用機率
算錢財、未來與愛情

第四組（數四乙）

410831201　許藝瀠　　410831203　沈宥旻
410831215　陳映竹　　410831236　鄭光呂
410831238　張顥繻　　410831246　唐翊庭

content

01

錢錢篇－
生活中的機率

賭博真的賺錢嗎，世足買到破產真的可能翻身？

Question

銅板問題

設甲，乙二人在賭博，甲投擲二公正銅板且不讓乙看到結果。此時丙從旁經過，忍不住說他看見有一銅板朝上。甲聞言對丙說:你破壞了我們的賭局，我的朋友乙正要猜兩個銅板朝上的面是相同還是相異。試問丙提供的資訊是否對乙有幫助？乙應猜相同還是相異？

A：有幫助，相異

Answer

銅板可能的情況：{(上,上),(上,下),(下,上),(下,下)}

設$P(W)$為兩銅板所有可能情況的機率

$P(A)$為兩銅板朝上的面相同的機率

$P(B)$為兩銅板朝上的面相異的機率

W_1 ：{(上,上),(上,下),(下,上),(下,下)}

A ：{(上,上),(下,下)}

B ：{(上,下),(下,上)}

$$P(A)=\frac{2}{4}=\frac{1}{2},P(B)=\frac{2}{4}=\frac{1}{2}$$

$\longrightarrow$

W_2 ：{(上,上),(上,下),(下,上)}

A ：{(上,上)}

B ：{(上,下),(下,上)}

$$P(W_2)=\frac{3}{4},P(A)=\frac{1}{4},P(B)=\frac{2}{4}$$

$$P(A|W_2)=\frac{\frac{1}{4}}{\frac{3}{4}}=\frac{1}{3},P(B|W_2)=\frac{\frac{2}{4}}{\frac{3}{4}}=\frac{2}{3}$$

Question

生日問題

一個班級中有n位學生，且班上無雙胞胎的情況下，這個班級中兩個人生日的在同一天的可能性要超過50%，這個班級最少有幾位學生？

A：23位

Answer

假設每個生日的機率均等，且不計閏年。

I. 反轉問題，計算沒有相同生日的機率，先考慮人數少的組計算只有一對人有不同生日的機率。

假設 $f(n)$是n位學生中，每位學生生日皆在不同天的機率

Ex.

$$n=2,\ f(2)=\frac{365}{365}\times\frac{364}{365}\approx 0.997=99.7\%$$

$$n=3,\ f(3)=\frac{365}{365}\times\frac{364}{365}\times\frac{363}{365}\approx 0.992=99.2\%$$

$$\boxed{\Rightarrow f(n)=\frac{P_n^{365}}{365^n}}$$

…

Answer

II. 計算出 $f(n)$ 後，再用 $100\% - f(n)$ 可得至少有兩位同學同一天生日的機率

Ex.

$$n = 2, 100\% - f(2) \approx 100\% - 99.7\% = 0.3\%$$

$$n = 3, 100\% - f(3) \approx 100\% - 99.2\% = 0.8\%$$

$$\cdots$$

$$\boxed{\Rightarrow 100\% - f(n) = 100\% - \frac{P_n^{365}}{365^n}}$$

Answer

III. 在 n＝23 時，我們可得

$$f(23)=\frac{P_{23}^{365}}{365^{23}}\approx 0.4927=49.27\%$$

$$100\%-f(23)=100\%-\frac{P_{23}^{365}}{365^{23}}$$

$$\approx 100\%-49.27\%=50.73\%$$

因此，一個班級中，在無雙胞胎的情況下，兩個人生日的在同一天的可能性要超過50%，這個班級最少有 **23** 位學生。

Question

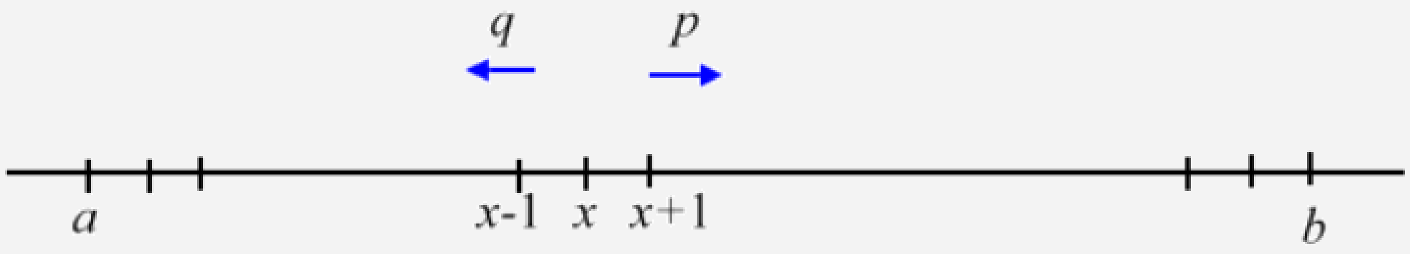
賭博問題

甲乙二人進行一次一元，不輸光不停止的賭博，甲原有 m 元本錢，乙原有 n 元本錢，甲贏一次的機率為 p，乙贏一次的機率為 q，問甲將乙贏光的機率為多少？

A：$\dfrac{1-\left(\frac{q}{p}\right)^{m}}{1-\left(\frac{q}{p}\right)^{m+n}}$

Answer

設 $f(x)$表示目前在 x 點而到達 b 的機率（讓乙輸光光）



已知 $f(a)=0$（甲先輸光，甲贏光乙的機率為0）

$f(b)=1$（乙已經輸光，甲贏光乙的機率為1）

$$\boxed{f(x)=p\cdot f(x+1)+q\cdot f(x-1)}$$

Answer

$$(p+q)\cdot f(x) = p\cdot f(x+1) + q\cdot f(x-1)$$

$$\Rightarrow p\cdot f(x) + q\cdot f(x) = p\cdot f(x+1) + q\cdot f(x-1)$$

$$\Rightarrow p\left[f(x+1) - f(x)\right] = q\left[f(x) - f(x-1)\right]$$

$$\Rightarrow f(x+1) - f(x) = \frac{q}{p}\left[f(x) - f(x-1)\right]$$

因此$f(x) - f(x-1) = \dfrac{q}{p}\left[f(x-1) - f(x-2)\right]$

Answer

$$f(x)-f(x-1)=\frac{q}{p}\left[f(x-1)-f(x-2)\right]=\left(\frac{q}{p}\right)^2\left[f(x-2)-f(x-3)\right]$$

$$=\ldots=\left(\frac{q}{p}\right)^{x-a-1}\left[f(a+1)-f(a)\right]=\left(\frac{q}{p}\right)^{x-a-1}\left[f(a+1)\right]$$

