臺灣二ＯＯ八年國際科學展覽會

科　　　　別：數學

作 品 名 稱：別鬧了，辛普森先生

學校 / 作者：國立新竹高級中學　　　　　　李建慶
國立新竹高級中學　　　　　　王翔生

作者簡介

大家好，我們是李建慶和王翔生，新竹高中高二的學生。我們從小熱愛數學和物理，對於未知的事物，都想要努力的去追尋，雖然時常遇到挫折，不過從挫折中我們也學到許多事情，最快樂的事是把想很久的問題終於想出答案或解決的辦法時，那種豁然開朗的感覺。

Abstract

We investigate the machinery producing successive Simpson's paradoxical reverse. Taking advantage of algebraic and geometric techniques, we obtain the following results.

Take playing baseball for example. In our study, we find that Simpson's paradox only occurs when the hitter's hits over 3 times in one game. Set n equal to the times I will hit in one game. If my batting average in each game is at least $\frac{n-1}{2}$ times higher than the others'; then I am sure that my total batting average would not be invert by the others.

In order to find how many the lattice points in the triangle, we use Pick's formula. But sometimes, the Pick's formula is not appropriate to triangles whose vertex are not all lattice points. So we develop New Pick's formula to estimate the number of lattice points in such kind of triangles.

Besides, we also find an iterative algorithm to produce successive "Simpson reverse" phenomenon by using C^{++} language, and we can therefore produce as many "Simpson's set of four sequences" terms as we like(not beyond the computers' upper limit).Moreover, if both sequences of ratios converge, then they must have the same limit.

摘要

我們探討了一般人乍看之下顯得頗弔詭的**辛普森詭論**。我們配合GSP作圖，用解析幾何、設立直角座標系和C^{++}程式的運算，找出在特殊情況下或一般情況下所產生的**辛普森數列組**和特殊的性質，並且以棒球場上的打擊率爲例子來做印證。通常一場棒球賽中，每個人平均上場3次～4次，經過我們的討論，發現要發生逆轉的機會只有在上場達到4次或以上時才會發生。

爲了求出在直角座標系中可以滿足的格子點個數，我們用了Pick公式，但爲了更準確的估計，我們引進了虛點的概念，重新推導出了新Pick公式。

另外，我們還發現，假設兩個人上場比賽，若打了2場，且每場最多上場打擊K次，其中的一個人的打擊率只要是另一個人的$\frac{K-1}{2}$倍以上就保證不會被逆轉。

我們又找到了連續產生辛普森逆轉的演算法，利用C^{++}寫出程式，經由演算法和遞迴式，製造出項數可任意多(只要電腦能夠承受)的辛普森數列組，且我們發現若兩個比值數列接收斂，則極限趨近於同一個數值。

壹、　前言

一、研究動機：

一個特別的機會，看到了辛普森詭論，便對其看似顯然而卻不當然的結果感到好奇，而且一般人是以機率和統計的觀點來看待此詭論，我們卻對如何從幾何和代數來產生詭論中連續不斷的逆轉現象十分感興趣，於是對之展開探討。

二、研究目的：

設$\{p_n\}$ $\{q_n\}$ $\{r_n\}$ $\{s_n\}$皆是正整數所構成的數列，滿足

$$1 \ge \frac{q_n}{p_n} > \frac{s_n}{r_n} \ge 0 \text{ 且 } \frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i} < \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} \text{，} \forall n \ge 2$$

則我們稱$\{p_n, q_n; r_n, s_n\}$爲「辛普森數列組」，

我們希望討論：

（一）兩組組合的辛普森詭論

1. 極端分母解
2. 被極端分母解包含的所有解個數
3. 對於一般情形解的個數估計

（二）辛普森數列組

1. 如何產生辛普森數列組
2. 辛普森數列組的分類
3. 極限值的探討

（三）倍率的探討

貳、　研究過程及方法

一、研究過程

（一）何謂辛普森詭論？

在兩項統計中，甲的機率皆大於乙，但總合起來，即是將甲在兩項統計中所有的值先相加(如第一場的打擊數再加上第二場的打擊數)，再做統計(如算出打擊率)，卻會發現在某些情況下，甲總和的機率將會小於乙總和的機率，產生逆轉(如簡介時所提到的例子)，這就是辛普森詭論。

舉個例子，假設小陳和小王上場比賽打棒球，小陳的打擊率在第一場時是20%，小王卻是0%。第二場時，小陳的打擊率竟高達100%，但小王只有80%，那總合起來，小王的打擊率可否高於小陳？

答案是肯定的，見下圖表：

	小陳	小王
第一場	1/5(20%)	0/1(0%)
第二場	4/4(100%)	4/5(%)
總和	5/9(55.6%)	4/6(66.7%)

這就是辛普森詭論，可以發現最後小陳竟會被小王逆轉。

（二）探討 n=2 的狀況(n 爲數列組的總項數)

以一場棒球比賽爲例：

設甲、乙第 n 場打擊率（$\dfrac{\text{第n場安打數}}{\text{第n場總打數}}$）分別爲 a_n和b_n；**（不可約分）**

且甲、乙的總打擊率（$\dfrac{\text{總安打數}}{\text{總打數}}$）爲 A、B，則求出滿足條件的辛普森數列組（即滿足 $a_n > b_n$ 且 A＜B 時的整數解。）

令第 n 場總打數最高上限爲 K：(分母的最大值)

$\boxed{K=1}$時，　$a_n=\dfrac{1}{1}$；$b_n=\dfrac{0}{1}$，故必不逆轉。

$\boxed{K=2}$時，　$a_n=\dfrac{1}{1}$、$\dfrac{2}{2}$、$\dfrac{1}{2}$；$b_n=\dfrac{1}{2}$、$\dfrac{0}{1}$、$\dfrac{0}{2}$，

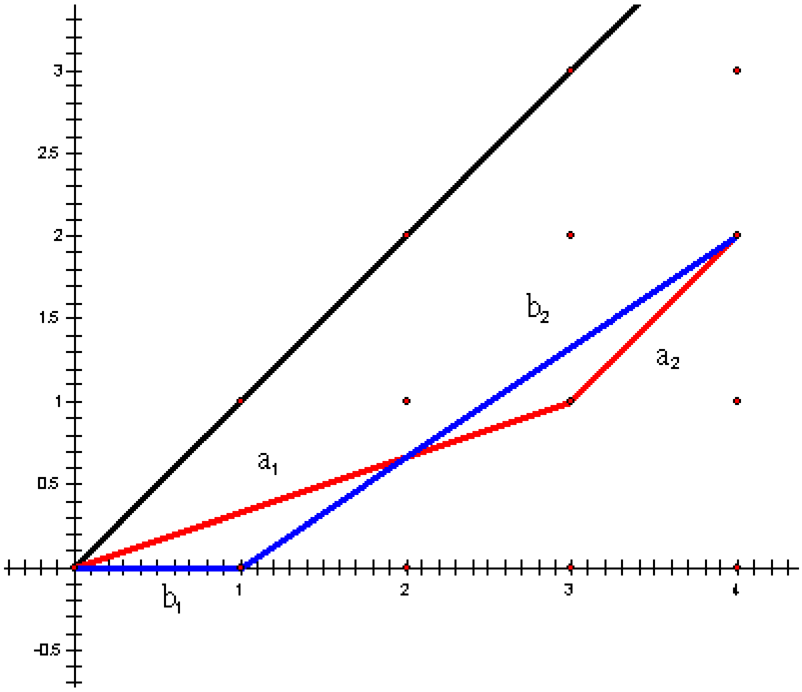
可推得

$$1\ge A\ge\frac{1}{2}\;;\;\frac{1}{2}\ge B\ge 0，$$

因此 $A\ge B$，故此情況也不逆轉。

$\boxed{K=3}$ 時，利用圖形法，我們可以將打擊率視為一個在 XY 平面上的斜率，分子為 Y 軸的截線段長，分母為 X 軸的截線段長，為了方便計算和討論，我們都先令 $a_2 > a_1$

（若 $a_2 = a_1$，則 A= $a_1 = a_2$，但 $\begin{matrix} a_1 = a_2 > b_1 \\ a_1 = a_2 > b_2 \end{matrix} \Rightarrow A > B$，一定無辛普森逆轉），當 $a_2 > a_1$ 時，結果如下頁圖：



可知若 $a_1 > b_1$ 且 $a_2 > b_2$，如果要產生辛普森反轉，就要讓 a_1 的 X 截線段長盡量長，使總斜率不會太大（因 $a_2 > a_1$），而 B 則剛好相反，若要使 B 最大，最小的斜率 b_1 的 X 截線段長一定要最短，而最大的斜率 b_2 的 X 截線段長要最長，才會使總斜率最大，當 K =3 時，最理想的狀況如圖，為了產生能讓 b_2 最大，同時 X 截線段長最小，a_2 需為 $\frac{1}{1}$。如此便會產生：

$$a_1 = \frac{1}{3}, b_1 = \frac{0}{1}$$
$$a_2 = \frac{1}{1}, b_2 = \frac{2}{3}$$
$$A = \frac{1+1}{3+1} = \frac{2}{4}, B = \frac{0+2}{1+3} = \frac{2}{4}$$
$$A = B$$

，所以在 K =3 時，頂多只會使 A=B，而不會產生辛普森逆轉。

$\boxed{K \geq 4}$時，此時情況較多且十分複雜，無較固定作法，所以我們爲了討論方便而令：

甲的第 n 場的打擊數和安打數爲 p_n、q_n，乙的第 n 場的打擊數和安打數爲 r_n、s_n

則 $a_n = \dfrac{q_n}{p_n}, b_n = \dfrac{s_n}{r_n}$

代入 A＝$\dfrac{q_1+q_2}{p_1+p_2}$；B＝$\dfrac{s_1+s_2}{r_1+r_2}$

可得 A＝$a_1 \times \dfrac{p_1}{p_1+p_2} + a_2 \times \dfrac{p_2}{p_1+p_2}$；B＝$b_1 \times \dfrac{r_1}{r_1+r_2} + b_2 \times \dfrac{r_2}{r_1+r_2}$

