第三部 機遇

第19章 模擬

授課教師: 吳漢銘
淡江大學 數學系 資統組
http://www.hmwu.idv.tw
hmwu@mail.tku.edu.tw

個案研究

- 賽馬時，馬匹的位置靠近賽場內側的比較有利。為了公平，檔位是在賽前抽籤隨機決定。
- 2007年，俄亥俄州賽馬委員會注意到，某位練馬師在35場比賽裡面，有30次抽到內側三個位置之一。(一場比賽可能有6到10匹馬參賽，其中多半時候是9匹馬參賽。因此對大部分的比賽來說，抽中內側三個位置之一的機率是1/3 。)
- 委員會認為，在35場比賽當中，至少有30場抽到內側三個位置之一的機率非常小。因此懷疑有人作弊。
- 但是該練馬師在數年之內已讓他的馬參加了近1000場比賽，本來在某個時期就不免出現連續35場比賽當中至少有30次抽到內側三個位置之一的情況。表面看來是令人驚異的巧合之一。或許作弊的指控並沒有根據。
- 本章教你計算在1000場比賽當中，會出現連續35場比賽至少有30次抽到內側三個位置之一的機率。

機率從何而來

- 擲銅板時得到正面和反面的機率都很接近1/2。原則上來說，這些機率是根據擲很多次的數據得來的。
- 擲一個銅板10次，會出現連續3次正面的機率要如何計算呢？
    - 可以利用描述擲銅板情況的模型，計算出這個機率。
    - 一旦我們根據數據，找到擲銅板的機率模型後，就不需在每次要找一個新事件的機率時，都再重新開始。
- 用機率模型的一大優點，是讓我們可以只先對一些如「擲一次銅板得到正面」的簡單事件分配機率，就能夠計算一些複雜事件的機率。不論機率模型反映的是根據數據得到的機率還是個人機率，上面的優點都成立。
- 只要我們有機率模型，就可以用電腦來模擬重複許多次的狀況。

模擬

- 利用隨機數字表或者電腦軟體中的隨機數字，來模仿機遇現象就叫做模擬(simulation)。
- 科學家與工程師用「模擬」，替複雜的情形算出機率。
    - 有了可靠的機率模型，模擬是找出複雜事件發生機率的有效工具。
- 模擬會迫使我們仔細思考機率模型的意義。
- 模擬困難的部分: 建立模型。
- 模擬容易的部分: 利用電腦多次重複的結果，計算機率。
- 一個事件在這些重複結果中發生的比例，遲早會接近它的機率，所以模擬可以對機率做適當的估計。

例1: 如何執行模擬

擲一個銅板10次。結果中會出現至少3個連續正面或是至少3個連續反面的機率是多少？

- 第1步：提出機率模型。
  - 每一次擲，正面和反面的機率各為0.5。
  - 每次投擲之間，彼此是獨立的。(知道某次的結果，不會改變其他任一次所擲結果的機率。)
- 第2步：分配隨機數字以代表不同的結果。
  - 表A中的隨機數字會以符合第1步之機率的方式，來代表各種結果。(表中的每一個數字會是0、1、2、3、4、5、6、7、8、9中任一個的機率都是0.1，而且表中數字之間是互相獨立的。)
  - 擲銅板結果分配數字的方法之一：
    - 每個數字模擬擲一次銅板的結果。
    - 奇數代表正面，偶數代表反面。
  - 這樣子的分配可行，是因為5個奇數使正面機率恰好是5/10。表中的連續數字可模擬多次獨立的投擲。

例1: 如何執行模擬 (conti.)

- 第3步：模擬多次重複。
  - 表A中的10個連續數字模擬了一組擲銅板10次的狀況。
  - 在表A中讀取許多組的連續10個數字，就模擬了重複多次的狀況。
  - 每次重複時，要記錄事件(至少連續3個正面或反面)有沒有發生。

	第一組	第二組	第三組
字	1 9 2 2 3 9 5 0 3 4	0 5 7 5 6 2 8 7 1 3	9 6 4 0 9 1 2 5 3 1
/反面	正正反反正正正反正反	反正正正 反反反 正正正	正反反反正正反正正正
續3個？	有	有	有

