 廣義重力模式

A.G. Wilson 的重力模式採用負指數曲線，



但依以往的經驗，運輸走廊的成本函數與旅次量的關係常呈不規則分佈，常見的是三峯的型態；左峯為極短程旅次，右峯為極長程旅次，而中央峯則通常是在兩端與運輸走廊中點的旅次（見下左圖）；在一些都會區中却又往往是 Gamma 曲線（見下右圖），因而 $e^{-\beta x}$ 就無法充分解釋這一類型的現象。



1983 年，R.H.Gray 與 A.K.Sen 提出了一個一般化的重力模式架構，容許在模式中使用更多解釋變數。Gray-Sen 的模式為

$$t_{ij} = a_i b_j \exp (Z_{ij} \beta)$$

$$\sum_j t_{ij} = o_i$$

$$\sum_i t_{ij} = d_j$$

此地 Z_{ij} 是 i 與 j 之間的解釋變數，可以不止一個。因為總旅行成本不易估計，所以在 Gray-Sen 模式中去除了總旅行成本的限制條件。如果 Z_{ij} 為單一成本 C_{ij} ，則模式的指數部份就變成 Wilson 模式，如果 Z_{ij} 定義成 $[\ln C_{ij}, C_{ij}]$ ，則為 BPR 模式中的 Gamma 曲線，因為

$$\exp(a \cdot \ln C_{ij} + b C_{ij}) = C_{ij}^a \exp(b C_{ij}), \quad \beta = \begin{bmatrix} a \\ b \end{bmatrix}$$

因為 $Z_{ij} \beta$ 是一個比較廣泛的解釋因素，我們可以使用兩地的吸引，產生，成本，時間，方便性，可達性等因素，從而重力模式除了不能改變其根本結構之外，幾乎可以用任何有意義的資料來解釋 i 與 j 之間的旅次費。

5.4.1 Gray-Sen 的理論

估計 Gray-Sen 的模式並不是一件困難的工作。因為

$$t_{ij} = a_i b_j \exp(Z_{ij} \beta)$$

所以

$$\frac{t_{ij}}{t_{ii}} = \frac{b_j}{b_i} \cdot \frac{\exp(Z_{ij} \beta)}{\exp(Z_{ii} \beta)}$$

$$\frac{t_{ji}}{t_{jj}} = \frac{b_i}{b_j} \cdot \frac{\exp(Z_{ji} \beta)}{\exp(Z_{jj} \beta)}$$

兩者相乘，則得

$$\begin{aligned} \frac{t_{ij} \cdot t_{ji}}{t_{ii} \cdot t_{jj}} &= \frac{\exp(Z_{ij} \beta) \cdot \exp(Z_{ji} \beta)}{\exp(Z_{ii} \beta) \cdot \exp(Z_{jj} \beta)} \\ &= \exp((Z_{ij} + Z_{ji} - Z_{ii} - Z_{jj}) \beta) \end{aligned}$$

這樣就全部把平衡因子 a_i 與 b_j 消去。如果把上式取對數，

$$\ln \left(\frac{t_{ij} \cdot t_{ji}}{t_{ii} \cdot t_{jj}} \right) = (Z_{ij} + Z_{ji} - Z_{ii} - Z_{jj}) \beta$$

則很顯然地是一個線性迴歸模式，自變數為 $Z_{ij} + Z_{ji} - Z_{ii} - Z_{jj}$ ，應變數為 $\ln \left(\frac{t_{ij} \cdot t_{ji}}{t_{ii} \cdot t_{jj}} \right)$ ，從而估計出 β ，於是就得到一個矩陣 $[\exp(Z_{ij} \beta)]$ ，再以 $[o_i]$ ， $[d_j]$ 為限制條件，經 Furness 法平衡後就可以得到 $[t_{ij}^*]$ ， a_i 與 b_j 。但是，解 β 就與整合性的對數機率比模式一樣，它並非一個能夠滿足迴歸模式條件的問題。下面的幾節就討論這些問題。