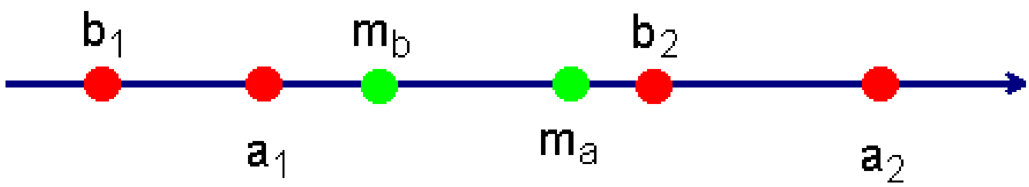
令：

$$\frac{p_1}{p_1+p_2} = [\![p_1]\!] \text{，} \frac{p_2}{p_1+p_2} = [\![p_2]\!]$$

$$\frac{r_1}{r_1+r_2} = [\![r_1]\!] \text{，} \frac{r_2}{r_1+r_2} = [\![r_2]\!]$$

則 A＝$a_1[\![p_1]\!] + a_2[\![p_2]\!]$；B＝$b_1[\![r_1]\!] + b_2[\![r_2]\!]$

如下方數線所示，滿足 $a_1 > b_1$、$a_2 > b_2$、$a_2 > a_1$，令 m_a、m_b 爲 $\overline{a_1a_2}$、$\overline{b_1b_2}$ 的中點，顯然可知 $m_a > m_b$



當$\boxed{p_1 > p_2}$時，根據分點公式可知 A 會在 $\overline{a_1m_a}$ 上，而若$\boxed{p_1 < p_2}$時，A 會在 $\overline{a_2m_a}$ 上，而當$\boxed{p_1 = p_2}$時，A＝m_a，也就是說 A 點會根據 p_1、p_2 的大小而改變位置，同理，B 也一樣。

現在，依 p_1、p_2、r_1、r_2 是否等於 K，我們發現，

在 $a_1 > b_1$、$a_2 > b_2$、$a_2 > a_1$ 時可分爲 16 種情況。

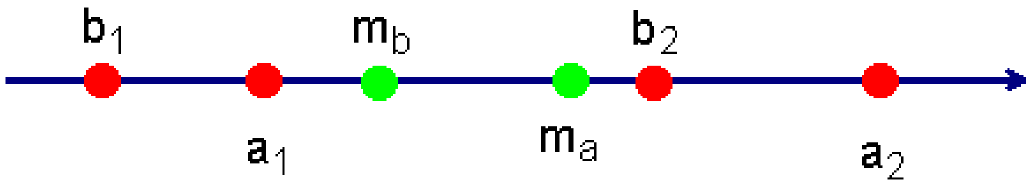
爲了方便底下的討論，我們採用以下的符號：

如下例的左圖，若 a_1 的分母爲 K，則將 a_1 的格子塗黑如下例的右圖。

a_1	b_1
a_2	b_2

	b_1
a_2	b_2

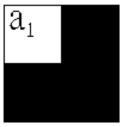
下列為無解情況



討論一：a_1、b_1、a_2、b_2的打擊數（分母）皆為K

$$\begin{matrix} A = m_a \\ B = m_b \end{matrix} \Rightarrow A > B$$，顯然不可能產生辛普森。

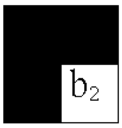
討論二：只有b_1、a_2、b_2的打擊數（分母）為K



因為$p_1 > p_2$，

$$\begin{matrix} A > m_a \\ B = m_b \end{matrix} \Rightarrow A > B$$，不可能產生辛普森。

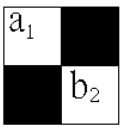
討論三：只有a_1、b_1、a_2的打擊數（分母）為K



因為$r_1 < r_2$，

$$\begin{matrix} A = m_a \\ B < m_b \end{matrix} \Rightarrow A > B$$，不可能產生辛普森。

討論四：只有b_1、a_2的打擊數（分母）為K



因為$p_1 > p_2$、$r_1 < r_2$

$$\begin{matrix} A > m_a \\ B < m_b \end{matrix} \Rightarrow A > B$$，不可能產生辛普森。

以下為有解，為了方便找出解並精確的估計有幾個解，我們需要用到兩件工具，

第一個是在座標上找出極端分母解的方法，第二個是 Pick 公式

（1）Pick 公式

在直角座標系上，給定任意多邊形，其中多邊形的頂點皆為格子點且多邊形的邊並沒有打結那麼，令 Area=多邊形的面積，b=多邊形邊上的格子點，i=多邊形內部的格子點，則

$$Area = \frac{b}{2} + i - 1$$

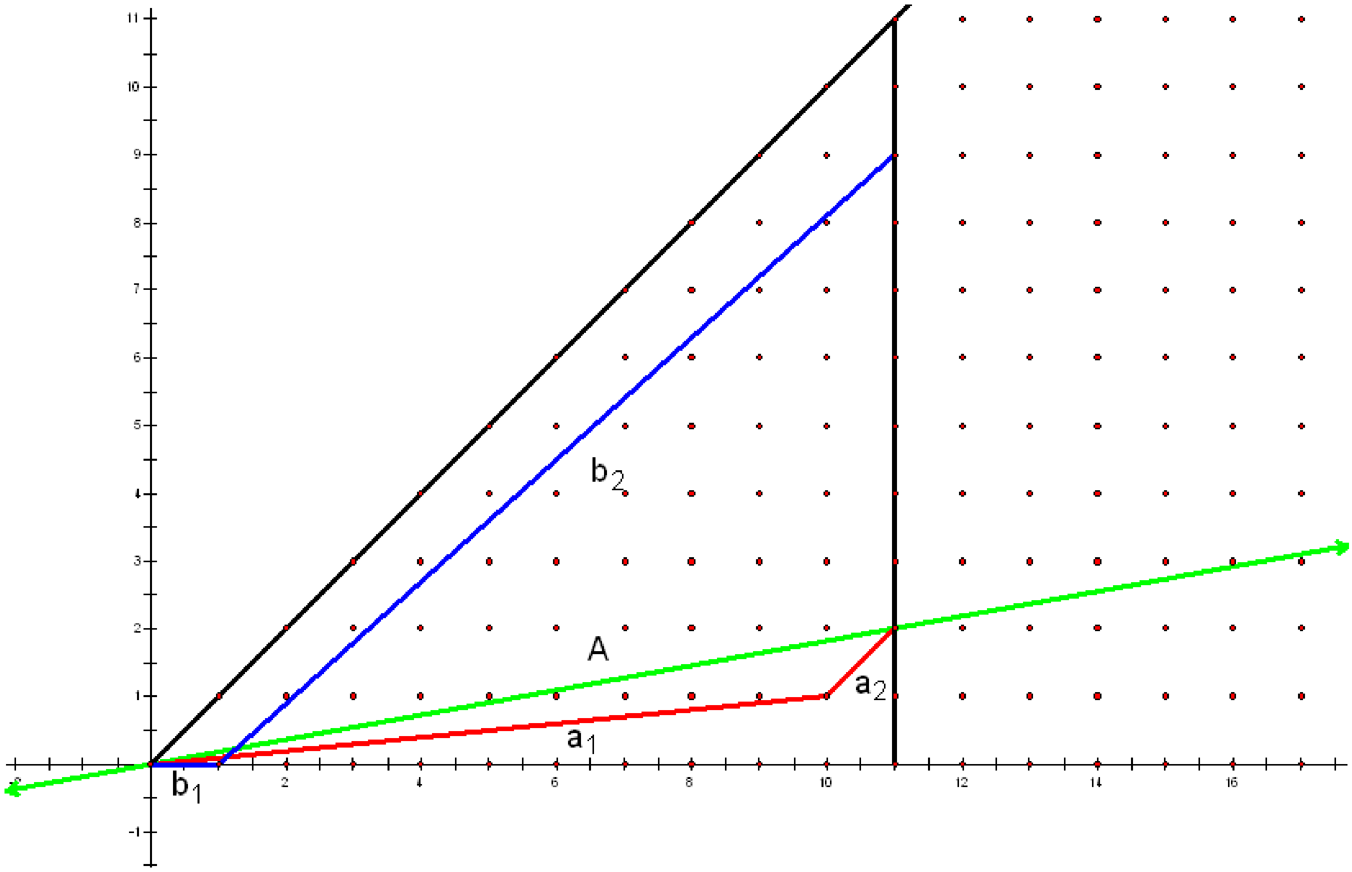
利用 Pick 公式，我們可以估計出由 A、b_2 所夾三角形區域中格子點的個數，就可以求出有幾組解。

如上頁的圖，根據解析幾何可以求出 A 和 b_2 所夾的面積等於 34.11，邊上的點有 8 個，所以可計算出三角形中所包含的點個數有 31.11 個，所以可估計出約有 31 個點。而實際慢慢找，發現格子點有 31 個。

（2）極端分母解

在作圖從幾何圖形中找解時，我們發現 K 增加時，所出現的解的個數也會跟著變多，但是有一種情況很特殊，**這種情況會使(B-A)的值最大**，也就是點(r_1, s_1)和(r_1+r_2, s_1+s_2)所連成的線段和斜率為 A 且過原點的直線以及鉛直線 $x = r_1 + K$ 所夾的圖型中所有的格子點都是會造成辛普森反轉的解。

對於造成此種狀況的(r_1, s_1, r_2, s_2)，我們就稱之為極端分母解。



見上圖，紅色線段爲a_1和a_2，綠色線段爲 A，藍色線段則b_1和b_2

從一開始的假設中，我們知$a_1 > b_1$、$a_2 > b_2$、$a_2 > a_1$，所以爲了讓 A 的值盡量小，也就是在數線上要讓 A 往a_1靠近，根據之前的結論，方法就是讓a_1的 X 截線段長(p_1)要越大越好，而a_2則恰恰相反，因爲$a_2 > a_1$，所以a_2的 X 截線段長(p_2)要越小越好(也就是 1 個格子點的距離)。