Answer

$$\text{令} k = f(a+1), \text{則} f(x) - f(x-1) = k \cdot \left(\frac{q}{p}\right)^{x-a-1}$$

$$f(x) - f(x-1) = k \cdot \left(\frac{q}{p}\right)^{x-a-1}$$

$$f(x-1) - f(x-2) = k \cdot \left(\frac{q}{p}\right)^{x-a-2}$$

$$\vdots$$

$$f(a+1) - f(a) = k \cdot \left(\frac{q}{p}\right)^{0}$$

將其相加，可得

$$\boxed{f(x) = k + k \cdot \left(\frac{q}{p}\right) + \ldots + k \cdot \left(\frac{q}{p}\right)^{x-a-2} + k \cdot \left(\frac{q}{p}\right)^{x-a-1}}$$

Answer

$$f(x)=k+k\cdot\left(\frac{q}{p}\right)+\ldots+k\cdot\left(\frac{q}{p}\right)^{x-a-2}+k\cdot\left(\frac{q}{p}\right)^{x-a-1}\Rightarrow f(x)=\frac{k\left[1-\left(\frac{q}{p}\right)^{x-a}\right]}{1-\left(\frac{q}{p}\right)},\text{又 } f(b)=1$$

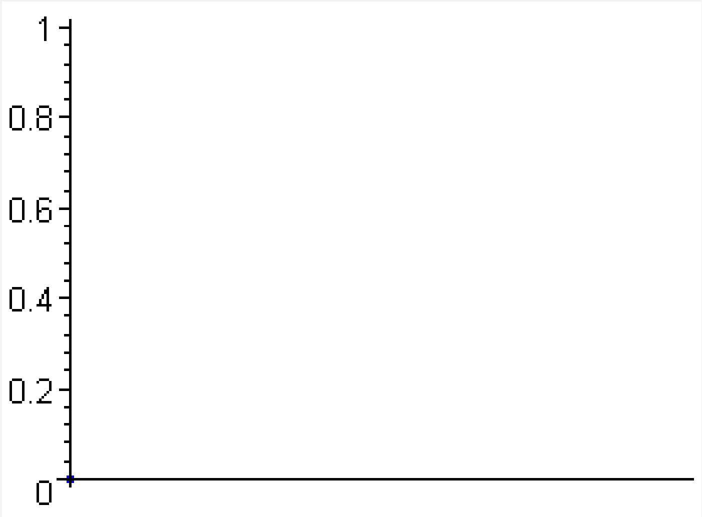
$$\text{故}1=f(b)=\frac{k\left[1-\left(\frac{q}{p}\right)^{b-a}\right]}{1-\left(\frac{q}{p}\right)}\Rightarrow k=\frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^{b-a}}$$

$$\text{因此 }\boxed{f(x)=\frac{1-\left(\frac{q}{p}\right)^{x-a}}{1-\left(\frac{q}{p}\right)^{b-a}}}$$

Answer

$a=0, b=m+n$

故可得 $\boxed{f(m)=\dfrac{1-\left(\dfrac{q}{p}\right)^{m}}{1-\left(\dfrac{q}{p}\right)^{m+n}}}$



02

未來篇—
貝氏定理

給我現在有的數據，我可以用貝氏定理推出你的未來！

貝氏定理

托馬斯·貝葉斯
(Thomas Bayes, 1702-1761)

〈流數學說的介紹，以及對《分析學家》作者的反對提出數學家的防禦〉：**我早就認為流數法的基本原理和規則，需要更為全面且清楚的解釋和證明。**

在機率論上的研究：〈《機率論》中一個問題的解決〉表明他想解決的問題是給定某未知事件(指發生的機率未知)發生與未發生的次數，求在一次試驗中發生的機率值介於兩個指定機率值之間的可能性(機率)。

有歷史學者認為他可能受過棣美弗(Abraham de Moivre, 1667-1754)的指導，而《機率論》是棣美弗有關機率的數學著作。

全機率定理(Law of Total Probability)

定義1.1(分割, partition)

設$A_1, A_2, \ldots, A_n$為樣本空間S中之一組事件，若滿足下面兩個條件:

① 兩兩互斥:$A_i \cap A_j = \emptyset, \forall \mathrm{i} \neq \mathrm{j}$

② 集體包含:$\cup_{i=1}^{n} A_i = S$

定理1.2(全機率定理)

設$A_1, A_2, \ldots, A_n$為樣本空間S中之一組分割，則對於樣本空間S中任意事件B而言，我們有

$\mathrm{P(B)}=\sum_{i=1}^{n} \mathrm{P}(A_i \cap B)=\sum_{i=1}^{n} \mathrm{P}(B|A_i)\mathrm{P}(A_i)$

貝氏定理

定理2.1(貝氏定理)

設$A_1, A_2, \ldots, A_n$為樣本空間S中之一組分割，B是樣本空間S上之任意事件，則

$$\mathrm{P}(A_k|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{B}|A_k)\mathrm{P}(A_k)}{\sum_{i=1}^{n}\mathrm{P}(\mathrm{B}|A_i)\mathrm{P}(A_i)}$$

其中k可以是任意1, 2, ⋯, n

證明:

$$\mathrm{P}(A_k|\mathrm{B}) = \frac{\mathrm{P}(A_k \cap B)}{\mathrm{P}(\mathrm{B})} = \frac{\mathrm{P}(A_k \cap B)}{\sum_{i=1}^{n}\mathrm{P}(A_i \cap B)} = \frac{\mathrm{P}(\mathrm{B}|A_k)\mathrm{P}(A_k)}{\sum_{i=1}^{n}\mathrm{P}(\mathrm{B}|A_i)\mathrm{P}(A_i)}$$

① 可以看出貝式定理的分母就是一個全機率定理。

② 事前機率(priorprobability):$\mathrm{P}(A_1), \mathrm{P}(A_2), \dots, \mathrm{P}(A_n)$
指的是事件B發生前之機率，此即未考慮事件B之下，$A_1, A_2, \dots, A_n$各別事件發生的**非條件機率**。

③ 事後機率(posteriorprobability):$\mathrm{P}(A_1|B), \mathrm{P}(A_2|B), \dots, \mathrm{P}(A_n|B)$
指的是已知事件B發生後之機率，此即給定事件B已發生之下，$A_1, A_2, \dots, A_n$各別事件發生的**條件機率**。

④ 貝氏定理的意義與精神:當得到新的訊息，也就是認知到事件B已經發生的新事實，此時事件$A_1, A_2, \dots, A_n$不宜再用事前機率來評估，我們既知事件B已經發生，就要想辦法重新估算一下，再給定事件B已經發生的前提條件之下，事件$A_1, A_2, \dots, A_n$發生的條件機率，而貝氏定理可以幫我們完成這件事。

Question

某石油公司根據地質資料，評估某地有20%的機率實際蘊藏有石油，又根據經驗，如一地實際蘊藏有石油，則在該地鑽依探勘油井有80%的機率會挖出石油。現在，該公司在已經連續鑽兩座探勘油井挖不到油的情況下，計算該地實際蘊藏有石油的機率為何？