5.4.2 偏差

t_{ij} 是 i 到 j 旅次量的觀測值，令 x_{ij} 為 i 到 j 的理論值（為未知），則 $x_{ij} = t_{ij} + \epsilon_{ij}$ ，此地 t_{ij} 為一隨機數，表示誤差項，當然 x_{ij} 也是一個隨機數，因此 $E(x_{ij}) = t_{ij}$ 。一個很合理的假設，就是 ϵ_{ij} 是一個 Poisson 分配，因此，

$$E(\epsilon_{ij}) = 0, \text{ 且 } \text{Var}(\epsilon_{ij}) = E(\epsilon_{ij}^2) = x_{ij}$$

上一節末的迴歸模式中牽涉到把應變數取對數的技巧，但是因為 $E(\ln x) \neq \ln E(x)$ ，所以會引起偏差。換句話說，當我們用 $\ln \left(\frac{t_{ij} \cdot t_{ji}}{t_{ii} \cdot t_{jj}} \right)$ 做應變數執行迴歸的技巧，而得一應變數估計值 y 時， $\exp(y)$ 並不是 $(t_{ij} \cdot t_{ji}) / (t_{ii} \cdot t_{jj})$ 的不偏估計量。因此，經對數轉換後，必須要有一個方法來平衡（或說是降低）這一項偏差。最簡單而且最常用的方法就是加上 0.5。

考慮 $E(x_{ij})$ ，使用對數函數的無窮級數，我們有

$$\begin{aligned} E(\ln x_{ij}) &= E(\ln(t_{ij} + \epsilon_{ij})) \\ &= E\left(\ln\left(t_{ij} \cdot \left(1 + \frac{\epsilon_{ij}}{t_{ij}}\right)\right)\right) \end{aligned}$$

$$\begin{aligned}
&= E \left(\ln t_{ij} + \ln \left(1 + \frac{\epsilon_{ij}}{t_{ij}} \right) \right) \\
&= E \left(\ln t_{ij} \right) + E \left(\ln \left(1 + \frac{\epsilon_{ij}}{t_{ij}} \right) \right) \\
&= E \left(\ln t_{ij} \right) + E \left(\frac{\epsilon_{ij}}{t_{ij}} - \frac{\epsilon_{ij}^2}{2 t_{ij}^2} + \frac{\epsilon_{ij}^3}{3 t_{ij}^3} - \dots \right) \\
&= E \left(\ln t_{ij} \right) + \frac{E(\epsilon_{ij})}{t_{ij}} - \frac{E(\epsilon_{ij}^2)}{2 t_{ij}^2} + \dots \\
&\approx E \left(\ln t_{ij} \right) - \frac{1}{2 t_{ij}} \quad (\text{注意 } E(\epsilon_{ij}) = 0, E(\epsilon_{ij}^2) = t_{ij}) \\
&\approx E \left(\ln t_{ij} \right)
\end{aligned}$$

但是如果考慮加上 0.5 後的結果，我們有下面的式子，

$$\begin{aligned}
E \left(\ln \left(x_{ij} + \frac{1}{2} \right) \right) &= E \left(\ln \left(t_{ij} + \epsilon_{ij} + \frac{1}{2} \right) \right) \\
&= E \left(\ln \left(t_{ij} \cdot \left(1 + \frac{\epsilon_{ij} + \frac{1}{2}}{t_{ij}} \right) \right) \right) \\
&= E \left(\ln t_{ij} \right) + E \left(\frac{t_{ij} + \frac{1}{2}}{t_{ij}} - \frac{(t_{ij} + \frac{1}{2})^2}{2 t_{ij}^2} + \dots \right) \\
&= E \left(\ln t_{ij} \right) + \frac{E(\epsilon_{ij}) + \frac{1}{2}}{t_{ij}} \\
&\quad - \frac{E(\epsilon_{ij}^2) + E(\epsilon_{ij}) + \frac{1}{4}}{2 t_{ij}} + \dots \\
&= E \left(\ln t_{ij} \right) + \frac{1}{2 t_{ij}} - \frac{1}{2 t_{ij}} - \frac{1}{8 t_{ij}^2} + \dots
\end{aligned}$$