反觀b_1，因爲要產生辛普森逆轉，所以b_1和b_2都要很接近a_1和a_2，又因爲$a_2 > a_1$所以$b_2 > b_1$是最好的選擇。將他們畫在數線中，要使 B 盡量往b_2的地方移動，所以b_1的 X 截線段長(r_1)要越小越好(也就是 1 個格子點的距離)，而b_2則恰恰相反，因爲$b_2 > b_1$，所以b_2的 X 截線段長(r_2)要越大越好。

另外，會影響總斜率的不只是長度，本身的斜率大小(a_1、a_2、b_1、b_2)也要考量，從總斜率來看，A 主要是受a_1的影響(因長度較長)，B 主要是受b_2的影響(因長度較長)，所以a_1要盡量小，b_2要盡量大(才會較容易產生辛普森反轉)，所以會迫使a_2的斜率=1，b_1的斜率=0。

證明

我們可以先假設另一條$a_2^{'}$，若$a_2^{'}$的端點在上圖a_2的端點的右邊一格，

綠色線段的斜率會 A 會下降，同時，藍色線段b_2也會跟著下降，但b_2減小所造成三角型中格子點個數減小的量遠較 A 減小所造成三角型中格子點個數所增加的量多，因爲b_2的長度>a_2的長度，所以當轉動固定格子點的角度時，b_2所劃過的面積必大於a_2，所以減少的格子點個數也較多。

同理，若將a_1往上揚，爲了使夾的格子點個數變多，b_1也要跟著往上揚，

但 a_1 的長度> b_1 ，所以因 a_1 上揚所減少的格子點個數必大於 b_1 上揚所增加的格子點個數。所以，極端分母解就是 b_2 末端的點。

綜觀上面的討論，假設每人上場最多為 K 次，我們可以歸納出

$$a_1=\frac{1}{K},b_1=\frac{0}{1},a_2=\frac{1}{1},b_2=\frac{K-1}{K}$$ 為極端分母解

而得到 $$A=\frac{2}{K+1},B=\frac{K-1}{K+1}$$

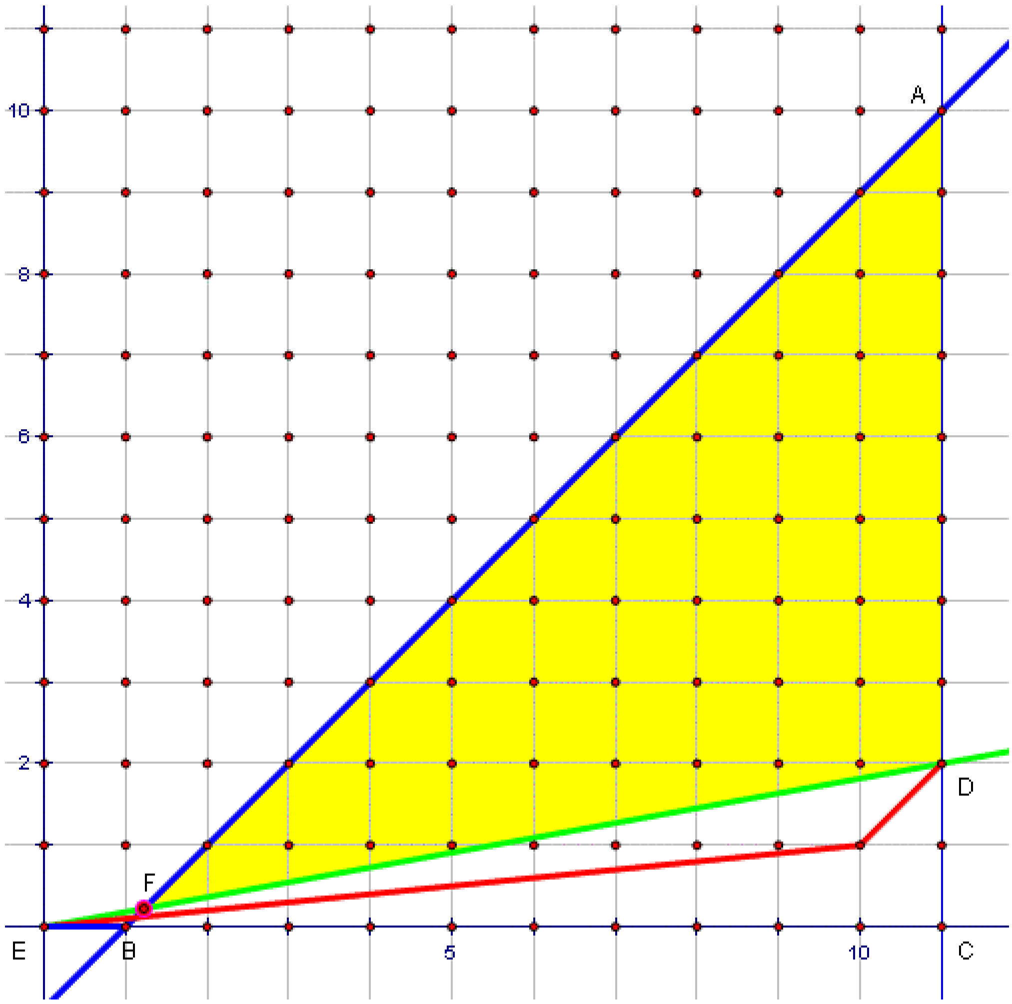
所以當 $K\geq 4$ 時，就有產生辛普森逆轉的可能

我們以上的証明，$\boxed{\text{只適用於 } a_1 \text{、} b_1 \text{、} a_2 \text{、} b_2 \text{ 中任三個在改變}}$，至於四個都在改變時，我們尚未證明出它是最多的解，也就是在

$$a_1\neq\frac{1}{K},b_1\neq\frac{0}{1},a_2\neq\frac{1}{1},b_2\neq\frac{K-1}{K}$$

的狀況下，使否會產生一種 a_1 、 a_2 、 b_1 、 b_2 ，他所圍出的區域中，格子點的個數比極端分母解多。不過我們用 C^{++} 程式跑 K=4~10 的情況中，發現它產生解的個數都小於極端分母解所產生的解的個數，而且格子點個數的差值越來越大。所以我們仍大膽推測極端分母解所包含的區域是造成辛普森反轉中，格子點最多的。但證明還在努力中。

極端分母解數公式



將極端分母解 $\overline{A_1}=(K,1)$ 、 $\overline{A_2}=(1,1)$ 、 $\overline{B_1}=(1,0)$ 、 $\overline{B_2}=(K,K)$ 代入，如上圖所示，爲一極端分母解的範例，而極端分母解的解爲圖中黃色 ΔADF 的內格子點，也就是將 ΔABC 的內格子點扣掉 ΔCDE 的內格子點再加上右邊 $\overline{AD}$ 上的格子點

（不包括 A 點 D 點），最後若 $K+1$ 是偶數則須扣掉 $\overline{DE}$ 中點。

則 A $(K+1,K)$ 、B $(1,0)$ 、C $(K+1,0)$ 、D $(K+1,2)$ 、E $(0,0)$ ，可由直角三角形內格子點公式推得 ΔABC 的內格子點

$$=\frac{K^2-(K,K)-(K+K)+2}{2}=\frac{K^2-3K+2}{2}=\frac{(K-2)(K-1)}{2},$$

ΔCDE 的內格子點

$$=\frac{2(K+1)-(K+1,2)-(K+1+2)+2}{2}=\frac{K+1-(K+1,2)}{2}$$

$\overline{AD}$ 上的格子點（不包括 A 點 D 點）$=K-3$

$\overline{DE}$ 中點上的格子點數$=(K+1,2)-1$

則極端分母解 ΔADF 的內格子點

$$=\frac{(K-2)(K-1)}{2}-\frac{K+1-(K+1,2)}{2}+K-3-(K+1,2)+1=\frac{K^2-2K-3-(K+1,2)}{2}$$

此即爲極端分母解數公式。

（3）$\boxed{\text{格子點的探討}}$

由於極端分母解的解爲直角座標系上三條相異直線圍成三角形內的格子點，爲了求出極端分母解的ㄧ般式，需要先解決此一問題，然後不斷嘗試方法解決，卻沒有更完整的突破，在此無法精準無誤求出格子點數的困境之下，只好另尋其他較精準、誤差較小的方法。

格子點問題雖然沒有完全解決，但在思考何討論的過程中也有了一些想法和結論，以下是介紹格子點的相關性質。

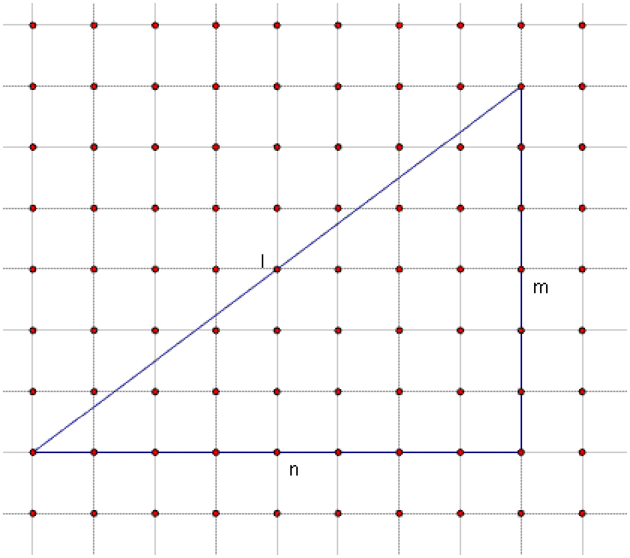
格子點性質 1：直角座標系上過一格子點的直線斜率若是有理數則必過無限多個格子點，若是無理數則只過原本通過的格子點。