  - 一共做了25組，其中23組有連續至少3個或更多的正面或反面。用比例來估計這個事件的機率為23/25=0.92。
  - 可以利用電腦重複做好幾千次，得到真正的機率大約是0.826。
- 大部分的人認為連續正面或反面不太容易發生，但連我們這個很小規模的模擬結果所顯示的，擲10次銅板大部分時候都會出現連續3個同樣的結果，這已足以導正我們直覺的錯誤。

獨立

- 整個模擬過程最困難的部分，通常就是建立機率模型。
- 擲銅板這個例子中的模型能代表許多機率問題，因為它是由許多**獨立**的試驗(擲銅板)構成，而每次試驗都有一樣的可能結果及機率。
  - 投籃10次和觀察10個孩子的性別，也有類似的模型，也可用幾乎一樣的方法模擬。
- **獨立**
  如果「知道兩個隨機現象其中之一的結果」，並不會改變另一個結果的機率，就稱這個兩**隨機現象是獨立的**（independent ）。
- 獨立就和機率的其他性質一樣，一定要重複觀察很多次，才能證實。
- 重複投擲銅板應該是獨立的(銅板沒有記憶)，經過觀察後也發覺事實確是如此。
  - 要說一個籃球員的前後投球之間彼此獨立，就不那麼可信，不過觀察顯示，至少滿接近獨立的。

例2: 分配數字以執行模擬

- (a) 從就業人口占70%的一群人當中，隨機選一個人。一個數字代表一個人：

  0, 1, 2, 3, 4, 5, 6 = 有就業者
  7, 8, 9 =未就業者

- (b) 從就業人口占73%的一群人當中，隨機選一個人。現在得用兩個數字來模擬一個人：

  00, 01, 02, …, 72 = 有就業者
  73, 74, 75, …, 99 = 未就業者

- (c) 從50%有就業、20%失業以及30%不屬於勞動人口的一群人中隨機選一人。一個數字模擬一個人：

  0, 1, 2, 3, 4 =就業者
  5, 6 =失業者
  7, 8, 9 =不屬於勞動人口

Ex19.1 隨機抽牌

- 一副標準撲克牌的52張牌裡面，有13張黑桃、13張紅心、13張方塊和13張梅花。
- 如果你要模擬從一副52張撲克牌裡面隨機抽一張、記錄其花色(黑桃、紅心、方塊或梅花)，應該要怎樣分配隨機數字？

獨立性

- 獨立性質(或「不獨立」性質)必須經由重複觀察隨機現象許多次之後才能證實。
- 兩個隨機現象不獨立要怎樣才能看出來呢？
  - 舉例：我們要怎樣才知道，一枚均勻銅板(正反面的機率都是0.5)的各次投擲結果之間並不獨立呢？
- 考慮運用「獨立」的定義：
  - 當我們考慮擲一枚均勻銅板很多次的一序列結果時，可以計算整個序列當中，相鄰的前後兩次出現相同結果之發生比例。
  - 即正面之後接著正面、或反面之後接著反面的發生頻率。
  - 如果各次投擲之間互相獨立(如此則每次投擲結果不會影響下一次結果的機率)，觀察的次數又夠多的話，上述比例應該接近0.5。

例3: 研究獨立性

- 假設我們擲一枚均勻銅板15次，得到以下一序列結果：

  順序：1 2 3 4 5　6 7 8 9 10 11 12 13 14 15

  結果：正正正反正　正反反反反　反　正　正　反　反

- 前14次的投擲結果當中有9次，接下來擲出的結果相同。前後兩次出現相同結果的發生比例為9/14=0.64
- 對於擲這麼少次的結果來說，這個比例距0.5 的差距不算大。
- 如果正面之後接著正面、或反面之後接著反面的發生比例很接近0.5的話，卻未必代表各次投擲之間互相獨立。

獨立性

- 舉例：假設依照以下模式把銅板正面朝上或反面朝上依序擺設：

順序：	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
結果：	正	正	反	反	正	正	反	反	正	正	反	反	正	正	反

  - 先是兩個正面、接著兩個反面、再來兩個正面，依此類推。如果知道前幾次的結果，我們就確知下一次的結果是什麼。
  - 連續結果之間並不獨立。然而若觀察前14次的結果，此序列當中正面之後接著正面、或反面之後接著反面的發生比例是7/14=0.5。
- 當上述比例和0.5差很多時，可辨別出不獨立的情況。但若比例接近0.5，則並不能做出「獨立」的結論。
- **利用相關係數**：計算兩個隨機現象觀察數據的相關係數。
  - 如果這兩個隨機現象之間互相獨立，就不會有直線關聯，相關係數應接近0。
  - 如果兩個隨機現象之間的相關係數是0 ，並不代表二者之間必定互相獨立。