A：$\dfrac{0.2^2 \times 0.2^2}{0.2^2 \times 0.2 + 1 \times 0.8}$

Answer

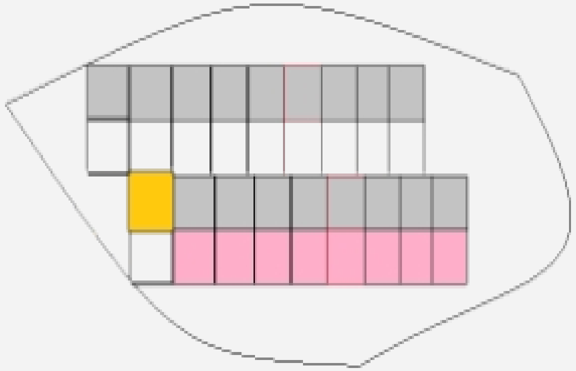
令A事件表示該地蘊藏有石油

令B事件表示連續兩座探勘油井挖不到石油

則$\mathrm{P(A)} = 0.2$, $\mathrm{P}(\mathrm{A}') = 0.8$, $\mathrm{P(B|A)} = 0.2^2$, $\mathrm{P}(\mathrm{B}|\mathrm{A}') = 1^2$

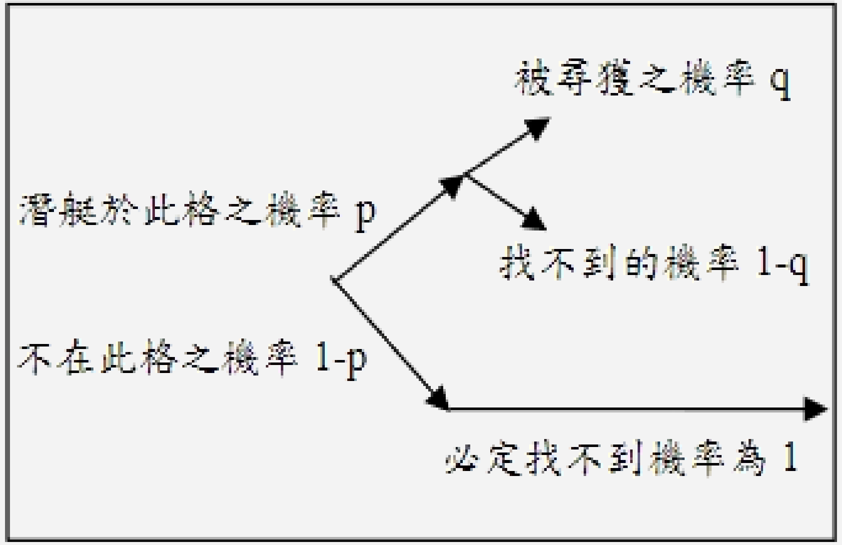
$$\mathrm{P(A|B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(\mathrm{B}|\mathrm{A}')P(\mathrm{A}')} = \frac{0.2^2 \times 0.2^2}{0.2^2 \times 0.2 + 1 \times 0.8}$$

貝氏定理的應用



貝氏搜尋理論(Bayesiansearchtheory)

美國的潛艇在大西洋海域失蹤，根據各類專家的意見，確認出潛艇最可能掉落在某海域，而此海域被細分成許多的小方格，每個方格有兩個重要的機率值，分別為 p 與 q，其中 p 表示潛艇落於此一方格的機率，此稱為事前機率，為未加入新資訊更新前的主觀機率，而若潛艇落於此方格會被找尋到的機率為 q。



Question

即使潛艇真的落在黃色方格，但能找到的機率只有q，所以當黃色方格搜尋後，如果找不到但失落物仍存於此方格的機率為一條件機率為何?

Answer

令A事件為潛艇在黃色方格

令B事件為此方格找不到潛艇，則B'為此方格找不到潛艇

則

$$\mathrm{P}(\mathrm{A}|\mathrm{B}') = \frac{\mathrm{P}(\mathrm{A} \cap \mathrm{B}')}{\mathrm{P}(\mathrm{B}')} = \frac{p(1-q)}{p(1-q)+(1-p)} = \frac{p(1-q)}{1-pq} = \mathrm{p} \times \frac{1-q}{1-pq} < p$$

依此結果可以得知比原先認定的 p 還要小

Question

對其他的方格而言，在未搜尋前，若潛艇落於其他方格的事前機率為1-p，則當黃色的方格搜尋未果的情況下，潛艇落於其他方格的機率應該會提升，至於提升為何？可以考慮，在黃色方格找不到的條件下，潛艇落在其他方格的機率應該會調整為何？

Answer

令A事件為潛艇在其他方格
令B事件為此方格找不到潛艇，則B'為此方格找不到潛艇
則

$$\mathrm{P}(\mathrm{A}|\mathrm{B}') = \frac{\mathrm{P}(\mathrm{A}\cap\mathrm{B}')}{\mathrm{P}(\mathrm{B}')} = \frac{1-p}{p(1-q)+(1-p)} = \frac{1-p}{1-pq} > 1-p$$

依此結果可知，機率比原先認定的1-p還要大。也就是說，當黃色方格搜尋未獲後，其他方格的機率會變動，接著就尋找機率最高的方格，依此原則，直到尋獲為止。

商學上的應用-賽局理論

最基本的賽局是兩方參賽者，依據雙方行動順序可分成，靜態賽局表示參與者同時行動或先後行動但不知道前者採取何種行動；而動態賽局則為參與者有先後順序而且可以觀察到先行動者的行動。另外依據對對方資訊的了解可分成有完全認知的「完全資訊」，否則即為「不完全資訊」。