$$\approx E(\text{Ln } t_{ij}) - \frac{1}{8 t_{ij}^2}$$

如果 t_{ij} 充分大， $E(\text{Ln } x_{ij})$ 與 $E(\text{Ln}(x_{ij} + 0.5))$ 中的二次項可以略而不計。在 $E(\text{Ln } x_{ij}) \approx E(\text{Ln } t_{ij}) - 1/(2 t_{ij})$ 中， $1/(2 t_{ij})$ 顯然要比 $E(\text{Ln}(x_{ij} + 0.5)) \approx E(\text{Ln } t_{ij}) - 1/(8 t_{ij}^2)$ 中的 $1/(8 t_{ij}^2)$ 來得大，所以 $E(\text{Ln}(x_{ij} + 0.5))$ 更接近於 $E(\text{Ln } t_{ij})$ 。

因為

$$\text{Ln}\left(\frac{t_{ij} \cdot t_{ji}}{t_{ii} \cdot t_{jj}}\right) = \text{Ln}(t_{ij}) + \text{Ln}(t_{ji}) - \text{Ln}(t_{ii}) - \text{Ln}(t_{jj})$$

四項都用 Ln 來轉換，因此比照上述，在 t_{ij} 充分大的情況下（比如 $t_{ij} > 1$ ），我們可以把每一個 $\text{Ln}(t_{ij})$ 換成 $\text{Ln}(t_{ij} + 0.5)$ ，以求降低偏差；所以，令

$$Z_{ij} = \text{Ln}(t_{ij} + 0.5) + \text{Ln}(t_{ji} + 0.5) - \text{Ln}(t_{ii} + 0.5) - \text{Ln}(t_{jj} + 0.5)$$

而用 Z_{ij} 做應變數。

但是，若 t_{ij} 並不很大時又如何呢？我們在下面會有一節討論。

5.4.3 異質性

當 t_{ij} 充分大時，我們不難大略估計 $\text{Ln}(x_{ij} + \frac{1}{2})$ 的變異數。因為 t_{ij} 充分大，而前述又以說明

$$E(\text{Ln}(x_{ij} + \frac{1}{2})) \approx E(\text{Ln } t_{ij}) - \frac{1}{8 t_{ij}^2}$$

$$\approx E(\text{Ln } t_{ij})$$

$$= \text{Ln } t_{ij}$$

所以變異數 $\text{Var} \left(\text{Ln} \left(x_{ij} + \frac{1}{2} \right) \right)$ 為

$$\text{Var} \left(\text{Ln} \left(x_{ij} + \frac{1}{2} \right) \right) = E \left(\left[\text{Ln} \left(x_{ij} + \frac{1}{2} \right) - E \left(\text{Ln} \left(x_{ij} + \frac{1}{2} \right) \right) \right]^2 \right)$$

$$\approx E \left(\left[\text{Ln} \left(x_{ij} + \frac{1}{2} \right) - \text{Ln } t_{ij} \right]^2 \right)$$

$$= E \left(\left[\text{Ln} \left(t_{ij} + \epsilon_{ij} + \frac{1}{2} \right) - \text{Ln } t_{ij} \right]^2 \right)$$

$$= E \left(\left[\text{Ln} \left(t_{ij} \left(1 + \frac{\epsilon_{ij} + \frac{1}{2}}{t_{ij}} \right) \right) - \text{Ln } t_{ij} \right]^2 \right)$$

$$= E \left(\left[\text{Ln } t_{ij} + \text{Ln} \left(1 + \frac{\epsilon_{ij} + 0.5}{t_{ij}} \right) - \text{Ln } t_{ij} \right]^2 \right)$$

$$= E \left(\left[\text{Ln} \left(1 + \frac{\epsilon_{ij} + 0.5}{t_{ij}} \right) \right]^2 \right)$$

