統計，讓數字說話！

Statistics

concepts and controversies

Chapter 7

機率：討論可能性的語言

2009/07/11

隨機並不是混亂，而是一種秩序。

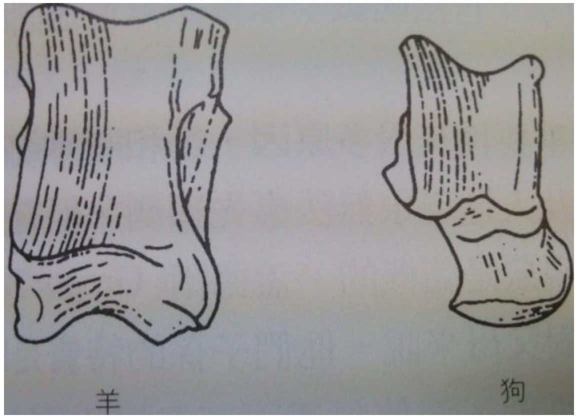
我們最該操心的反而是我們安排的隨機 ——
不是上帝的骰子，而是賭場的骰子。

隨機

隨機（Random）：

符合以下情況的現象，我們稱為隨機。

- 確實的結果事前無法預知。
- 雖然如此，但有可預測的長期型態，可以用很多次試驗結果的分佈來描述。



- Ex. 1:

  擲骰子、發洗好的牌、轉輪盤。

- Ex. 2:

  古代-擲骨頭（距骨是相當規則的實心骨頭，取自動物腳跟）

隨機 (Cont.)

- 機率的重要性有很多原因，卻和賭博或產生數據的關係不大。
- 許多自然以及人造現象無法事先預測，但有長期型態，此現象為隨機。
- 機率用來描述遺傳學、物理學及其研究領域的各種現象。
- Ex.