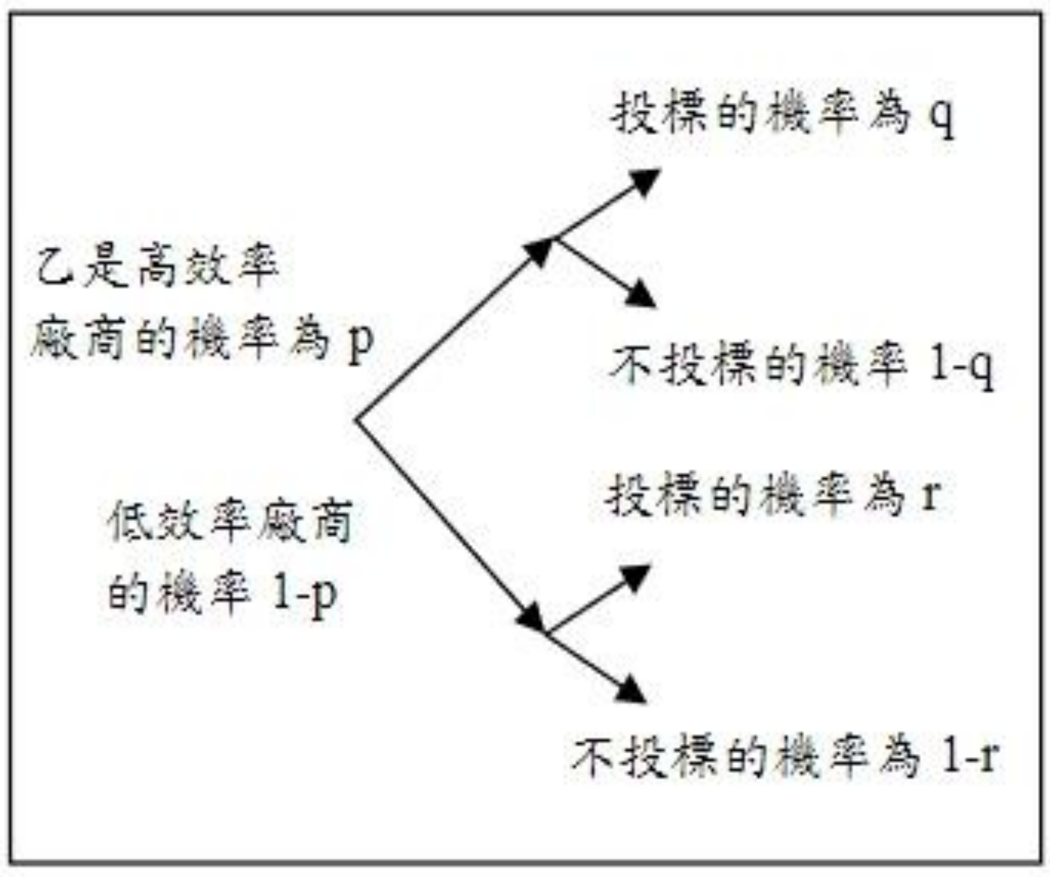
如果是動態不完全資訊的賽局，則可以觀察到對方的行動，當對方行動時，可利用這些訊息，利用貝氏定理進而釐定出策略，而取得優勢。

假設最單純的情況，僅有甲、乙兩家競爭的廠商，共同參與某項投標活動，規定以價格最低者得標。把價格壓低，得標機會大增，但獲利縮水

Question

現在假設，廠商就只有兩種類型，分別為高效率與低效率兩種，甲廠商根據經驗判斷乙為高效率廠商的機率為p，則有1-p的機率為低效率廠商，此即為先事前率。而對類似的競標案，根據過去的經驗，高效率廠商會競標的機率為q，低效率廠商為r，其中（q＞r＞0），樹狀圖應該表示為何？在確知乙參加競標案的條件下，乙會是高效率廠商的機率為何？

Answer


投標的機率為 q
乙是高效率
廠商的機率為 p
不投標的機率 1-q
投標的機率為 r
低效率廠商
的機率 1-p
不投標的機率為 1-r

Answer

令A事件為乙是高效率廠商
令B事件為乙有參加競標

$$\mathrm{P(A|B)}=\frac{\mathrm{P(A\cap B)}}{\mathrm{P(B)}}=\frac{pq}{pq+(1-p)r}>\frac{pq}{pq+(1-p)q}=p$$