上式展開，因 t_{ij} 充分大，因而可以不理會 $\frac{1}{t_{ij}}$ 以後的項，從而得到

$$\text{Var} \left(\text{Ln} \left(x_{ij} + \frac{1}{2} \right) \right) \approx E \left(\left[\frac{\epsilon_{ij} + 0.5}{t_{ij}} \right]^2 \right)$$

$$= E \left(\frac{\epsilon_{ij}^2 + t_{ij} + 0.25}{t_{ij}^2} \right)$$

$$\approx \frac{E(\epsilon_{ij}^2) + E(\epsilon_{ij})}{t_{ij}^2}$$

$$= \frac{E(\epsilon_{ij}^2)}{t_{ij}^2}$$

$$= \frac{1}{t_{ij}}$$

因此可以看到當 t_{ij} 加大時， $\ln(x_{ij} + \frac{1}{2})$ 變小，而當 t_{ij} 降低時 $\ln(x_{ij} + \frac{1}{2})$ 增加，換言之，應變數項 Z_{ij} 的變異數與

$$\frac{1}{t_{ij}} + \frac{1}{t_{ji}} + \frac{1}{t_{ii}} + \frac{1}{t_{jj}}$$

成反比，因此不滿足迴歸模式中對變異數為常數的假定，所以要用加權 (Weight) 的迴歸技巧來處理它；而權數就是

$$\frac{1}{\sqrt{\frac{1}{t_{ij}} + \frac{1}{t_{ji}} + \frac{1}{t_{ii}} + \frac{1}{t_{jj}}}}$$

如果 t_{ij}, t_{ji}, t_{ii} 與 t_{jj} 有一個趨近於 0，則上式的權數也趨於零，因此對應的 Z_{ij} 也就等於不發生作用，因而 Z_{ij} 可以省略不計。

5.4.4 較小的 t_{ij}

固然 $t_{ij} + 0.5$ 可以取對數以計算 Z_{ij} ，但是當 t_{ij} 很小或為零時，就不能充分地彌補 $E(\ln x)$ 所帶來的偏差。不過當 t_{ij} 很小或零時，乘到 Z_{ij} 去的權數也不大，所以對迴歸模式的影響也不大；而且當 t_{ij} 為零時，會把 Z_{ij} 去除，因此就根本沒有影響。然而，當有大量的 t_{ij} 值很小或為零時，使用迴歸（也就是 Gray-Sen）模式就不是很理想的技巧了。

5.4.5 估計方法

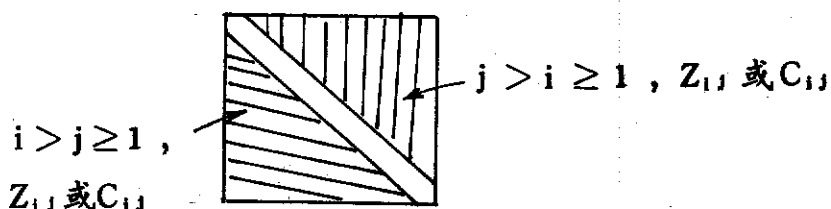
如前所述， $Z_{ij} = \ln(t_{ij} + 0.5) + \ln(t_{ji} + 0.5) - \ln(t_{ii} + 0.5) - \ln(t_{jj} + 0.5)$ ，同理，自變數數也要做相對的轉換，令

$$C_{ij} = Z_{ij} + Z_{ji} - Z_{ii} - Z_{jj}$$

因此迴歸模式為

$$Z_{ij} = C_{ij} \beta + \epsilon_{ij}$$

但是， $Z_{ij} = Z_{ji}$ ， $C_{ij} = C_{ji}$ ；而且 $Z_{ii} = C_{ii} = 0$ ，所以我們只需要考慮比一半還要少的資料即可，亦即考慮上半或下半三角形即得，前者為 $j > i \geq 1$ ，後者為 $i > j \geq 1$ 。



在計算 Z_{ij} ($j > i \geq 1$) 時可以同時算出權數 W_{ij} 如下：

$$W_{ij} = \frac{1}{\sqrt{\frac{1}{t_{ij}} + \frac{1}{t_{ji}} + \frac{1}{t_{ii}} + \frac{1}{t_{jj}}}}$$

當有任一 $t_{ij}, t_{ji}, t_{ii}, t_{jj}$ 為零，就刪除 Z_{ij} 與 C_{ij} 這一對觀測值，最後，利用

$$\left[\frac{Z_{ij}}{W_{ij}} \right] = \left[\frac{C_{ij}}{W_{ij}} \right] \beta + \epsilon_{ij}$$