證明：若直線斜率爲一有理數，則其必可表示斜率 $m=\dfrac{b}{a}$ $\quad a,b\in Z$

故若已知經過一點 (x_0,y_0) ，則此條直線上另一點 (x_0+a,y_0+b) 亦爲格子點。
同理可證，若一直線過兩點以上，則其斜率必爲有理數。

格子點性質 2：兩條過格子點水平距離等於 1 斜率爲 $\dfrac{1}{n}$ $\quad n\in Z$ 的平行線間無格子點。

格子點直角三角形的內格子點數

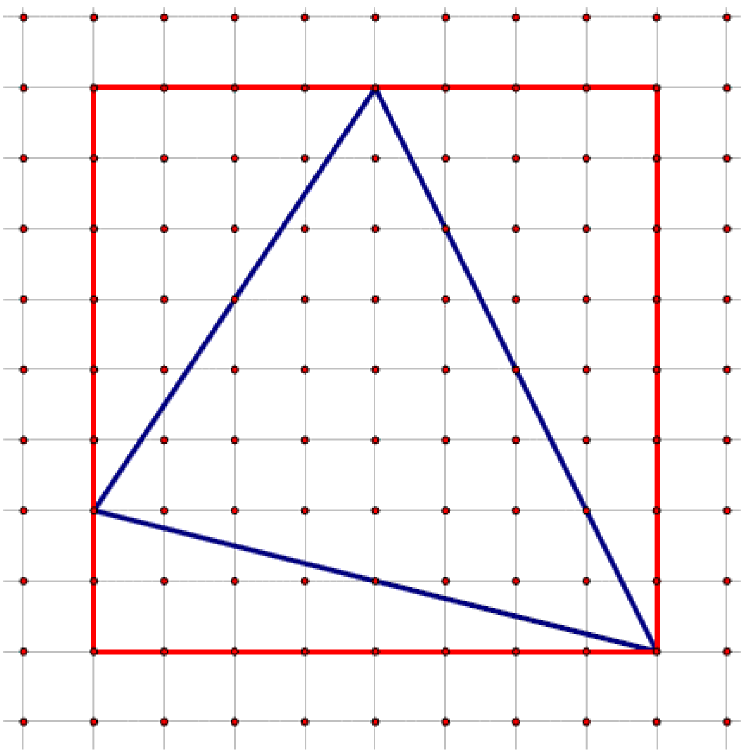


如圖，若三格子點圍成直角三角形，且兩股長分別為 n、m，則其斜邊上的格子點數$=(n,m)+1$，兩股上的格子點數$=n+m+1$，而三邊上的總格子點數$=(n,m)+n+m-1$

而其內部的格子點數等於邊長為 n、m 的長方形扣掉四邊及對角線上的格子點數的一半$=\dfrac{(n+1)(m+1)-(n,m)-2(n+m)+1}{2}=\dfrac{nm-(n,m)-(n+m)+2}{2}$

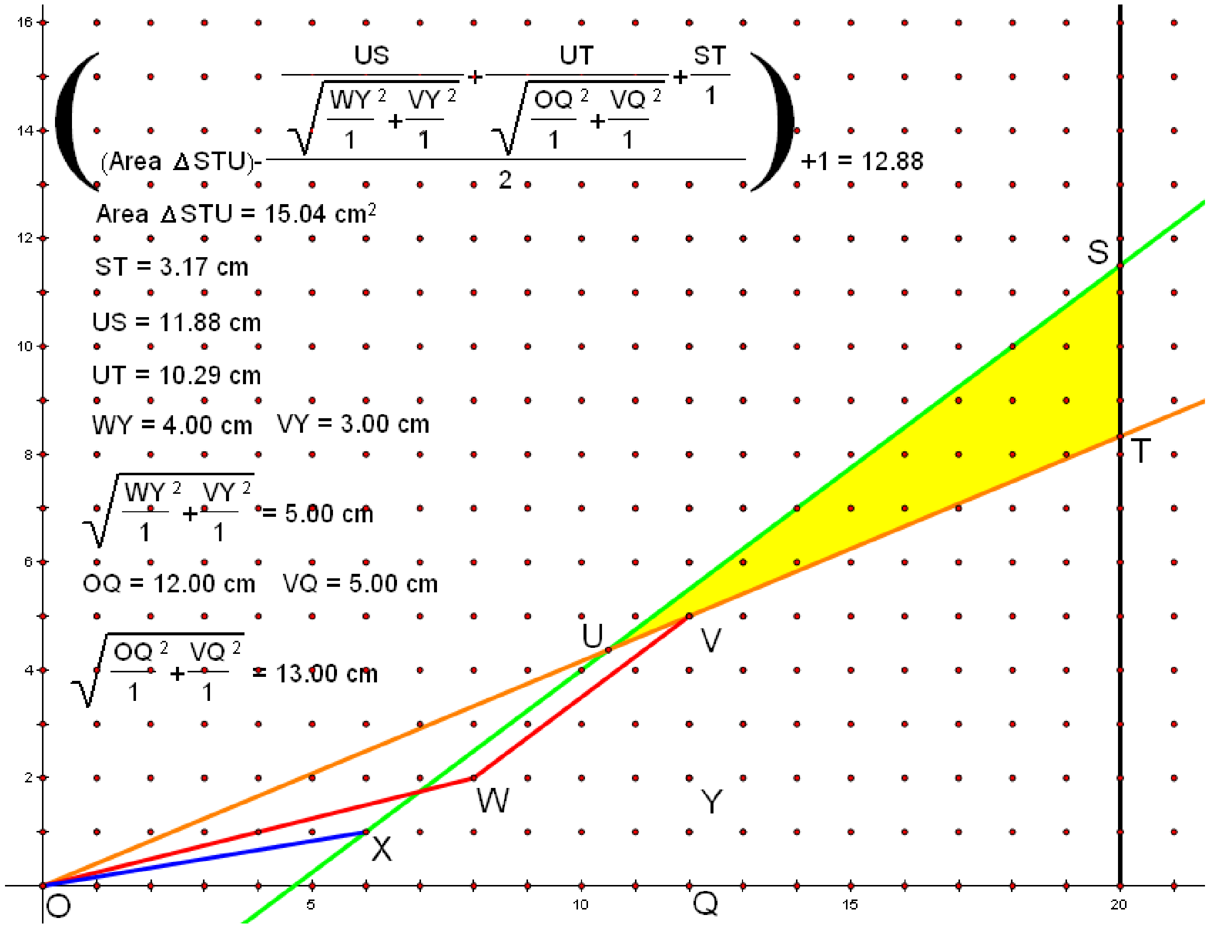
(稱之為$\boxed{\text{直角三角形內格子點公式}}$)

格子點三角形內格子點數



而若不是直角三角形則如左圖，過三頂點作水平線或鉛直線交於四點形成長方形，則將長方形的格子點扣掉三個直角三角形的格子點便是此三角形的內格子點。

（4）$\boxed{\text{新 Pick}}$



對於其他的三角形(非極端分母解)，最末端的點，如上圖的 S 和 T，大都非格子點，爲了能更方便利用 Pick 公式，我們做了一個假設「虛點」。

虛點的意思是不完全的點，就像上圖的 T，它並非在格子點上，但我們可以從線段 $\overline{ST}$ 中發現，格子點和格子點的間距皆爲 1cm，經由計算可知 T 的座標爲 $(20,8\frac{1}{3})$，而在 $\overline{ST}$ 上，距離 T 最近的格子點和 T 的距離是 $1-\frac{1}{3}=\frac{2}{3}$，而在 $\overline{ST}$ 上每隔 1cm 就有一個點，所以我們便假設這個虛點 T 是 $\dfrac{(\frac{2}{3})}{1}=\dfrac{2}{3}$ 個點。

所以我們定義虛點爲 $\dfrac{l}{d}$，l 爲距離頂點處最近的格子點距離，d 是在同一線上相鄰兩格子點間的距離。

而在一條線段 L 上，若不計算起始點，包含在 L 上的格子點爲

$$\frac{\text{線段總長}}{\text{線段相鄰兩格子點間的距離}}=\frac{\text{線段總長}}{\text{d}}$$

我們設上圖之三角形的三邊長分別爲 L_1、L_2、L_3，格子點間的距離爲 d_1、

d_2 、 d_3 ，所以可得三角形邊上的點(包含虛點)總共有 $\frac{L_1}{d_1}+\frac{L_2}{d_2}+\frac{L_3}{d_3}$ 個。

以上圖爲例，我們令

$$\begin{cases} L_1 = \overline{SU} \\ L_2 = \overline{UT} \\ L_3 = \overline{ST} \end{cases} \qquad \begin{cases} d_1 = \sqrt{\left(\frac{\overline{VY}}{(\overline{VY},\overline{WY})}\right)^2 + \left(\frac{\overline{WY}}{(\overline{VY},\overline{WY})}\right)^2} = \sqrt{\left(\frac{3}{1}\right)^2 + \left(\frac{4}{1}\right)^2} = 5 \\ d_2 = \sqrt{\left(\frac{\overline{VQ}}{(\overline{VQ},\overline{OQ})}\right)^2 + \left(\frac{\overline{OQ}}{(\overline{VQ},\overline{OQ})}\right)^2} = \sqrt{\left(\frac{5}{1}\right)^2 + \left(\frac{12}{1}\right)^2} = 13 \\ d_3 = 1 \end{cases}$$

因此可估計出三角形內的個子點有 12.88 個，若用原本的 Pick 會得到 13.04 個，然而實際上只有 12 個，所以用此方法來做較精確的估計。

如上頁圖，固定其中三個值，求第四個解，令分母最大爲 K ，以向量表示

$\overline{a_n} = (A_{nx}, A_{ny})$ A 第 n 場時(總打數,安打數)，

$\overline{b_n} = (B_{nx}, B_{ny})$ B 第 n 場時(總打數,安打數)。

則可設 A、B 兩場的打擊爲

$\overline{a_1} = (A_{1x}, A_{1y})$ 、 $\overline{b_2} = (A_{2x}, A_{2y})$ 、 $\overline{b_1} = (B_{1x}, B_{1y})$ 以及爲了求最多解，作 $\overleftrightarrow{XS}$ 平行 $\overleftrightarrow{WV}$ ，

則 $\overleftrightarrow{OV}$ 交 $\overleftrightarrow{XS}$ 於 U 點，交過 S 的鉛直線於 T 點，則 ΔUST 內的格子點即爲 b_2 滿足