依此結果可以得知，機率值比原先認定的p還要大，而此P(A|B)稱為事後機率，為利用資訊後調整的機率。得到事後機率後，甲可以根據此值來擬定投標的金額，增加勝算。

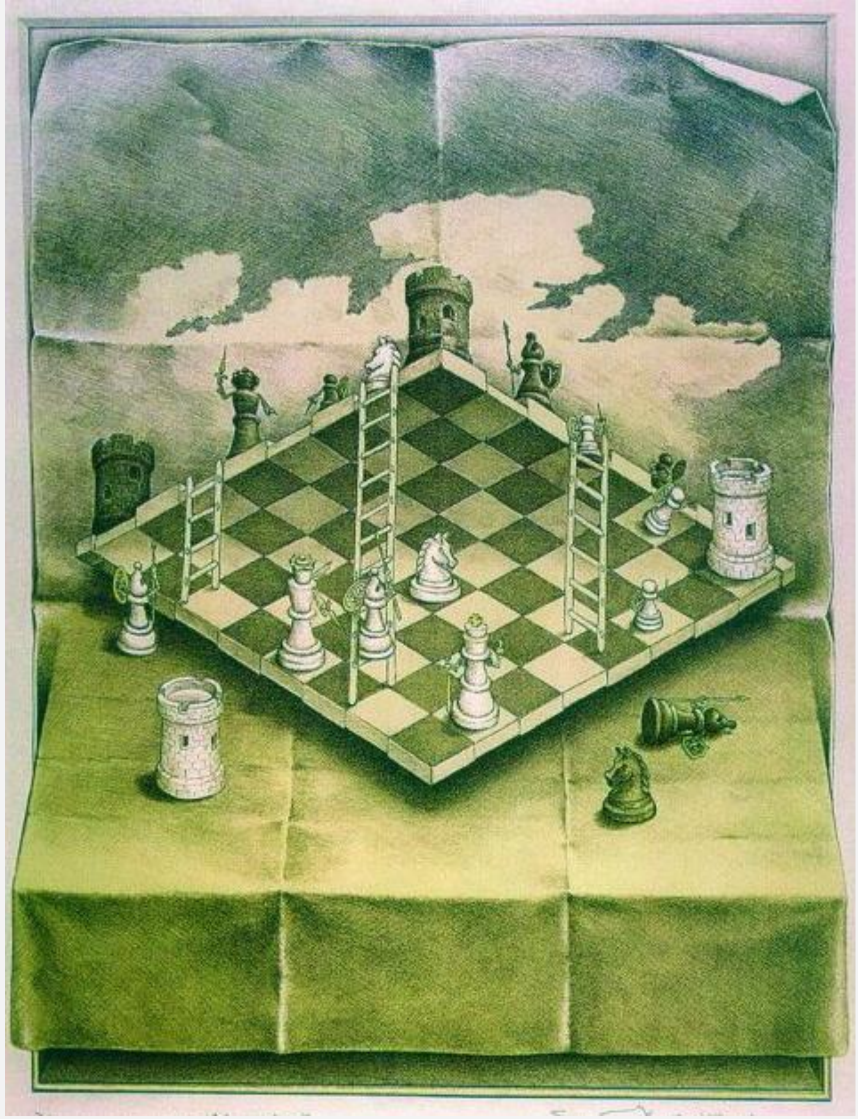
03

愛情篇—
機 率 悖 論

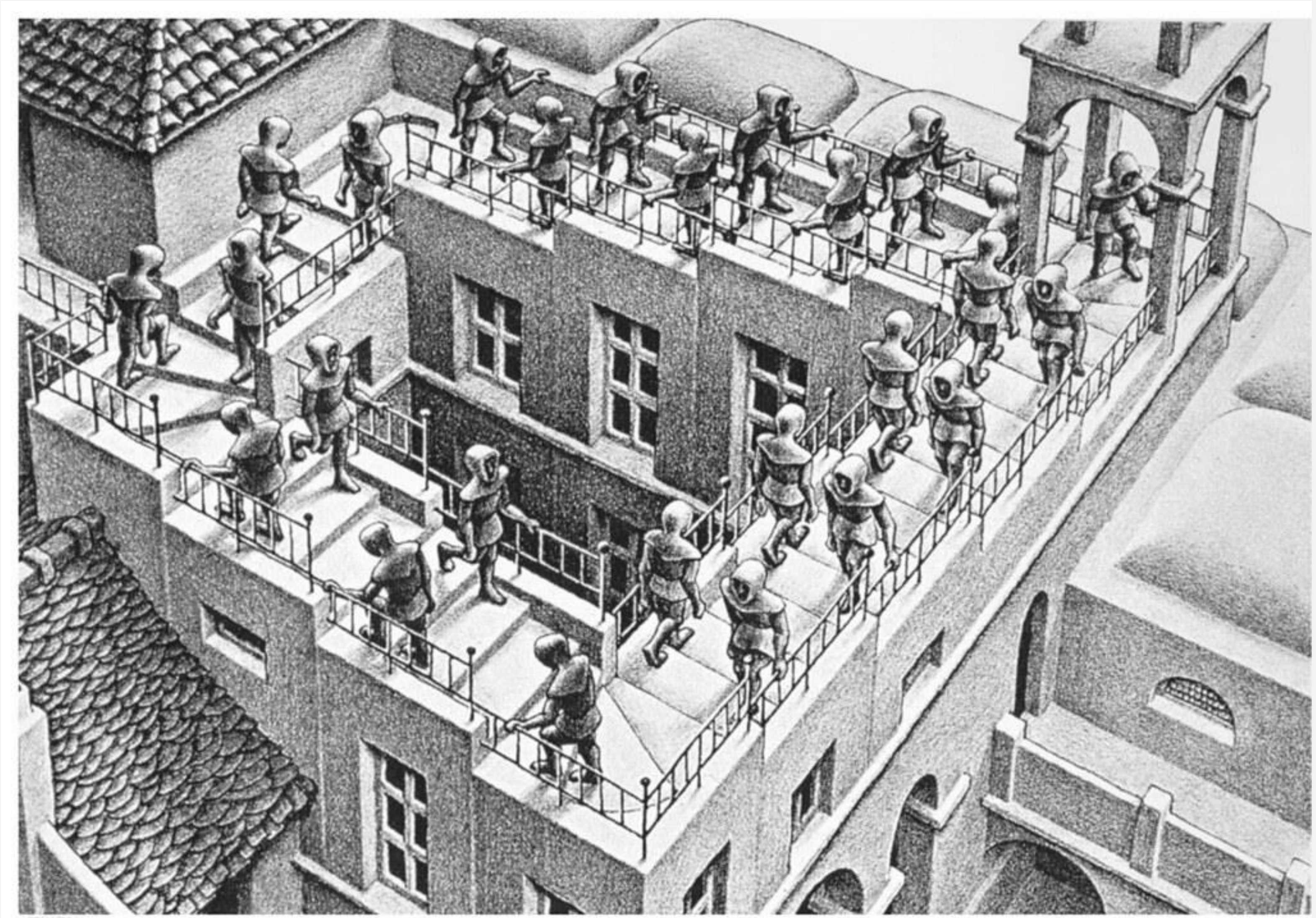
明明只是問法不同，怎麼同一個題目可以給我好幾種答案？如同愛情一般永遠不能用邏輯解答...

何謂悖論（Paradox）？

悖論是指一種導致矛盾的命題。通常從邏輯上無法判斷正確或錯誤稱為悖論，有時違背直覺的正確論斷也稱為悖論。悖論的英文paradox一詞來自希臘語，意思是「未預料到的」、「奇怪的」。 如果承假設它是真的，經過一系列正確的推理，卻又得出它明顯是假的；如果假設它是假的，經過一系列正確的推理，卻又得出它明顯是真的。古今中外有不少著名的悖論，它們震撼了邏輯和數學的基礎，激發了人們求知和精密的思考，吸引了古往今來許多思想家和愛好者的注意力。解決悖論難題需要創造性的思考，悖論的解決又往往可以給人帶來全新的觀念。



何謂悖論



說謊者悖論

(A)這個語句為假

- 如果(A)為真，那麼「這個語句為假」為真 →(A)一定為假。

  從(A)為真的假設推導出(A)為假的結論，矛盾。

- 如果(A)為假，那麼「這個語句為假」為假 →(A)一定為真。

  從(A)為假的假設推導出(A)為真的結論，矛盾。

這句話是假的

理髮師悖論

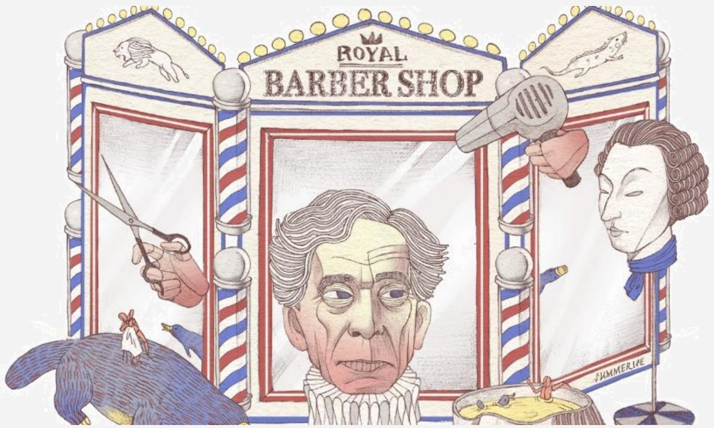
理髮師說：我要為城裡人刮鬍子，而且一定只要為城裡所有「不為自己刮鬍子的人」刮鬍子。
想一想：理髮師該為自己刮鬍子嗎？

- 如果他為自己刮鬍子，按照他說「只為城裡所有不為自己刮鬍子的人刮鬍子」

→ 他不應該為自己刮鬍子，矛盾。

- 如果他不為自己刮鬍子，按照他說「一定要為城裡所有不為自己刮鬍子的人刮鬍子」

→ 他應該為自己刮鬍子，矛盾。



何謂悖論

「悖論」是指一種命題，聲稱某項內容超出甚至反對「通常的見解」。
抛開悖論的各種含義，通常所說的導致矛盾的悖論，應當滿足如下條件：

1. 有一個命題 A，稱為悖論命題。
2. 有一個邏輯系統 L，稱為相關系統。
3. 有一組命題 E，稱為背景命題。背景命題都是相關系統中的真命題。相關系統被簡化為背景命題，背景命題成為悖論證明的依據。
4. 相關系統存在兩個證明可以獲得悖論命題A的真值，其中一個證明 A 為真，而另一個證明 A 為假，從而出現矛盾。

