

從費氏臨床試驗談近代生物統計

張憶壽·熊 昭

一. 楔子

在許多人的觀念中，凡是與醫學、藥學、動物、植物等有關的隨機現象均在生物統計 (Biostatistics) 的知識範圍之內。從這個角度來看，我們當能理解生物統計給人的拼盤印象。這情形與理論數學中的主題鮮明，層次井然固然不同，而與統計學中某些分支所表現的精雕細琢也不甚一樣。但是這種拼盤印象似乎逐漸成爲過去了。生物統計學的內容在過去三、四十年間有了相當的充實與改變。這種充實與改變一方面是基於醫藥研究上的急切需要，另一方面則是由於近代機率論的適時成長。而代表這種充實與改變的兩個主要分支便是存活分析與序列分析。

在本文裡，我們要先介紹一個著名的費氏白血病患存活數據，然後指出要能充分利用類似這樣的數據而得出科學性結論便是當代存活分析與序列分析的主要目標。希望這個簡單的介紹能夠幫助具有大學乃至高中數學程度的人士瞭解生物統計的要旨。

二. 費氏白血病患存活數據

費氏 (Freireich) 等人在 1963 年報告了一個關於白血病患的存活數據。這套數據給許多生物統計學家研究過。我們現在參考 Andersen, Borgan, Gill & Keiding (1993) 及 Siegmund (1985) 來陳述這套數據。

當時，在美國的十一家醫院裡，有 62 個患有急性白血病 (acute leukemia) 的病患，這些病患是在接受類固醇 (prednisone) 治療後得到部分或完全緩解的小孩。緩解 (remission) 在這是指骨髓中的病況消失。這 62 個病患接受了一個緩解維持治療方案 (remission maintenance therapy)。它的做法是這樣的：以醫院爲單位，把在同一醫院內的這些病患依照緩解狀況加以配對，然後在每一對中隨機抽出一位給他施予 6-MP 治療，而剩下的一位則給他安慰劑 (placebo)。

下表所列的是 21 個完整配對的緩解時段數據。緩解時段 (remission length) 是指病患配對後 (約爲緩解發生時刻) 至白血病復發時刻的時段長度。譬如說，第 5 對病患中用安慰劑的是 8 週後開始復發，用 6-MP 的是 22 週後復發。

對(<i>i</i>)	安慰劑	6-MP	勝負	淨贏 $W(i)$
1	1	10	1	1
2	22	7	-1	0
3	3	32+	1	1
4	12	23	1	2
5	8	22	1	3
6	17	6	-1	2
7	2	16	1	3
8	11	34+	1	4
9	8	32+	1	5
10	12	25+	1	6
11	2	11+	1	7
12	5	20+	1	8
13	4	19+	1	9
14	15	6	-1	8
15	8	17+	1	9
16	23	35+	1	10
17	5	6	1	11
18	11	13+	1	12
19	4	9+	1	13
20	1	6+	1	14
21	8	10+	1	15

表一.

表中第一行為對數, 第二行為某對中使用安慰劑的病患的緩解時段, 第三行為使用6-MP之緩解時段, (皆以週為單位), 第四行為該對勝負狀況, 1表示6-MP得勝, -1表示安慰劑得勝, 第五行為 $W(i)$ 。例如第7列說第7對病人中使用安慰劑的緩解期為2週, 使用6-MP的是16週, 所以6-MP得勝, 而在第7對的勝負分曉時, 6-MP一共淨贏3分。

基於人道考量, 這個試驗是所謂的序列試驗; 即, 要在能看出6-MP與安慰劑之優劣後儘早停止試驗, 以免病患成為劣等治療法的無謂犧牲品。因此, 試驗開始之前便如下規定: 如果一個使用安慰劑的病患比他配對的病患早產生復發現象, 則記安慰劑缺點一次; 反之, 則記6-MP缺點一次。而當雙方缺點次數的差距大到一個程度時, 則即刻停止試驗。

由於停止試驗的時候, 很可能還有病患處於緩解期間。對這些病患, 我們沒法知道他的緩解時段有多長, 而僅知道他在緩解開始至試驗結束止的這一時段內沒有復發。數據表裡第3對病患中服用6-MP的那位便是在32週後, 試驗終止了, 而仍處在緩解期中的一個例子。這種情形的數據是一種所謂的設限數據。如果一個病患因為其他因素在白血病復發前死亡, 或中途離去而不在試驗觀察之中, 則所觀察到的數據也稱為設限數據。