此種 $(\overline{a_1}, \overline{a_2}, \overline{b_1})$ 情況的解，因此假設 $\overline{b_2} = (K, K\frac{A_{2y}}{A_{2x}})$ 。

由以上兩場的結果便可假設座標 $W\ (A_{1x}, A_{1y})$ 、 $V\ (A_{1x}+A_{2x}, A_{1y}+A_{2y})$ 、

$X\ (B_{1x}, B_{1y})$ 、 $S\ (B_{1x}+K, B_{1y}+K\frac{A_{2y}}{A_{2x}})$ 。

可推得三條直線的方程式

$\overleftrightarrow{UT}$ ： $(A_{1y}+A_{2y})x-(A_{1x}+A_{2x})y=0$ ………(1)

$\overleftrightarrow{US}$ ： $(y-B_{1y})=(x-B_{1x})\frac{A_{2y}}{A_{2x}}$ ………(2)

$\overleftrightarrow{ST}$ ： $x=B_{1x}+K$ ………(3)

兩兩解聯立方程式，將(3)代入(2)得 S 點 $x=B_{1x}+K$ ， $y=B_{1y}+K\frac{A_{2y}}{A_{2x}}$

將(3)代入(1)得 T 點 $x=B_{1x}+K$，$y=(B_{1x}+K)\dfrac{A_{1y}+A_{2y}}{A_{1x}+A_{2x}}$，再利用克拉瑪公式解 U 點座標。

$$\begin{cases}(A_{1y}+A_{2y})x-(A_{1x}+A_{2x})y=0\\ A_{2y}x-A_{2x}y=A_{2y}B_{1x}-A_{2x}B_{1y}\end{cases}$$

求得

$$\triangle=\begin{vmatrix}A_{2y} & -A_{2x}\\ A_{1y}+A_{2y} & -(A_{1x}+A_{2x})\end{vmatrix}=A_{2x}A_{1y}-A_{1x}A_{2y}$$

$$\triangle y=\begin{vmatrix}A_{2y} & -A_{2x}B_{1y}+A_{2y}B_{1x}\\ A_{1y}+A_{2y} & 0\end{vmatrix}=(A_{2x}B_{1y}-A_{2y}B_{1x})(A_{1y}+A_{2y})$$

$$\triangle x=\begin{vmatrix}A_{2x}B_{1y}-A_{2y}B_{1x} & -A_{2x}\\ 0 & -(A_{1x}+A_{2x})\end{vmatrix}=A_{1x}A_{2y}B_{1x}+A_{2x}A_{2y}B_{1x}-A_{1x}A_{2x}B_{1y}-A_{2x}{}^{2}B_{1y}$$

$$=(A_{2x}B_{1y}-A_{2y}B_{1x})(A_{1x}+A_{2x})$$

得 U 點 $x=\dfrac{\triangle x}{\triangle}=\dfrac{(A_{2x}B_{1y}-A_{2y}B_{1x})(A_{1x}+A_{2x})}{A_{2x}A_{1y}-A_{1x}A_{2y}}$

$$y=\frac{\triangle y}{\triangle}=\frac{(A_{2x}B_{1y}-A_{2y}B_{1x})(A_{1y}+A_{2y})}{A_{2x}A_{1y}-A_{1x}A_{2y}}$$ 。

爲方便以下的公式推導，先以 $S(S_x,S_y)$、$T(T_x,T_y)$、$U(U_x,U_y)$ 三點導出，而此三點座標皆已推出。

可得 $\triangle UST$ 的面積$=\dfrac{(S_y-T_y)(T_x-U_x)}{2}$，

而 $\triangle UST$ 邊上的點個數等於 $\dfrac{\text{線段總長}}{\text{線段相鄰兩格子點間的距離}}$

$$=\frac{\overline{ST}}{1}+\frac{\overline{US}}{(A_{2x},A_{2y})}+\frac{\overline{UT}}{(A_{1x}+A_{2x},A_{1y}+A_{2y})}$$

其中(A_{2x},A_{2y}) 、$(A_{1x}+A_{2x},A_{1y}+A_{2y})$表示其最大公因數。

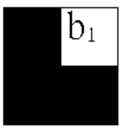
$$=(S_y-T_y)+\sqrt{\frac{(S_x-U_x)^2+(S_y-U_y)^2}{(\frac{A_{2x}}{(A_{2x},A_{2y})})^2+(\frac{A_{2y}}{(A_{2x},A_{2y})})^2}}+\sqrt{\frac{(T_x-U_x)^2+(T_y-U_y)^2}{(\frac{A_{1x}+A_{2x}}{(A_{1x}+A_{2x},A_{1y}+A_{2y})})^2+(\frac{A_{1y}+A_{2y}}{(A_{1x}+A_{2x},A_{1y}+A_{2y})})^2}}$$

再由 PICK 定理 $Area=\frac{b}{2}+i-1$ 得$\triangle UST$內的格子點數

$$i=Area-\frac{b}{2}+1$$

$$=\frac{(S_y-T_y)(T_x-U_x-1)+\sqrt{\frac{(S_x-U_x)^2+(S_y-U_y)^2}{(\frac{A_{2x}}{(A_{2x},A_{2y})})^2+(\frac{A_{2y}}{(A_{2x},A_{2y})})^2}}+\sqrt{\frac{(T_x-U_x)^2+(T_y-U_y)^2}{(\frac{A_{1x}+A_{2x}}{(A_{1x}+A_{2x},A_{1y}+A_{2y})})^2+(\frac{A_{1y}+A_{2y}}{(A_{1x}+A_{2x},A_{1y}+A_{2y})})^2}}}{2}+1$$

$\boxed{\text{討論五}}$：只有a_1 、a_2 、b_2的打擊數（分母）爲K



根據$\boxed{\text{極端分母解}}$，會產生

$$a_1=\frac{1}{K},b_1=\frac{0}{1},a_2=\frac{K}{K},b_2=\frac{K-1}{K}$$

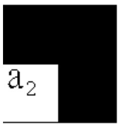
推得$A=\frac{1+K}{2K},B=\frac{K-1}{K+1}$，若逆轉，表示$A=\frac{1+K}{2K}<B=\frac{K-1}{K+1}$

得到

$$K^2+2K+1<2K^2-2K$$
$$K^2-4K-1>0$$
$$(K-2)^2>5$$

所以當這種情況有解時，分母至少要等於 5 才有可能。

討論六：只有 a_1 、 b_1 、 b_2 的打擊數（分母）為 K



根據極端分母解，會產生

$$a_1=\frac{1}{K},b_1=\frac{0}{K},a_2=\frac{1}{1},b_2=\frac{K-1}{K}$$

推得 $A=\dfrac{2}{K+1},B=\dfrac{K-1}{2K}$ ，若逆轉，表示 $A=\dfrac{2}{K+1}<B=\dfrac{K-1}{2K}$

得到

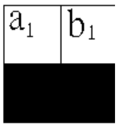
$$4K<K^2-1$$

$$K^2-4K-1>0$$

$$(K-2)^2>5$$

所以當這種情況有解時，分母至少要等於 5 才有可能。

討論七：只有 a_2 、 b_2 的打擊數（分母）為 K



根據極端分母解，會產生

$$a_1=\frac{1}{K-1},b_1=\frac{0}{1},a_2=\frac{K}{K},b_2=\frac{K-1}{K}$$

推得 $A=\dfrac{K+1}{2K-1},B=\dfrac{K-1}{K+1}$ ，若逆轉，表示 $A=\dfrac{K+1}{2K-1}<B=\dfrac{K-1}{K+1}$

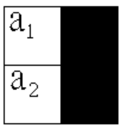
得到

$$K^2+2K+1<2K^2-3K+1$$

$$K^2-5K>0$$

所以當這種情況有解時，分母至少要等於 6 才有可能。

討論八：只有 b_1 、 b_2 的打擊數（分母）為 K



根據極端分母解，會產生

$$a_1=\frac{1}{K-1},b_1=\frac{1}{K},a_2=\frac{1}{1},b_2=\frac{K-1}{K}$$

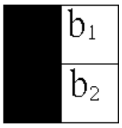
推得 $A=\dfrac{2}{K},B=\dfrac{K}{2K}$ ，若逆轉，表示 $A=\dfrac{2}{K}<B=\dfrac{K}{2K}$

得到

$$4<K$$

所以當這種情況有解時，分母至少要等於 5 才有可能。

討論九：只有 a_1 、 a_2 的打擊數（分母）爲 K



根據極端分母解，會產生

$$a_1=\frac{1}{K}, b_1=\frac{0}{1}, a_2=\frac{K}{K}, b_2=\frac{K-2}{K-1}$$

推得 $A=\dfrac{K+1}{2K}, B=\dfrac{K-2}{K}$ ，若逆轉，表示 $A=\dfrac{K+1}{2K}<B=\dfrac{K-2}{K}$

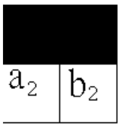
得到

$$K^2+K<2K^2-4K$$

$$K^2-5K>0$$

所以當這種情況有解時，分母至少要等於 6 才有可能。

討論十：只有 a_1 、 b_1 的打擊數（分母）爲 K



根據極端分母解，會產生

$$a_1=\frac{1}{K}, b_1=\frac{0}{K}, a_2=\frac{1}{1}, b_2=\frac{K-2}{K-1}$$

推得 $A=\dfrac{2}{K+1}, B=\dfrac{K-2}{2K-1}$ ，若逆轉，表示 $A=\dfrac{2}{K+1}<B=\dfrac{K-2}{2K-1}$

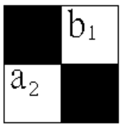
得到

$$4K-2<K^2-K-2$$

$$K^2-5K>0$$

所以當這種情況有解時，分母至少要等於 6 才有可能。

討論十一：只有 a_1 、 b_2 的打擊數（分母）爲 K



根據極端分母解，會產生

$$a_1=\frac{1}{K}, b_1=\frac{0}{1}, a_2=\frac{1}{1}, b_2=\frac{K-1}{K}$$

推得 $A=\dfrac{2}{K+1}, B=\dfrac{K-1}{K+1}$ ，若逆轉，表示 $A=\dfrac{2}{K+1}<B=\dfrac{K-1}{K+1}$