→ 在計算機率時若產生隨機變數的「機制」或「方法」沒有定義嚴謹，機率也將無法得到良好的定義。

1

悖論簡單實例

Ex：彩票悖論

彩票悖論

https://www.bilibili.com/video/BV18h411f7py/

彩票悖論又稱凱伯格悖論，源於小亨利·E·基伯格考慮的是公平的1,000張彩票只有一張中獎彩票。如果對彩票的執行了解這麼多，那麼接受某些彩票會中獎是合理的。

首先根據假設檢驗，如果原假設概率非常小，就可以拒絕原假設。假設0.001就是一個非常小的概率，組織一次公正的1,000張彩票抽獎活動，按照之前的假設，1號彩票中獎的概率是0.001，是要拒絕的，依次類推，我們可以拒絕所有的彩票，那麼就沒有彩票可中獎，但現實情況是總會有中獎的彩票，這是統計和邏輯不相符的一個例子。

不要把統計關係當作因果關係，邏輯和統計本身卻是大不相同。

2

常見迷思和悖論的差別

Ex：三門問題、金幣箱子問題

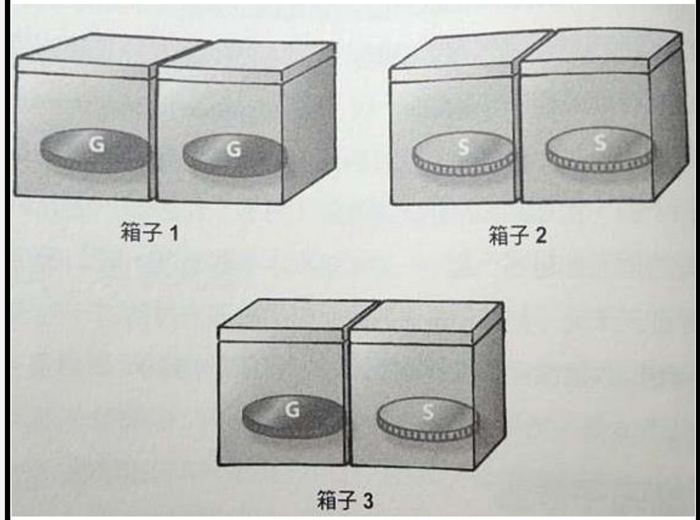
蒙提霍爾悖論

蒙提霍爾問題亦稱為蒙特霍問題、山羊問題或三門問題，是一個源自賽局理論的數學遊戲問題。

參賽者會看見三扇門，其中一扇門的裏面有一輛汽車，選中裏面是汽車的那扇門，就可以贏得該輛汽車，另外兩扇門裏面則都是一隻山羊。當參賽者選定了一扇門，主持人會開啟另一扇是山羊的門；並問：「要不要換一扇門？」

換，當參賽者轉向另一扇門而不是繼續維持原先的選擇時，贏得汽車的機會將會加倍。

伯特蘭箱子悖論



「似是而非的悖論」的第二個例子由十九世紀法國數學家**約瑟夫・伯特蘭**提出。

有三個箱子，每個箱子裡各有兩枚硬幣，放置方式如下：每個箱子都隔成兩半；每一半各放一枚硬幣，而且蓋子可以單獨打開來查看裡頭的硬幣種類（但不允許查看另一枚）。第一個箱子裡放了兩枚金幣（代號GG），第二個箱子裡放了兩枚銀幣（代號SS），第三個箱子則有金幣和銀幣各一枚（代號GS）。請問你選到內有金幣跟銀幣的箱子機率有多少？