孟德爾：觀察父母與子孫的特質為隨機，遺傳科學因而發展。

什麼是機率？

- 機率是從觀察到有些現象是隨機的開始。
- 隨機是對只有長時間才出現的某種法則之描述。
- Ex. 擲銅板

  把銅板丟到半空中，掉下來時會是正面還是反面？有時正面有時反面。我們沒法說出下次的結果會是什麼。但是如果我們擲很多次，就會出現一種型態。

擲銅板試驗

- **Ex. 1**

  法國自然主義者佈豐伯爵（Count Buffon；18世紀）擲銅板4040次。結果：2048個正面，得正面比率為2048/4040=0.5069。

- **Ex. 2**

  英國統計學家皮爾生（Karl Pearson；1990）神勇的擲銅板24,000次。結果：12,012次正面，得正面比率為0.5005。

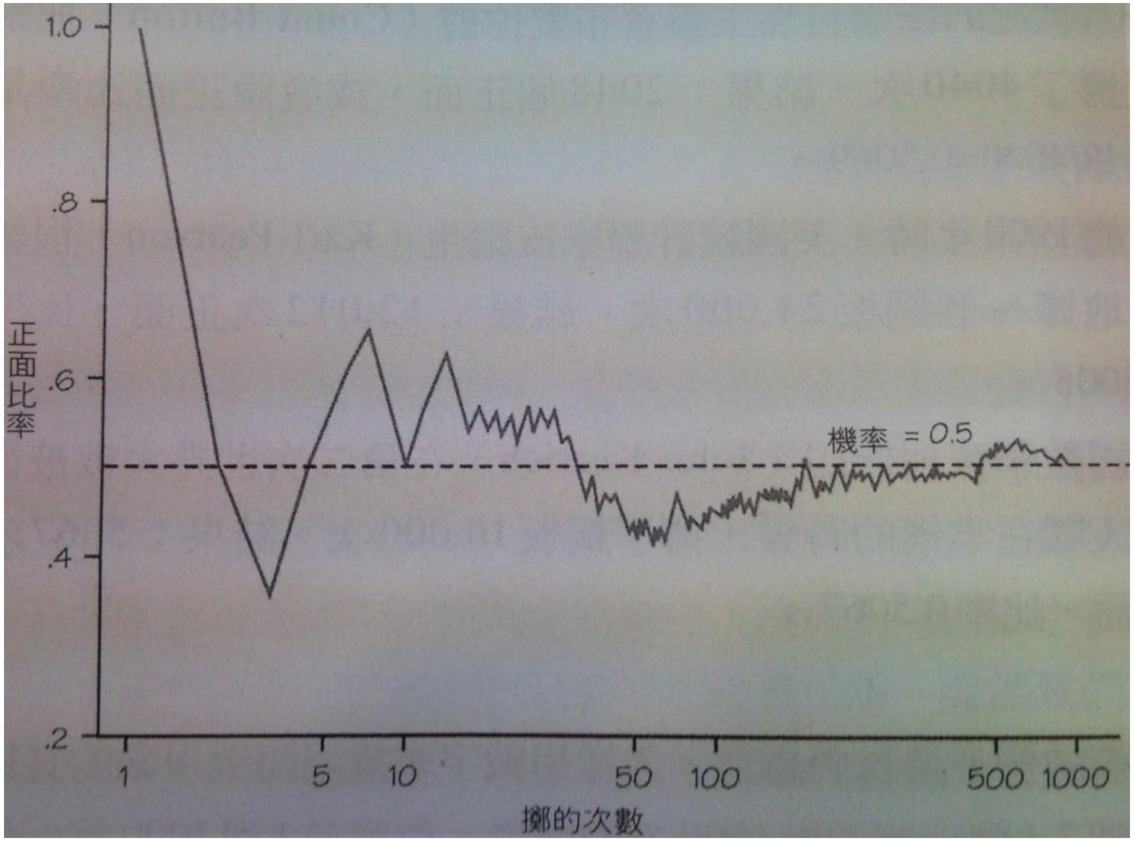
- **Ex. 3**

  英國數學家柯瑞屈（John Kerrich）在第二次世界大戰被德國人關於牢中，擲銅板10,000次。結果：5067次正面，比率0.5067。

機率

機率（Probability）

- 一個隨機現象任一結果的機率是：在重複很多很多次的情形下，該結果應會出現的比率。



死亡的機率

- **Ex. 死亡的機率**
  我們無法預測特定的人是否明年會死。但是如果我們觀好幾百萬人，死亡就是隨機了。25-34歲的男性中，明年會死的比率差不多為0.0021，即為年輕男子明年會死的「機率」；對於同年齡層婦女則為0.0007。

- →保險公司若賣人壽險給25-34歲的人，公司即知：賣男性的保險大約有0.21%要理賠， 賣女性約有0.07%要理賠。因此男性的保險費需多收一些，因為理賠的比率較高。

機率模型

機率模型（Probability model）

- 經由對不同的結果分配機率，來描述隨機現象。
- 機率理論的數學，始於列出所有正規機率模型該有的性質。只需幾個簡單的規則。

機率規則

任何對隨機現象各結果機率的合理分配，必須滿足以下規則：

- (A) 任何機率都介於0與1之間的數。。
- (B) 所有可能的結果合併起來，機率應該是1 。

一個事件（event）是任何一組結果的集合，事件的機率也遵循以下規則：

- (C) 一個事件不發生的機率是1減去該事件發生的機率。
- (D) 如果兩個事件當中沒有共同的結果，則該兩個事件中至少有一個會發生的機率是該兩事件個別機率的和。

隨機選取

- **Ex.** 隨機選取

  隨機選取一個25~29歲的美國婦女，被選取婦女的機會均相等。若選中的婦女已婚的機率？

  由普查局的婦女資料：

結果	單身	已婚	寡居	離婚
機率	0.352	0.577	0.003	0.068

隨機選取 (Cont.)