三. 序列分析

我們先引入一些數學符號來描述上述數據。

令 Y_i 是第 i 對病人配成對的時刻 (約為緩解發生的時刻)。令 X_{i1}, X_{i2} 分別代表第 i 對病人中採用6-MP與採用安慰劑的緩解時段。

第二節中所說的停止整個試驗的時間可以更清楚的敘述如下。

令 $W_i(t)$ 為第 i 對病人在時間 (日曆) t 時 6-MP 與安慰劑的勝負狀況。即當 $X_{i1} \wedge t > X_{i2} \wedge t$ 時, 令 $W_i(t) = 1$, 當 $X_{i1} \wedge t < X_{i2} \wedge t$ 時, 令 $W_i(t) = -1$, 當 $X_{i1} \wedge t = X_{i2} \wedge t$ 時, 令 $W_i(t) = 0$ 。(這裡 $a \wedge b$ 是指 a 與 b 二數中較小的那數)。

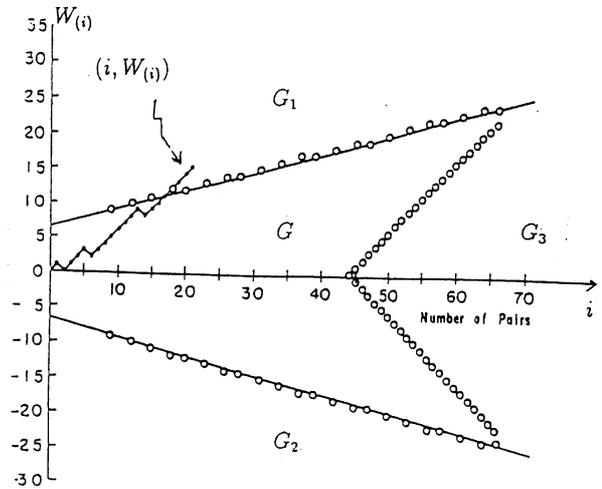
令 $W(t) = \sum_{i=1}^n W_i(t)$ 。則 $W(t)$ 代表了在 t 時刻, 6-MP 比安慰劑多得的點數。這裡 n 為一適當大的數, 大於可能配成的對數。令 $T_i = Y_i + X_{i1} \wedge X_{i2}$ 。令 $T_{(k)}$ 為 $\{T_1, T_2, \dots, T_n\}$ 中第 k 小的數。因此 $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$, 而且 $\{T_1, \dots, T_n\} = \{T_{(1)}, \dots, T_{(n)}\}$ 。

由於 $W(t)$ 只在當 $t = T_{(i)}$ 時才會變動它的值, 令 $W_{(i)} \equiv W(T_{(i)})$ 。 $W_{(1)}, W_{(2)}, \dots, W_{(n)}$ 這個有限數列就代表了 6-MP 與安慰劑的勝負記錄。為了方便, 我們可以把 $(i, W_{(i)}), i = 1, 2, \dots, n$ 劃在座標平面上。

當時, Freireich 等人便是預先把座標平面分成幾個區域, 分別叫 G, G_1, G_2, G_3 (如圖一), 然後規定好, 一旦有一個 i , 使得 $(i, W_{(i)})$ 離開 G , 則停止試驗。用數學符號來講便是定義 停止時刻

$$S = \inf \{i | (i, W_{(i)}) \notin G\}.$$

停止試驗時, 如果 $(i, W_{(i)})$ 是在 G_1 , 則說 6-MP 較好, 若在 G_2 , 則說安慰劑較好, 若在 G_3 , 則說兩者無甚差異。



圖一

這種一邊收集數據, 一邊做統計分析, 然後再決定是否要繼續收集數據的設計稱為序列設計, 相關的分析叫序列分析。它是當今統計領域中理論比較豐富的一支。但是, 在 1960 年代初期它的理論還是相當年輕的。關於這裡的情形, 就有許多問題是當時無法回答的。

Freireich 等人所選的這個 G, G_1, G_2 及 G_3 具有下述性質。令 $P = P(X_{i1} > X_{i2})$ 。如果 $P = 0.75$, 則這試驗會有 0.95 的機率說 6-MP 較好, 如果 $P = 0.25$, 則這試驗會有 0.95 的機率說安慰劑較好。

為什麼選用這樣的 $\{G, G_1, G_2, G_3\}$ 呢? 很自然的, 我們要問, 如果有另外一組 $\{G', G'_1, G'_2, G'_3\}$ 也具有上段所說的性質, 我們該選用那一組? 這時侯有一個簡單的選擇辦法, 即, 看那一組所需的樣本數 S 較少。用心的讀者不難發現這不是一個容易的問題。

四. 存活分析

上述判定6-MP與安慰劑優劣的辦法只是許多辦法之一。它是有些粗糙的，因為它只看每一對病人中的相對勝負，即 $X_{i1} > X_{i2}$ 或 $X_{i1} < X_{i2}$ ，而沒有把其他的量，譬如說， $X_{i1} - X_{i2}$ ，考慮進去。