正確答案是，**不論是否察看其中一枚硬幣，選到GS箱的機率一直都是三分之一，而非二分之一。**

3

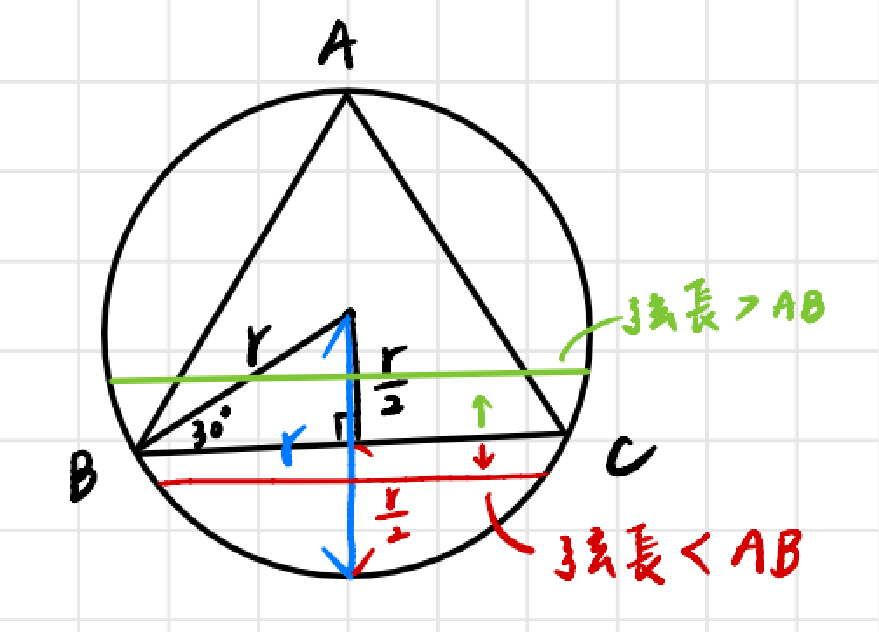
悖論問題

伯特蘭悖論、辛普森悖論、伯克森悖論

伯特蘭悖論（幾何機率悖論）

考慮一個內接於圓的等邊三角形。若隨機選圓上的弦，則此弦的長度比三角形的邊較短的機率為何？

想法一：隨機半徑（弦心距）

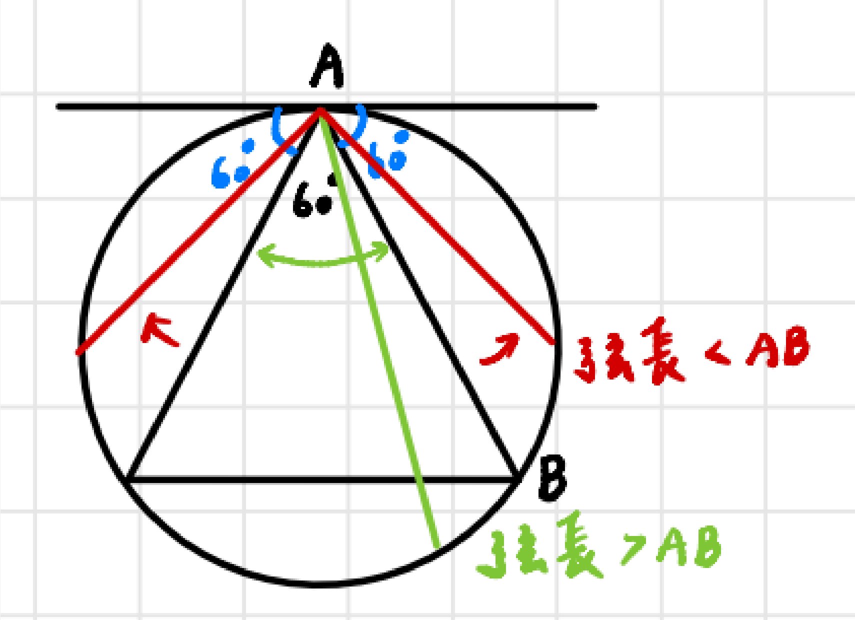


$$P(\text{弦長} < \overline{AB}) = \frac{\frac{r}{2}}{r} = \frac{1}{2}$$

伯特蘭悖論（幾何機率悖論）

考慮一個內接於圓的等邊三角形。若隨機選圓上的弦，則此弦的長度比三角形的邊較短的機率為何？

想法二：隨機端點（弦切角）



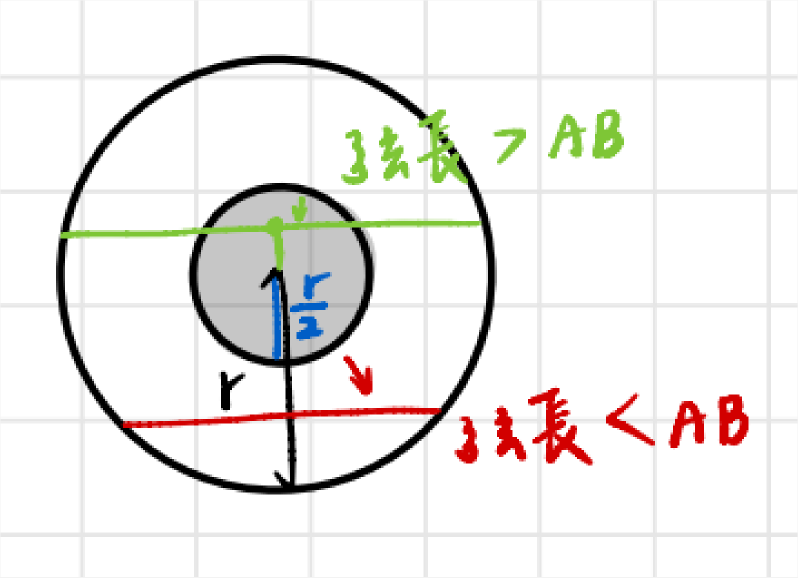
$$P(\text{弦長} < \overline{AB}) = \frac{120°}{180°} = \frac{2}{3}$$

伯特蘭悖論（幾何機率悖論）

考慮一個內接於圓的等邊三角形。若隨機選圓上的弦，則此弦的長度比三角形的邊較短的機率為何？

想法三：隨機中點（面積）

Idea : 有切到$\frac{r}{2}$為半徑之圓內部之弦弦長 $> \overline{AB} \rightarrow$ 用面積算機率

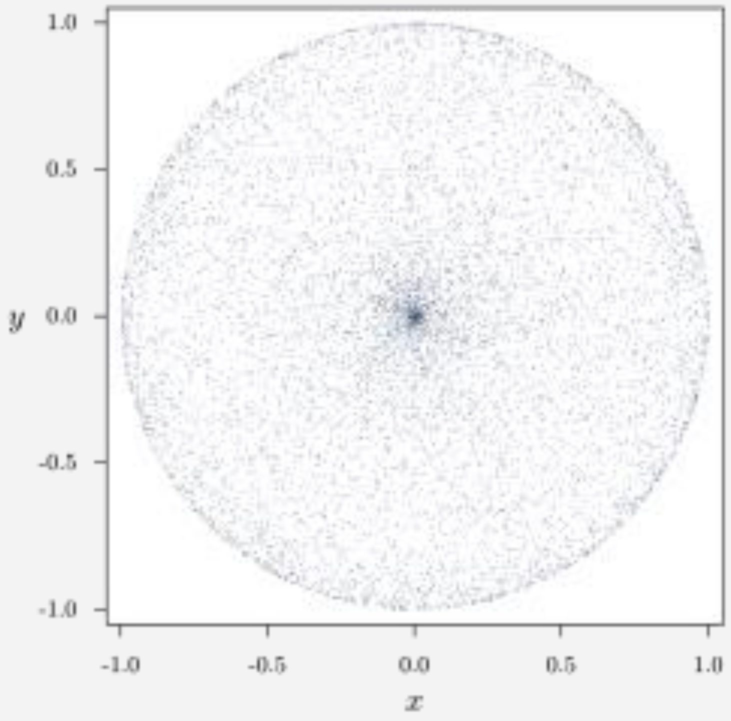


$$P(\text{弦長} > \overline{AB}) = \frac{\pi(\frac{r}{2})^2}{\pi r^2} = \frac{1}{4}$$

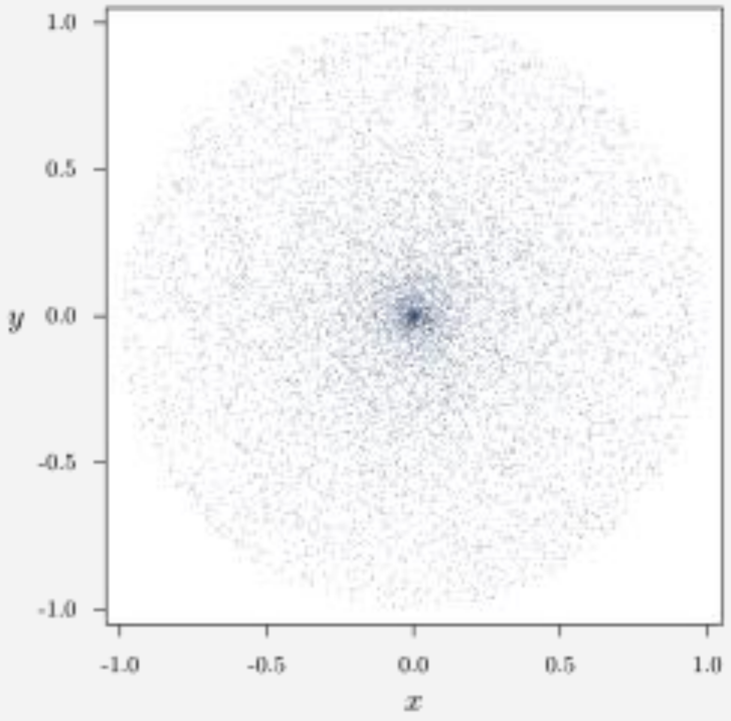
$$\Rightarrow P(\text{弦長} < \overline{AB}) = 1 - \frac{1}{4} = \frac{3}{4}$$