- 未婚的機率：

  0.352+0.003+0.068=0.423

- 運用規則(C)：

  未婚的機率=1-已婚的機率=1-0.577=0.423

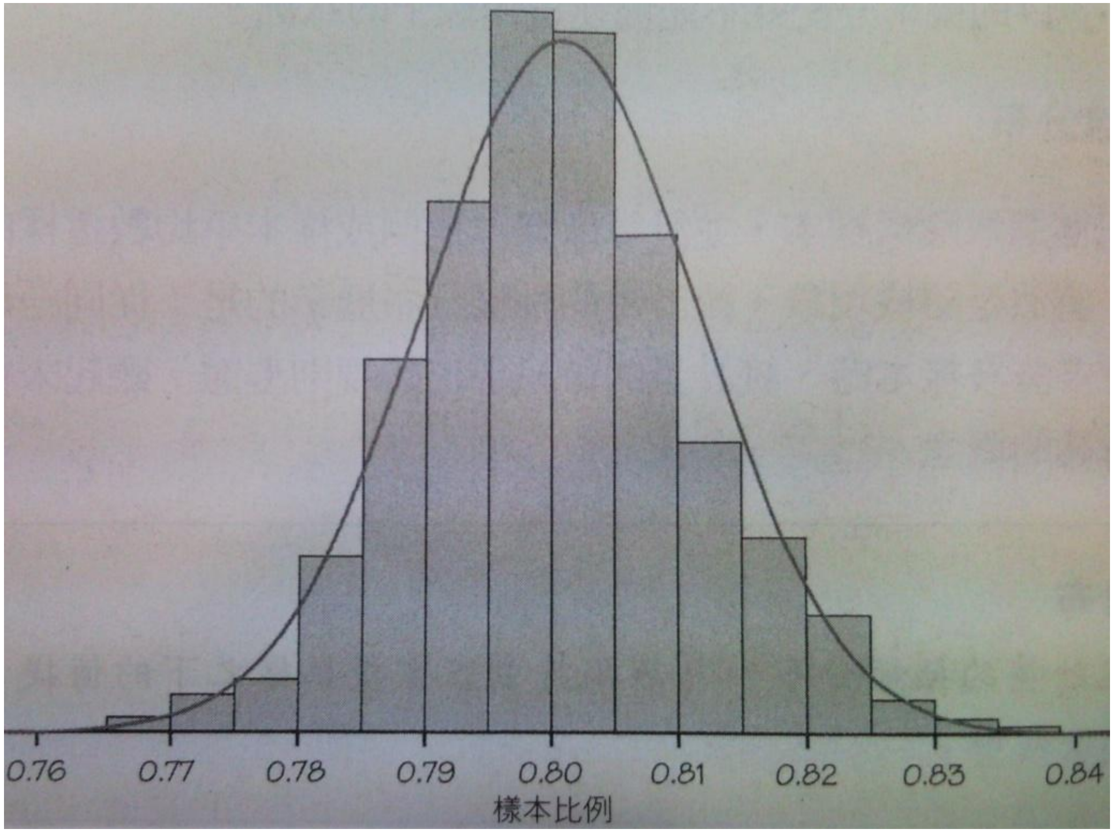
又見抽樣分佈

- 從母體選取隨機樣本，並計算像樣本比例或樣本平均數的統計量，即為隨機現象。
- 統計量的抽樣分佈展示的是：從同一母體抽取許多許多樣本時，統計量的值呈現出的規則型態。

抽樣分佈

抽樣分佈：

- 一個統計量的抽樣分佈，為該統計量在重複抽樣之下的值提供了機率模型。



怎樣分配機率

- 對每一個個別結果分配機率，使得規則(A)和(B)都滿足。任何事件的機率，是該事件所包含的個別結果之機率總和。
- 根據68-95-99.7規則或者表B，用常態曲線直接對結果的集合分配機率。

對於可能性的探討 (1/3)

- 機率的概念是，隨機現象「長期來說」是有規則的。不幸的是，我們對於隨機的直覺卻是說，隨機現象應該在短期就有規則。
- 當規則沒出現時，我們就會找解釋，而不把它當做是機遇變異。對了解機遇而言，我們的直覺會給我們很差的指引。

對於可能性的探討 (2/3)

- Ex. 什麼看起來像隨機的？

把一個銅板擲6次，並且把每次的正反記錄下來，以下那個結果比較可能發生？

“正反正反反正”　”反反反正正正”

幾乎每個人認為「正反正反反正」比較容易發生，因為「反反反正正正」看起來不隨機。

→事實上，兩者發生的機會一樣大。

對於可能性的探討 (3/3)

- **Ex.** 我們要個兒子

「親愛的艾比」專欄~有8個女兒的心煩意亂的母親的信，原來她和他先生只準備要4個孩子的，可是當4個都是女孩的時候他們就再試一次~又試一次，又試一次。在連續7個女兒之後，連她的醫生都向她保證：「根據平均數定律，成功和失敗會是100比1。」不幸的是~的確發生了。

獨立

獨立（Independent）：

- 如果”知道兩個隨機現象其中之一的結果”不會改變另一個結果的機率，就稱這兩個隨機現象為**獨立**。

→獨立就和機率的其他性質一樣，一定要**觀察很多次的重複才能證實**。

→重複擲銅板應該是獨立（銅板**沒有記憶**），經過觀察也證明如此。

→若要說一個籃球員的前後投球之間彼此獨立，就不那個可信，但是觀察顯示，至少滿接近獨立的。

機率不能夠這樣解釋

你可能會覺得下列兩句話的差別很小：

- "將一個平衡的銅板擲很多次以後，正面出現的比率會接近二分之一。"

  → (○)

- "將一個平衡的銅板擲很多次以後，正面出現的次數會接近總次數的二分之一。"

  → (×)

機率不能夠這樣解釋 (Cont.)