得到

$$2<K-1$$

$$K>3$$

所以當這種情況有解時，分母至少要等於 4 才有可能。

討論十二：只有b_2的打擊數（分母）為K

a_1	b_1
a_2	

根據極端分母解，會產生

$$a_1=\frac{1}{K-1},b_1=\frac{0}{1},a_2=\frac{1}{1},b_2=\frac{K-1}{K}$$

推得$A=\frac{2}{K},B=\frac{K-1}{K+1}$，若逆轉，表示$A=\frac{2}{K}<B=\frac{K-1}{K+1}$

得到

$$2K+2<K^2-K$$

$$K^2-3K-2>0$$

所以當這種情況有解時，分母至少要等於 4 才有可能。

討論十三：只有a_2的打擊數（分母）為K

a_1	b_1
	b_2

根據極端分母解，會產生

$$a_1=\frac{1}{K-1},b_1=\frac{0}{1},a_2=\frac{K}{K},b_2=\frac{K-2}{K-1}$$

推得$A=\frac{1+K}{2K-1},B=\frac{K-2}{K}$，若逆轉，表示$A=\frac{1+K}{2K-1}<B=\frac{K-2}{K}$

得到

$$K+K^2<2K^2-5K+2$$

$$K^2-6K+2>0$$

$$(K-3)^2-7>0$$

所以當這種情況有解時，分母至少要等於 6 才有可能。

討論十四：只有b_1的打擊數（分母）為K

a_1	
a_2	b_2

根據極端分母解，會產生

$$a_1=\frac{1}{K-1},b_1=\frac{1}{K},a_2=\frac{1}{1},b_2=\frac{K-2}{K-1}$$

推得$A=\frac{2}{K},B=\frac{K-1}{2K-1}$，若逆轉，表示$A=\frac{2}{K}<B=\frac{K-1}{2K-1}$

得到

$$4K-2<K^2-K$$

$$K^2-5K+2>0$$

所以當這種情況有解時，分母至少要等於 5 才有可能。

討論十五：只有a_1的打擊數（分母）為K

	b1
a2	b2

根據極端分母解，會產生

$$a_1=\frac{1}{K},b_1=\frac{0}{1},a_2=\frac{1}{1},b_2=\frac{K-2}{K-1}$$

推得$A=\dfrac{2}{K+1},B=\dfrac{K-2}{K}$，若逆轉，表示$A=\dfrac{2}{K+1}<B=\dfrac{K-2}{K}$

得到

$$2K<K^2-K-2$$

$$K^2-3K-2>0$$

所以當這種情況有解時，分母至少要等於 4 才有可能。

討論十六： a_1、b_1、a_2、b_2的打擊數（分母）皆不為K

a1	b1
a2	b2

此種狀況所包含的辛普森個數就是在打擊數為（K-1）時的總辛普森個數。

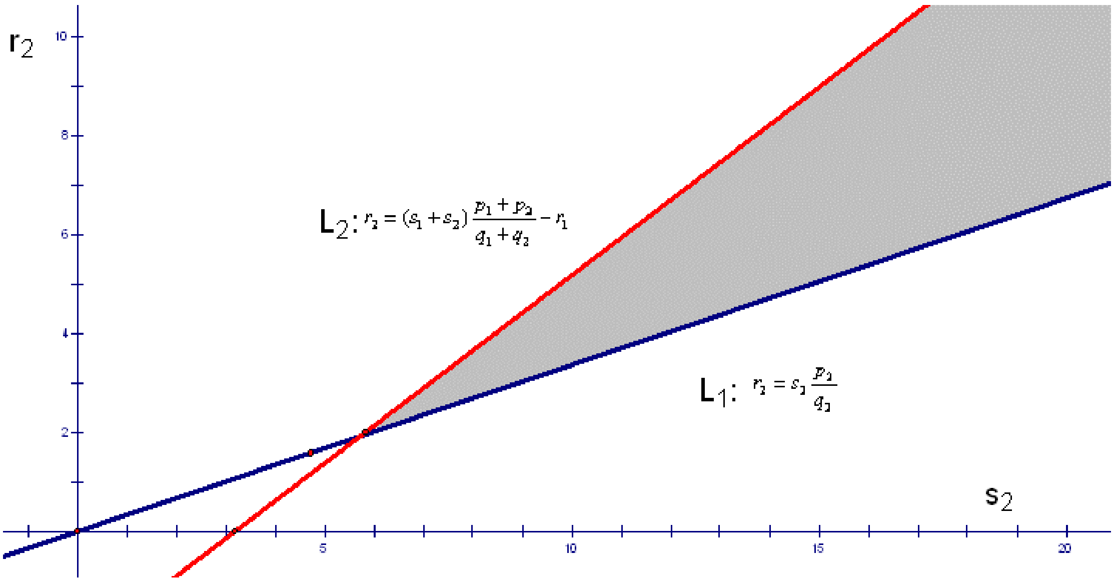
（三）如何產生辛普森數列組？

首先考慮兩組的情形，當已知 $\frac{q_1}{p_1} > \frac{s_1}{r_1}$ 時，找出一組 $\{p_2, q_2; r_2, s_2\}$ 滿足 $\frac{q_2}{p_2} > \frac{s_2}{r_2}$ 和 $\frac{q_1+q_2}{p_1+p_2} < \frac{s_1+s_2}{r_1+r_2}$ 的辛普森數列組 $\{p_n, q_n; r_n, s_n\}$。

則 $r_2 > s_2 \frac{p_2}{q_2}$ ， $r_2 < (s_1+s_2)\frac{p_1+p_2}{q_1+q_2} - r_1$

令 s_2 為橫軸、 r_2 為縱軸， L_1 為 $r_2 = s_2 \frac{p_2}{q_2}$ 、 L_2 為 $r_2 = (s_1+s_2)\frac{p_1+p_2}{q_1+q_2} - r_1$

設立直角座標系如下圖示



則 r_2 範圍為 L_1、L_2 和橫軸所圍成的區域，L_2 與橫軸的交點為 $(r_1\frac{q_1+q_2}{p_1+p_2} - s_1, 0)$

$\boxed{L_2\text{交橫軸於正向}}$：

若要使 L_2 交橫軸於正向，則

需要讓 $r_1\frac{q_1+q_2}{p_1+p_2} - s_1 > 0$

相當於 $\frac{q_1+q_2}{p_1+p_2} > \frac{s_1}{r_1}$

而當 $\frac{q_2}{p_2} > \frac{q_1}{p_1}$ 時，

可推得：

$$\frac{q_1+q_2}{p_1+p_2} > \frac{q_1}{p_1} \ ; \ \frac{q_1}{p_1} > \frac{s_1}{r_1}$$

故當 $\frac{q_2}{p_2} > \frac{q_1}{p_1}$ 時，L_2 交橫軸於正向。

若交於正向，要讓 r_2 有解， L_2 的斜率需要大於 L_1 的斜率，即需要滿足：

$$\frac{p_1+p_2}{q_1+q_2} > \frac{p_2}{q_2}$$

則在 $\frac{q_2}{p_2} > \frac{q_1}{p_1}$ 時，上式恆成立，必有解。

$\boxed{L_2\text{交橫軸於負向}}$：

若要使 L_2 交橫軸於負向，則

必須使 $r_1 \frac{q_1+q_2}{p_1+p_2} - s_1 < 0$

也就是 $\frac{q_1+q_2}{p_1+p_2} < \frac{s_1}{r_1}$

而又 $\frac{q_1}{p_1} > \frac{s_1}{r_1}$ ，所以 $\frac{q_1+q_2}{p_1+p_2} < \frac{s_1}{r_1} < \frac{q_1}{p_1}$ ，發生在 $\frac{q_2}{p_2} < \frac{q_1}{p_1}$ 的時候。

若 $\frac{q_2}{p_2} < \frac{q_1}{p_1}$ 時，必須滿足

$$\frac{q_1+q_2}{p_1+p_2} < \frac{s_1}{r_1} \qquad \text{才有解。}$$

依照辛普森數列組的關係式：

$$\begin{cases} \frac{q_2}{p_2} > \frac{s_2}{r_2} \\ \frac{q_1+q_2}{p_1+p_2} < \frac{s_1+s_2}{r_1+r_2} \end{cases},$$

可推得 $$\begin{cases} r_2 > s_2 \frac{p_2}{q_2} \\ r_2 < (s_1+s_2)\frac{p_1+p_2}{q_1+q_2} - r_1 \end{cases},$$

進一步得到 $s_2 \frac{p_2}{q_2} < r_2 < (s_1+s_2)\frac{p_1+p_2}{q_1+q_2} - r_1$ ，

若要讓 r_2 一定有整數解存在，則必須使 $(s_1+s_2)\frac{p_1+p_2}{q_1+q_2} - r_1 - s_2 \frac{p_2}{q_2} > 1$ ，

則在：

$$s_2 > \frac{1 + r_1 - s_1 \frac{p_1+p_2}{q_1+q_2}}{\frac{p_1+p_2}{q_1+q_2} - \frac{p_2}{q_2}} \qquad \text{時 } r_2 \text{ 有整數解。}$$

故給定第一組數之後，就可以依照關係式找出第二組數形成辛普森數列組。

以下探討多組解的狀況，先作以下假設：

$$P_n=\sum_{i=1}^{n}p_n \text{ 、 } Q_n=\sum_{i=1}^{n}q_n \text{ 、 } R_n=\sum_{i=1}^{n}r_n \text{ 、 } S_n=\sum_{i=1}^{n}s_n \text{ 。}$$

則須滿足

$$\begin{cases}\dfrac{q_n}{p_n}>\dfrac{s_n}{r_n}\\ \dfrac{Q_n}{P_n}<\dfrac{S_n}{R_n}\end{cases}\text{，n>1}$$