近代存活分析的主要貢獻之一便是提出了處理這個問題的一個辦法。下邊我們先介紹這個辦法，再提一下它的背景。為此，我們引入隨機過程

$$N_{ik}(t) = 1_{[X_{ik}, \infty)}(t \wedge T).$$

這裡 T 是整個臨床試驗停止的時刻。 $N_{ik}(t)$ 記錄了第 i 對第 k 人在 t 時刻是否仍在緩解期。

假設 X_{ik} 的風險函數 (hazard rate) 是 $\lambda_{ik}(t)$ 。 $\lambda_{ik}(t)$ 的意思是一個緩解期大於或等於 t 的人會馬上復發的機率。假設 $\lambda_{i1}(t) = \lambda_{i2}(t) \cdot e^\theta$ ，則 θ 便是一個比較6-MP與安慰劑優劣的一個指標，在存活分析裡，這叫做相對風險係數 (relative risk coefficient)。有的文獻稱之為療效 (treatment effect)。令

$$G(\theta, t) = \sum_{i=1}^{21} \sum_{k=1}^2 \int_0^t \left(Z_{ik} - \frac{\sum_{k=1}^2 1_{(0, X_{ik}]}^{(s)} e^{\theta Z_{ik}} Z_{ik}}{\sum_{k=1}^2 1_{(0, X_{ik}]}^{(s)} e^{\theta Z_{ik}}} \right) dN_{ik}(s)$$

其中 $Z_{i1} = 1, Z_{i2} = 0$ 。

依據數據，只要知道 θ 的值，則 $G(\theta, t)$ 的值可以明確的算出。我們要用來估計 θ_0 的

估計子 $\hat{\theta}$ 是 $G(\theta, \infty) = 0$ 的根。這個估計子與 Cox (1972) 所提出的 MPLE 是屬於同一類型的。我們趁此把近代存活分析的歷史背景介紹一下。

上述的估計法是當代醫學統計中的一個重要進展。它的理論基礎是隨機積分、點過程之鞅論及連續鞅之中央極限定理。這些東西是60年代至80年代機率理論的主要課題之一。它是40年代 Ito 積分與 Doob 鞅論為起始之近代機率中的一個主要分支，而似乎不是以醫學統計之需要為目標。但是，由於它的適時出現，使得存活分析得到及時的成長。這不能不說是一種巧合。這部分的主要貢獻者包括 Meyer 等人。在點過程方面特別有貢獻的大概是 Jacod。

在以這些機率為基礎來建立存活分析的第一個重要工作是 Aalen 的博士論文。目前的存活分析已經有相當豐富的理論，而不容許我們在這做詳盡介紹。但是，它的基本想法很簡單，以我們這邊的例子來看，它的意思是 Cox 所說的“ $G(\theta_0, t)$ 的期望值是 0”。Cox (1972) 花了相當的篇幅去解釋 $G(\theta_0, t)$ 的期望值是 0。事實上， $G(\theta_0, t)$ 是一個鞅 (martingale)。(據說，當時 Cox 便這樣認為)。後來，在 Aalen 的工作之後，Andersen & Gill (1982) 利用鞅論證明了當時 Cox 所提出之 MPLE 確是一個好的估計子。有興趣的讀者可以參閱 Andersen, Borgan, Gill & Keiding (1993)，我們在這部份也有一些工作，其中之一便是證明上述之 $\hat{\theta}$ 為 θ_0 之最佳估計子。

總之，我們可以用 $\hat{\theta}$ 來估計 θ_0 。我們知道 $\hat{\theta}$ 與 T 有關， T 愈大，則 $\hat{\theta}$ 愈能精確的

估計 θ_0 ；但是， T 愈大，使用劣等治療法的時間愈久。怎樣的 T 最好？這又是序列分析的課題了。

五. 結語

在第三節、第四節中我們藉著 Freireich 數據介紹了一點點的序列分析及存活分析的概念。這兩門學問都有相當的內容，一方面它們成功的扮演了科學之僕的角色，另一方面它們也造就了自己的理論。但是，Freireich 的試驗所能提出的科學結論就只這樣嗎？

Freireich 等人的論文中沒有把 Y_i 報告出來，只報告了 X_{ik} 及 X_{ik}^+ 。許多後來的統計學家便依照這樣的數據提出新的統計方法去分析 6-MP 的療效。除了 Sellke & Siegmund (1983) 之外，沒有人從序列分析的觀點來研究這數據。在 Sellke & Siegmund 的

工作之後，許多人瞭解到這工作的困難情形，也注意到當時 Freireich 沒發表 Y_i 是相當可惜的。事實上，Sellke & Siegmund 不是以研究這套數據為目標，他們是要建立更一般的序列分析理論。他們沒有考慮到配對的設計。如果把配對的設計考慮進來，該如何做呢？事實上，這時候比較好做，最近我們在這問題上得到了一些結論。

參考文獻

文中所引論文多可在以下兩本書中找到

1. Andersen, P. K., Borgan, ϕ , Gill, R. D., Keiding, N. (1993). Statistical Models Based on Counting Processes. Springer-Verlag.
2. Siegmund, D. O. (1985). Sequential Analysis. Springer-Verlag.

—本文作者任教於中央大學數學系與
中央研究院統計科學研究所—