例4: 我們要生女兒

一對夫婦計劃生孩子生到有女兒才停，或生了三個就停止。他們會擁有女兒的機率是多少？

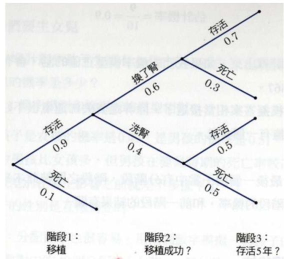
- 第1 步：機率模型
  - 每一個孩子是女孩的機率是0.49 ，是男孩的機率是0.51。各個孩子的性別是互相獨立的。
- 第2 步：分配隨機數字。
  - 用兩個數字模擬一個孩子的性別: 00, 01, 02, …, 48 ＝ 女孩; 49, 50, 51, …, 99 ＝ 男孩
- 第3 步：模擬生孩子策略
  - 從表A當中讀取一對一對的數字，直到這對夫婦有了女兒，或已有三個孩子。

6905	16	48	17	8717	40	9517	845340	648987	20
男女	女	女	女	男女	女	男女	男男女	男男男	女
+	+	+	+	+	+	+	+	−	+

- 10次重複中，有9次生女孩。會得到女孩的機率的估計是9/10=0.9。
- 如果機率模型正確的話，用數學計算會有女孩的真正機率是0.867。(我們的模擬答案相當接近了。除非這對夫婦運氣很不好，他們應該可以成功擁有一個女兒。)

例5: 腎臟移植

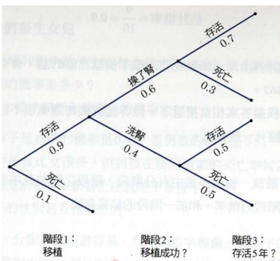
腎臟移植的病人資料：撐過移植手術的占90%，另外10%會死亡。在手術後存活的人中有60%移植成功，另外的40%還是得回去洗腎。五年存活率對於換了腎的人來說是70% ，對於回去洗腎的人來說是50%。計算能活過五年的機率。



- 第1 步：機率模型如圖19.1。
- 第2 步：對每個結果分配數字:
  - 階段1:
    0 ＝ 死亡;
    1, 2, 3, 4, 5, 6, 7, 8, 9 ＝ 存活
  - 階段2:
    0, 1, 2, 3, 4, 5 ＝移植成功
    6, 7, 8, 9 ＝ 仍需洗腎

例5: 腎臟移植 (conti.)

- 階段3，換了腎：
  0, 1, 2, 3, 4, 5, 6 ＝ 存活五年
  7, 8, 9 ＝ 未能存活五年
- 階段3 ，洗腎：
  0, 1, 2, 3, 4 ＝ 存活五年
  5, 6, 7, 8, 9 ＝ 未能存活五年
  (階段3的數字分配，和階段2的結果有關。所以二者間不獨立。)



- 第3步：多次模擬的結果。

	第1次	第2次	第3次	第4次
階段1	3 存活	4 存活	8 存活	9 存活
階段2	8 洗腎	8 洗腎	7 洗腎	1 新腎
階段3	4 存活	4 存活	8 死亡	8 死亡

- 在4次模擬中，有2次存活超過5年。
- 經由電腦多次重覆計算，五年存活機率是0.558。

Ex19.2 隨機抽牌

- 一副標準撲克牌的52張牌裡有13張黑桃、13張紅心、13張方塊和13張梅花。
- 執行模擬，估計從一副標準撲克牌中隨機抽出兩張牌時，兩張同花色的機率。
- 可依照例5的步驟來執行。共執行10次模擬，並利用你得到的結果來估計機率。

個案研究-隨堂考

- 賽馬時，假設每場比賽都是9匹馬參賽，所以如果用隨機方式決定的話，抽中內側三個位置之一的機率是1/3。
- 描述如何利用模擬來估計，在連續1000場比賽當中，會出現連續35場比賽至少有30次抽到內側三個位置之一的機率。
- 先說明你要怎樣模擬一組1000場的比賽，以及要怎樣判斷在這1000場比賽當中，是否出現連續35場比賽至少有30次抽到內側三個位置之一的情況。