則 $\begin{cases}r_n>s_n\dfrac{q_n}{p_n}\\ r_n<S_n\dfrac{P_n}{Q_n}-R_{n-1}\end{cases}$，

可得 $s_n\dfrac{q_n}{p_n}<r_n<S_n\dfrac{P_n}{Q_n}-R_{n-1}$，

如果要讓整數解一定存在，則

$$S_n\frac{P_n}{Q_n}-R_{n-1}-s_n\frac{p_n}{q_n}>1\text{ ，}$$

也就是

$$s_n>\frac{1+R_{n-1}-S_{n-1}\dfrac{P_n}{Q_n}}{\dfrac{P_n}{Q_n}-\dfrac{p_n}{q_n}}\text{ 。}$$

以上遞迴式稱之爲「辛普森遞迴式」，用此遞迴式可製造出連續不斷的辛普森反轉，附錄有我們用 C^{++} 所找出的辛普森數列組。

（四）安全倍率

套用$\boxed{\text{極端分母解}}$，會得到$a_1=\frac{1}{K}, b_1=\frac{0}{1}, a_2=\frac{1}{1}, b_2=\frac{K-1}{K}$

假設a_n的打擊率皆大於b_n的 t 倍，則式子可改成

$$a_1=\frac{1}{K}>t\frac{0}{1}=tb_1$$

$$a_2=\frac{1}{1}>t\frac{[\frac{K-1}{t}]}{K}=tb_2$$

所以得到
$$A=\frac{2}{K+1}, B=\frac{[\frac{K-1}{t}]}{K+1}$$

若不逆轉，表示 A>B

所以$2>[\frac{K-1}{t}]$，而得到$0<\frac{K-1}{t}<2$

所以$t>\frac{K-1}{2}$

所以在兩場比賽中，只要打擊率比對手多$\frac{K-1}{2}$倍，就保證不被逆轉

（五）倍率變化隨場數變化

(1)當一個越好一個越差

$\boxed{a_2\text{變大，}b_2\text{變小}}$

從數線上看，本來a_1就在b_1的右邊，若a_2變高，則a_2會在a_1的右邊，使 A 也在a_1的右邊。

反觀b_2，b_2會在b_1左邊，使 B 也在b_1左邊。

一開始a_1就在b_1的右邊，而 A 在a_1的右邊，B 在b_1左邊，所以 A 一定在 B 的右邊，所以不可能反轉。

$\boxed{a_2\text{變小，}b_2\text{變大}}$

一樣看數線，本來a_1在b_1的右邊，現在a_2變小，因爲 B 會介於b_1和b_2之間，所以a_2的範圍必須是

$$0<a_2<b_1$$

才有可能產生反轉。

(2)兩個都越來越好

結果就是之前討論的極端分母解所表示出的範圍。

（六）無窮項數的探討

我們把附錄中的辛普森數列轉變成小數，發現一直跑下去，當兩邊都收斂時，左邊和右邊的值會趨近於同一個數值，便將此現象稱爲$\boxed{\text{相同極限性質}}$。

預備知識

1.根據辛普森遞迴式的運算規則，可知$\forall n \in \mathrm{N} \geq 2$

$$\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i} < \frac{q_{n+1}}{p_{n+1}} \quad , \quad \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \frac{s_{n+1}}{r_{n+1}}$$

2.從數線的觀點來看，因爲

$$\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i} < \frac{q_{n+1}}{p_{n+1}} \quad , \quad \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \frac{s_{n+1}}{r_{n+1}}$$

所以

$$\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i} < \frac{q_{n+1} + \sum_{i=1}^{n} q_i}{p_{n+1} + \sum_{i=1}^{n} p_i} = \frac{\sum_{i=1}^{n+1} q_i}{\sum_{i=1}^{n+1} p_i} < \frac{q_{n+1}}{p_{n+1}} \quad ,$$

$$\frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \frac{s_{n+1} + \sum_{i=1}^{n} s_i}{r_{n+1} + \sum_{i=1}^{n} r_i} = \frac{\sum_{i=1}^{n+1} s_i}{\sum_{i=1}^{n+1} r_i} < \frac{s_{n+1}}{r_{n+1}}$$

3.

因爲 $\dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \dfrac{s_{n+1}}{r_{n+1}}$ ，而且知道 $<\dfrac{s_n}{r_n}>$ 會收斂至 β ，會得到 $\dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \beta$ ，因爲當 $\dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} > \beta$ 時，根據辛普森運算法則，$\dfrac{s_{n+1}}{r_{n+1}}$ 就會大於 β ，從數線的觀點來看，

兩個數都大於 β ，分子相加除以分母相加必會在兩數之間，使 $\dfrac{\sum_{i=1}^{n+1} s_i}{\sum_{i=1}^{n+1} r_i} > \dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} > \beta$ ，所以下一項 $\dfrac{s_{n+2}}{r_{n+2}}$ 也會大於 β，很顯然的 $<\dfrac{s_n}{r_n}>$ 並不會收斂

至 β ，所以 $\dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i} < \beta$ 必成立。

證明

$$\frac{q_1}{p_1} > \frac{s_1}{r_1}$$
$$\frac{q_2}{p_2} > \frac{s_2}{r_2}$$
$$\frac{q_3}{p_3} > \frac{s_3}{r_3}$$
....
....
....
$$\frac{q_n}{p_n} > \frac{s_n}{r_n}$$
....
$. \downarrow ... \downarrow$
$..\alpha....\beta$

若左項和右項都收斂，假設左項收斂至 α ，右項收斂至 β

1. 若 $\alpha < \beta$

我們可假設 $\dfrac{q_n}{p_n} = \alpha + \varepsilon_{n\alpha}$ ，$\dfrac{s_n}{r_n} = \beta + \varepsilon_{n\beta}$ ，

而得到 $\alpha = \dfrac{q_n}{p_n} - \varepsilon_{n\alpha}$ ，$\beta = \dfrac{s_n}{r_n} - \varepsilon_{n\beta}$ ，

又因爲 $\alpha < \beta$ ，所以 $\dfrac{q_n}{p_n} - \varepsilon_{n\alpha} < \dfrac{s_n}{r_n} - \varepsilon_{n\beta}$ 。只要 n 夠大，$\varepsilon_{n\alpha}$ 和 $\varepsilon_{n\beta}$ 皆會趨近於零， 所以在 n 很大時，必可以找到一組 $\dfrac{q_n}{p_n}$ 和 $\dfrac{s_n}{r_n}$ ，使得 $\dfrac{q_n}{p_n} < \dfrac{s_n}{r_n}$ ，顯然和我們要求的矛盾。

2. 若 $\alpha > \beta$

一樣先假設 $\dfrac{q_n}{p_n} = \alpha + \varepsilon_{n\alpha}$ ，$\dfrac{s_n}{r_n} = \beta + \varepsilon_{n\beta}$ ，只要 n 夠大，一定能找到一個 n 滿足 $\forall i \in \mathrm{N} > n$ ，

$\left|\frac{q_i}{p_i}-\alpha\right|<\left|\varepsilon_{n\alpha}\right|$ 且 $\left|\frac{s_i}{r_i}-\beta\right|<\left|\varepsilon_{n\beta}\right|$ 且 $\alpha-\beta>\left|\varepsilon_{n\alpha}\right|+\left|\varepsilon_{n\beta}\right|$，即

$\left|\varepsilon_{i\alpha}\right|<\left|\varepsilon_{n\alpha}\right|$ 且 $\left|\varepsilon_{i\beta}\right|<\left|\varepsilon_{n\beta}\right|$ 且 $\alpha-\beta>\left|\varepsilon_{n\alpha}\right|+\left|\varepsilon_{n\beta}\right|$

我們令這個滿足的數爲C，若逆轉成立的話，則可知

$\frac{\sum_{i=1}^{C}q_i}{\sum_{i=1}^{C}p_i}<\frac{\sum_{i=1}^{C}s_i}{\sum_{i=1}^{C}r_i}$，若再繼續做下去，再根據前面的預備知識可知

$\frac{\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}p_i}=\frac{\sum_{i=1}^{C}q_i+\sum_{i=k+1}^{n}q_i}{\sum_{i=1}^{C}p_i+\sum_{i=k+1}^{n}q_i}>\frac{(\sum_{i=1}^{C}q_i)+(n-C)q_{C+1}}{(\sum_{i=1}^{C}p_i)+(n-C)p_{C+1}}>\frac{q_{C+1}}{p_{C+1}}$，當

$$\lim_{n\to\infty}\frac{(\sum_{i=1}^{C}q_i)+(n-C)q_{C+1}}{(\sum_{i=1}^{C}p_i)+(n-C)p_{C+1}}=\lim_{n\to\infty}\frac{(n-C)q_{C+1}}{(n-C)p_{C+1}}=\frac{q_{C+1}}{p_{C+1}}$$

所以只要 n 夠大，可得$\frac{\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}p_i}>\frac{q_{C+1}}{p_{C+1}}=\alpha+\varepsilon_{(C+1)\alpha}$，而

$\left|\varepsilon_{(C+1)\alpha}\right|<\left|\varepsilon_{C\alpha}\right|$，所以得

到$\varepsilon_{(C+1)\alpha}>-\left|\varepsilon_{C\alpha}\right|$，所以$\alpha+\varepsilon_{(C+1)\alpha}>\alpha-\left|\varepsilon_{C\alpha}\right|$。

又因爲$\alpha-\beta>\left|\varepsilon_{C\alpha}\right|+\left|\varepsilon_{C\beta}\right|$，所以$-\left|\varepsilon_{C\alpha}\right|>\beta+\left|\varepsilon_{C\beta}\right|-\alpha$，再代入上式，得

$\alpha+\varepsilon_{(C+1)\alpha}>\alpha-\left|\varepsilon_{C\alpha}\right|>\beta+\left|\varepsilon_{C\beta}\right|$，而得到

$$\frac{\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}p_i}>\frac{q_{C+1}}{p_{C+1}}=\alpha+\varepsilon_{(C+1)\alpha}>\alpha-\left|\varepsilon_{C\alpha}\right|>\beta+\left|\varepsilon_{C\beta}\right|$$