來估計出 β 。

有了 β 之後，如前所述而計算矩陣 $[\exp(X_{ij} \beta)]$ ，以 $[o_i]$ 與 $[d_j]$ 平衡之，而得到 $[t_{ij}^*]$ ，這就是結果。

5.4.6 計算方法

程序：build (M, N, $[t_{ij}]$, $[X_{ij}]$, beta)

輸入：

M 矩陣列數；
N 矩陣行數；
 t_{ij} 旅次矩陣；
 X_{ij} 解釋 i 到 j 的變數，可能為向量；

輸出：

beta 輸出的模式係數

工作用：

Z_k 工作用行向量，至多 $N * (M - 1) / 2$ 個元素；
 C_k 工作用矩陣，至多 $N * (M - 1) / 2$ 列，各列中元素與 X_{ij} 的個數相同。

計算程序：

begin

$k := 0$;

 for $i := 1$ to $M - 1$ do

 for $j := i + 1$ to N do

 begin

 if $t_{ij} * t_{ji} * t_{ii} * t_{jj} < > 0$ then

 begin

k

$W := 1 / \sqrt{1/t_{ij} + 1/t_{ji} + 1/t_{ii} + 1/t_{jj}}$;

$z_k := \ln(t_{ij} + 0.5) + \ln(t_{ji} + 0.5)$
 $- \ln(t_{ii} + 0.5) - \ln(t_{jj} + 0.5)$;

$z_k := z_k * W$;

$c_k := (x_{ij} + x_{ji} - x_{ii} - x_{jj}) * W$;

 end ;

•
end ;
beta := (c' . c)⁻¹ . (c' . z) ;
end ;

第六章 結論

如第一章所述，本計畫重點並不在於如何應用這些模式，而是放在發展這些模式體系上，但是在發展過程中却發現相當程度的困難。在開始時，所有軟體的規格全是以 IBM 大型電腦 VM/370-CMS 構想，後來轉移到 MVS 系統，但是因為行政院主計處電腦中心的作業方式以及系統反應速度並不宜從事這一類型的發展工作，因而又移植到 IBM PC 上。但是 IBM PC 的記憶體容量，CPU 速度都有相當限制，所以系統的規格又不得不迎合新機種而再次更動，是故原來的設計構想在 IBM PC 系統中早已消失，因而時間上產生浪費。基於此，有幾點結論與建議如下：

1 軟體發展（如果需要做的話）是一件長期性的工作，需要有長期性的策略。軟體發展可分成三方面討論。第一是收集它處軟體應該要有計畫、有固定預算、有專人（可能分在各組）負責、有定期的通訊報告測試評估；目前本所在這一方面自陷於微電腦，策略固不能說不對，但絕非正確途徑，蓋微電腦因為機器架構上的限制，固然速度與記憶體容量能夠與迷你電腦相比，但總工作量到目前為止，因為生產成本關係，仍然與大型或迷你電腦相去甚遠，其優點不過方便而已。所以，收集工作應大小機種並重，萬不可有目前微電腦足敷所需而排斥更上層樓的心態。第二是發展應用軟體的長程規劃。以往發展應用軟體總是在計畫開始之後發起，而在計畫結束之前終結，以致於應用軟體設計不良、考慮不週，生命期極端短暫，計畫結束後形同廢棄，殊為可惜，所以針對這一類問題，對於計畫中使用或發展的軟體，在正確性、可用性、移植性、延伸性上應該有所考核，報告書應有相關文獻、流程，使用方式等說明，以免流失

，並且於計畫結束後，酌情予以繼續發展，延長軟體生命期。第三是發展具有前瞻性或實驗性的新軟體，前兩點是依實際需求而生，但是問題在於這一類型軟體總是比較傳統、古典，而趨於保守；在從事規劃工作時，如果使用同一階層，類似理論背景的軟體所得來的結果總是相去不遠的，因之在做決定時可能已經引入偏見而不自知，所以如果有一些前瞻性、新穎的軟體負擔部份測試與對比的工作，那麼相比較之後（新方法當然有新理論背景，故可印證舊理論，或得出對比）所下的結論可能就比較踏實，或者是也可以從新、舊的比較而發現新問題與新答案。故謂前瞻性與實驗性的新軟體是不可或缺的，也應該有一定程度的助力、經費與支持；但這類型軟體無需過多，蓋新理論在尚未充分實證之前，仍然不過是理論而已。