- **Ex. 擲一個銅板100,000次**

假設我們擲一個銅板十萬次，並記錄100次、1,000次、10,000次及100,000次之後的正面次數：

擲的次數	正面次數	正面比率	正面次數和總次數一半的差距
100	51	0.51	1
1,000	510	0.51	10
10,000	5,100	0.51	100
100,000	51,000	0.51	1,000

機率能用來度量風險嗎？

- 科學家常用不好的事件發生之機率來描述風險。一般個人及社會則似乎忽略了機率。我們對某些機率很低的風險反應過度，卻對某些更有機會發生的事疏於注意。

機率能用來度量風險嗎？ (Cont.)

- **Ex. 學校裡的石綿**

  高度暴露於石棉是危險。低度暴露的風險是低的，例如，如果學校的暖氣管周圍隔熱材中有石綿，學校裡的老師和學生的風險很低。一位老師如果在一個有典型石綿含量的學校工作三十年，他會因石綿得癌症的機率約百萬分之十五。開車的人一輩子當中，會死於車禍的機率大約是百萬分之一萬五千。也就是說，經常開車的風險是在有石綿的學校教書的風險的一千倍。

→有風險沒讓我們停止開車。

→風險小的石綿，卻引發大規模的清除運動。

賭場的優勢：期望值

- **Ex. 比較賭法**

  假設有人提供你以下兩種賭法，賭注一樣多：賭A的話，你贏就付你10元，而你贏的機率是1/2。賭B的話，你可贏10,000元，而贏的機會是1/10。

→即使A勝算大，多半人還是會選擇B，因為若贏，則B付的錢較多。

→光根據贏的機率決定賭那一個，會是很笨的方法。

賭場的優勢：期望值 (Cont.)

- **Ex.** 比較賭法

  如果玩很多次，這兩種賭法的平均報償會是多少？

→A的平均報償：

（10元×1/2）+（0元× 1/2）=5元

→ B的平均報償：

（10,000元× 1/10）+（0元× 9/10）=1000元

→如果玩多次，當然應選B賭法。

期望值

期望值（Expected value）

- 有數值結果的隨機現象之期望值，是每一個結果乘上它的機率、再對所有可能結果加總而得。
- 如果用符號表示，假設可能結果是$a_1, a_2, \ldots, a_k$，它們的機率是$p_1, p_2, \ldots, p_k$，則期望值是

$$\text{期望值}=a_1p_1+a_2p_2+\ldots a_kp_k$$

期望值 (Cont.)

- **Ex. 原始的紐約彩券**

每賣出一百萬張彩券，紐約州彩券就提出下列彩金：

1個	50,000美金獎
9個	5,000美金獎
90個	500美金獎
900個	50美金獎

若你買一張，你贏50,000元的機會是1/1,000,000。全部得獎彩券共1,000張。則你贏錢的期望值是？

→(50,000美金)(1/1,000,000)+(5,000美金)(9/1,000,000)
+(500美金)(90/1,000,000)+(50美金)(900/1,000,000)
+(0美金)(999,000/1,000,000)=0.185美金

大數法則

大數法則（low of large numbers）

- 如果有數值結果的隨機現象獨立地重複許多次，實際觀測到的結果之平均值會趨近期望值。
- 需要多大的次數？大數法則並無定多少的次數才能保證平均結果會接近期望值？則依隨機結果的”變異性”決定。
- Ex. 大數法則解釋了：為什麼對個人來說是消遣或是嗜好的賭博，對賭場來說卻是生意。很大數量的客人平均贏的錢會很接近期望值。賭場經營者事先算好了期望值，且知長期下來的營收入，無需做牌來保證利潤。

隨機並非混亂，而是一種秩序

Thanks for your attention