再根據預備知識 3. $\frac{\sum_{i=1}^{n}s_i}{\sum_{i=1}^{n}r_i}<\beta$

可得$\dfrac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i} > \dfrac{q_{C+1}}{p_{C+1}} = \alpha + \varepsilon_{(C+1)\alpha} > \alpha - \left|\varepsilon_{C\alpha}\right| > \beta + \left|\varepsilon_{C\beta}\right| > \beta > \dfrac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} r_i}$

矛盾！！！

所以根據辛普森遞迴式做出來的數列組，左右都會收斂至相同的數值。

在分母有上限時可否無限制逆轉

根據極端分母解的定義，可求出$\dfrac{q_1}{p_1} = \dfrac{1}{K}$，$\dfrac{s_1}{r_1} = \dfrac{0}{1}$，$\dfrac{q_2}{p_2} = \dfrac{1}{1}$，$\dfrac{s_2}{r_2} = \dfrac{K-1}{K}$，

因爲極端分母解是最先出現的解，所以只要出現，就可無限制逆轉。

$\forall n \in \mathrm{N}$

令$\dfrac{q_{2n+1}}{p_{2n+1}} = \dfrac{q_1}{p_1} = \dfrac{1}{K}$ ， $\dfrac{q_{2n}}{p_{2n}} = \dfrac{q_2}{p_2} = \dfrac{1}{1}$

$\dfrac{s_{2n+1}}{r_{2n+1}} = \dfrac{s_1}{r_1} = \dfrac{0}{1}$ ， $\dfrac{s_{2n}}{r_{2n}} = \dfrac{s_2}{r_2} = \dfrac{K-1}{K}$

所以可以得到

$$A_{2n} = \frac{\sum_{i=1}^{2n} q_i}{\sum_{i=1}^{2n} p_i} = \frac{2n(2)}{2n(K+1)} = \frac{2}{K+1} \quad , \quad B_{2n} = \frac{\sum_{i=1}^{2n} s_i}{\sum_{i=1}^{2n} r_i} = \frac{2n(K-1)}{2n(K+1)} = \frac{K-1}{K+1}$$

當$K \geq 4$時，就可讓$B_{2n} > A_{2n}$而產生逆轉

$$A_{2n+1} = \frac{\sum_{i=1}^{2n+1} q_i}{\sum_{i=1}^{2n+1} p_i} = \frac{2n(2)+1}{2n(K+1)+K} = \frac{4n+1}{K(2n+1)+1} ,$$

$$B_{2n} = \frac{\sum_{i=1}^{2n+1} s_i}{\sum_{i=1}^{2n+1} r_i} = \frac{2n(K-1)}{2n(K+1)+1} = \frac{2nK-2n}{2nK+2n+1}$$

若$B_{2n+1} > A_{2n+1} \Leftrightarrow \dfrac{2nK-2n}{2nK+2n+1} > \dfrac{4n+1}{K(2n+1)+1}$

$$\Leftrightarrow K^2(2n+1)(2n) - K(2n+1)(2n) + 2nK - 2n > K(4n+1)(2n) + (2n+1)(4n+4)$$

$$\Rightarrow K > \frac{n(6n-1)+\sqrt{n(68n^3+20n^2+13n+2)}}{2n(2n+1)} \quad or \quad K < \frac{n(6n-1)-\sqrt{n(68n^3+20n^2+13n+2)}}{2n(2n+1)} \text{(負不合)}$$

$$\Rightarrow K > \frac{n(6n-1)+\sqrt{n(68n^3+20n^2+13n+2)}}{2n(2n+1)} = \frac{6n-1}{4n+2} + \frac{\sqrt{68n^2+20n+13+\frac{2}{n}}}{4n+2}$$ 經

由微分後可發現 n 在大於 1 時 $\frac{6n-1}{4n+2}+\frac{\sqrt{68n^2+20n+13+\frac{2}{n}}}{4n+2}$ 是嚴格遞增，且

$$\lim_{n\to\infty}\left(\frac{6n-1}{4n+2}+\frac{\sqrt{68n^2+20n+13+\frac{2}{n}}}{4n+2}\right)=\lim_{n\to\infty}\left(\frac{6n}{4n}+\frac{\sqrt{68}n}{4n}\right)=\frac{6+\sqrt{68}}{4}\approx 3.562<4$$

所以當 $K\geq 4$ 時，也可讓 $B_{2n+1}>A_{2n+1}$ 產生逆轉

所以當分母上限 ≥ 4 時，就可以無限制逆轉

（七）辛普森的分類

當場次在兩場次以上時，除了原先定義的『前 n 項皆有辛普森逆轉』的辛普森數列組之外，還有任意連續兩項、三項或是任意兩項、三項等等。爲了以下討論方便，在此先將辛普森數列組分爲三個類型。

型 1	只須滿足『前 n 項皆有辛普森逆轉』。
型 2	只須滿足『總項有辛普森逆轉』。
型 3	只須滿足『任意 n 項皆有辛普森逆轉』， 我們稱之爲 $\boxed{\text{超辛普森數列組}}$。

型 1 的數列演算法爲先前就已經探討過，直接就來討論型 2，以下是當分母有最高限制時，打 3 場的型 2 辛普森數列組解的個數。

-----------	型 1 三場	型 1 四場	型 2 三場	型 2 四場	型 3 三場	型 3 四場
$K=1$	0	0	0	0	0	0
$K=2$	0	0	0	0	0	0
$K=3$	0	0	0	0	0	0
$K=4$	20	56	60	453	0	0
$K=5$	705	5670	2036	37664	0	0
$K=6$	7697	139428	21320	821364	8	0
$K=7$	56340		155344	10780800	199	0
$K=8$	277947		766867		1535	
$K=9$	1123357		3095580		9878	
$K=10$	3790929		10462363		41876	

參、 結論及展望

一、 結論

以棒球爲例，設 K 爲每場打擊數上限。

1.對於兩場比賽來說，至少一場的上場次數大於等於 4 才可能造成辛普森逆轉。
2.對於兩場比賽來說，極端分母解所包含區域中格子點個數是最多的，故若有解則其區域必含格子點。
3.對於兩場比賽，極端分母解 $A=\dfrac{2}{K+1}, B=\dfrac{K-1}{K+1}$，故打擊數趨近於無窮大時，B 趨近於 1，A 趨近於 0。
4.在兩場比賽中，只要每一場的打擊率均大於對手的 $\dfrac{K-1}{2}$ 倍，就保證不會被逆轉。
5.給定一組起始數列組，則可用「辛普森遞迴式」產生連續不斷的辛普森數列組。
6.若 $<a_n>$ 和 $<b_n>$ 皆收斂，則 a_n 和 b_n 的值必須趨近於同一個數值，才可以無限項逆轉。
7.極端分母解所包含區域的格子點數 $=\dfrac{K^2-2K-3-(K+1,2)}{2}$
8.給定 a_1、a_2、b_1，則可用新 Pick 公式估計出 b_2 的解的個數。

二、 展望

1. 我們希望能擴展討論範圍，討論項數 n 大於 2 時的狀況。
2. 改變比賽的人數，使比賽的人大於 2。
3. 探討出對於任意正整數 K，所產生的辛普森逆轉個數的一般式。
4. 求出只知道部分資訊的情況下，被逆轉的機率。
5. 探討各種辛普森解的個數之間的關係，希望能進而推導出一般式。

肆、 參考資料

1. Simpson's Paradox (Stanford Encyclopedia of Philosophy)
   http://plato.stanford.edu/entries/paradox-simpson/
2. David S.Moore 鄭惟厚 譯《統計學的世界》(Statistics：Concepts and Controversies),天下文化書坊,2000 年

附錄

辛普森數列組	逆轉後的辛普森
$\frac{3}{7} > \frac{7}{25}$	———————
$\frac{4}{8} > \frac{78}{157}$	$\frac{7}{15} < \frac{85}{182}$
$\frac{9}{17} > \frac{118}{223}$	$\frac{16}{32} < \frac{203}{405}$
$\frac{19}{35} > \frac{242}{446}$	$\frac{35}{67} < \frac{445}{851}$
$\frac{36}{68} > \frac{470}{888}$	$\frac{71}{135} < \frac{915}{1739}$
$\frac{73}{137} > \frac{959}{1800}$	$\frac{144}{272} < \frac{1874}{3539}$
$\frac{147}{275} > \frac{1938}{3626}$	$\frac{291}{547} < \frac{3812}{7165}$
$\frac{292}{548} > \frac{4150}{7789}$	$\frac{583}{1095} < \frac{7962}{14954}$

辛普森數列組		逆轉後的辛普森	
a	b	A	B
0.428571	0.28		
0.5	0.496815	0.466667	0.467033
0.529412	0.529148	0.5	0.501235
0.542857	0.542601	0.522388	0.522914
0.529412	0.529279	0.525926	0.526164
0.532847	0.532778	0.529412	0.529528
0.534545	0.534473	0.531993	0.532031
0.532847	0.532803	0.53242	0.532433
0.533273	0.533253	0.532847	0.532854
0.533485	0.532878	0.533167	0.533173
0.533273	0.533267	0.533291	0.533221
0.533326	0.533325	0.533273	0.533274

$K=6$ 的超辛普森數列組

1/5 > 0/1 and 4/6 > 3/5 and 1/1 > 5/6

1/6 > 0/1 and 2/4 > 2/5 and 1/1 > 5/6

1/6 > 0/1 and 2/5 > 2/6 and 1/1 > 4/5

1/6 > 0/1 and 2/5 > 2/6 and 1/1 > 5/6

1/6 > 0/1 and 2/5 > 2/6 and 2/2 > 5/6

1/6 > 0/1 and 3/5 > 2/4 and 1/1 > 5/6

1/6 > 0/1 and 4/6 > 3/5 and 1/1 > 5/6

1/6 > 0/2 and 4/6 > 3/5 and 1/1 > 5/6

評語

研究精神可嘉，思考週延，表達方式也不錯，可惜受限於題目本身，較難多著力於數學的深度。