2. 建立軟體發展策略、程序與文獻標準，以往我們完全沒有一個可供遵行的標準，所以曾經發展出來的軟體程式缺乏明確目標，可能只能處理某一特殊問題而不能做更為開拓性的應用；也沒有程序與文獻標準，因此品質不但難以控制，而且在維護上也相當困難。如果本所有這樣的策略與標準，則在軟體發展上就會比較有例可循，也可以有考核的基準與維護上的依賴；我們應該體認到維護與生產並重的準則。
3. 建立寫作程序與各項標準規格，這一點是承襲上一點而來；縱使沒有大型系統，但至少目前的微電腦系統是可以先行辦理的。這一項建議重點在於：本所的標準作業系統為何？制式程式語言為何？程式寫作格式（Style）為何？註解方式為何？輸出／入界面為何？程式文獻為何？等等問題。選定標準作業系統會日益複雜，目前只有MS-DOS一項，但隨著IBM推出PS/2系列機種，以及OS/2與即將成為標準的UNIX系統，如果不能高瞻遠慮而定下標準，而任由發展人自行選定，則最終必然會走上軟體各不相容的地步，到時勢必常做移植工作，費時費力，而且也可能使軟體能力打折扣。制

式程式語言的重要性並不低於一個作業系統，各個程式的靜態、動態環境不一定相同，與作業系統的溝通方式互異，語言的設計目標也不一樣，因此本所在軟體發展上就應該以長遠目標選定一套合用的制式語言與它的輔助語言，以便利程式的流通與維護；除了選定語言之外，還需要選擇一個編譯程式，需知A廠與B廠的同一語言的編譯程式並不一定能夠相互溝通，而且其它廠牌的應用程式也有偏好某幾個編譯程式的傾向。以本所過往的發展看來，以Fortran與C做為主力語言的可能性是很大的，但亦不應排斥在未來使用Ada的可能。

4. 硬體方面亦應從速訂立準則。目前的現象是大眾只知道有IBM的系列機型，對其它廠牌機器的規格、性能知之甚少；但縱使是IBM PS / 2—80，使用Intel 80386 / 80387，其性能也未必會較更高一層面的微電腦工作站來得好。因之，我們應該有一個標準，依功能與作業內容劃分，到了某一層面之後，應該考慮到更為強力的機種，如Apollo，Sun，Micro Vax等；就發展軟體來看，一旦進入UNIX，則可用的工具將較使用PC級機種來得更有彈性、更豐富。不過，這些高層次機種不必多，而且在發展軟體時尚須注意移植的問題；至於UNIX與IBM大型機種、個人電腦之間的連繫，目前已經沒有任何困難。總之，硬體的採購不宜漫無目標，隨使用人要求，而應就機型、所處理的業務、是否為專業性用途等等，納入一個整體性的架構中考慮。

5. 目前本所正逐漸引入資料庫（Data Base）的技術，唯所有訓練都只注重在某些資料庫軟體（如dBASE）的操作方法與程式寫作，而較少就資料庫原理、規劃、設計、溝通上著手，因而建立的資料庫固有資料庫之實，却無資料庫的精神。事實上，國內的電腦訓練普遍存在這個現象，因此，本所往後舉辦、參加的各式訓練，應該以基礎訓練為主，而把技術層面的課題放到能夠掌握基本觀念之後，

如此才能避免見木不見林的後果。因為資料庫的建立固然由少數人進行，但是其影響却很大，甚至於全所，因此讓專業者與非專業者有正確的認識至關重要。

以上五點均為進行計畫時所遇到的困難，或是在可想見的將來可能會出現的問題，相信若能及早策劃、解決，對本所建立資料庫系統與未來的軟體發展均有莫大助益。