每一個弦都可以被其中點唯一決定

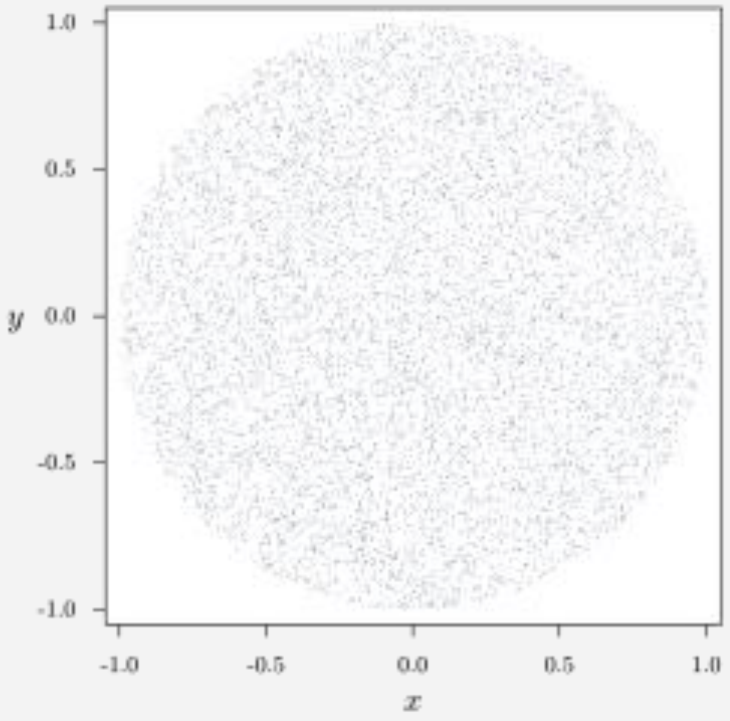
上述三種方法會給出不同中點的分佈



隨機的弦的中點
（方法一：隨機半徑）



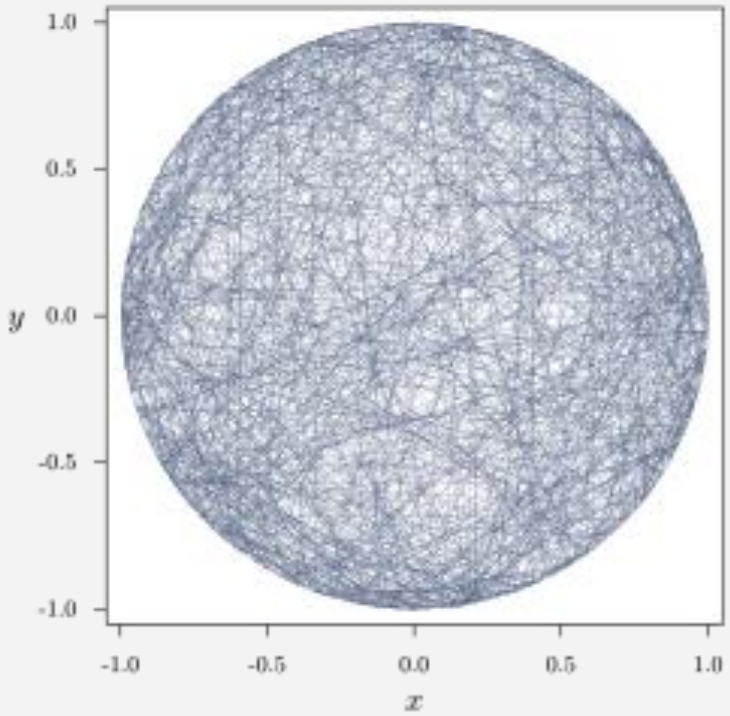
隨機的弦的中點
（方法二：隨機端點）



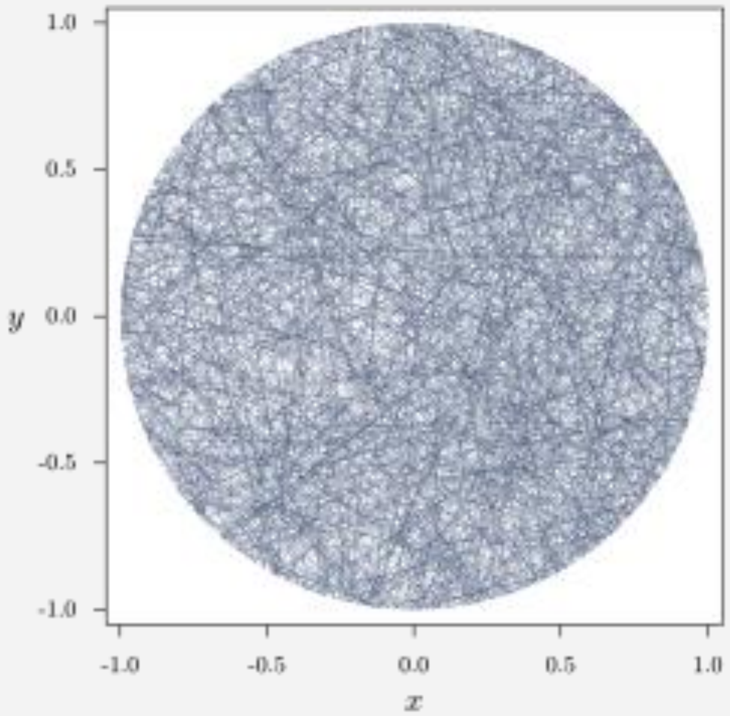
隨機的弦的中點
（方法三：隨機中點）

直接看弦的分佈

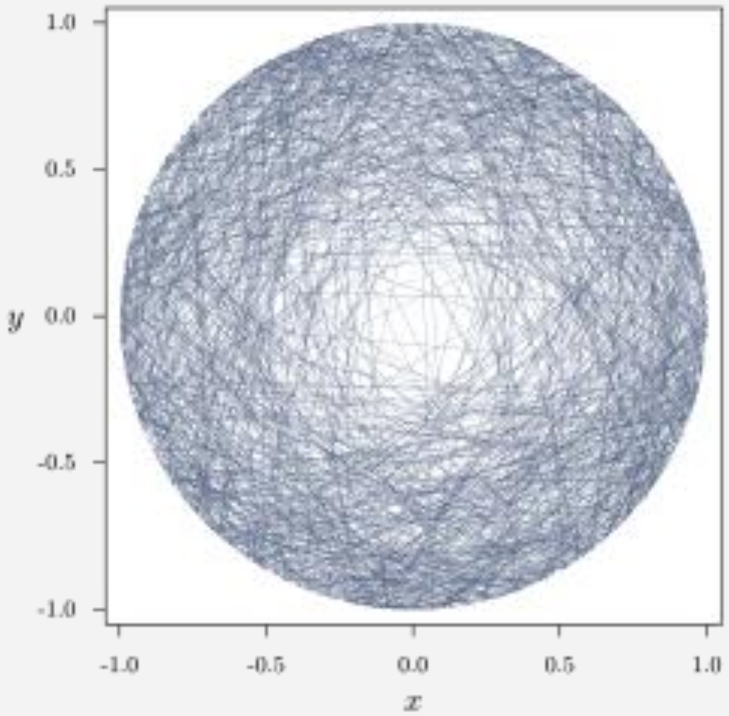
方法2的弦看起來比較均勻，方法1、方法3的弦則較不均勻。



隨機的弦
（方法一：隨機半徑）



隨機的弦
（方法二：隨機端點）

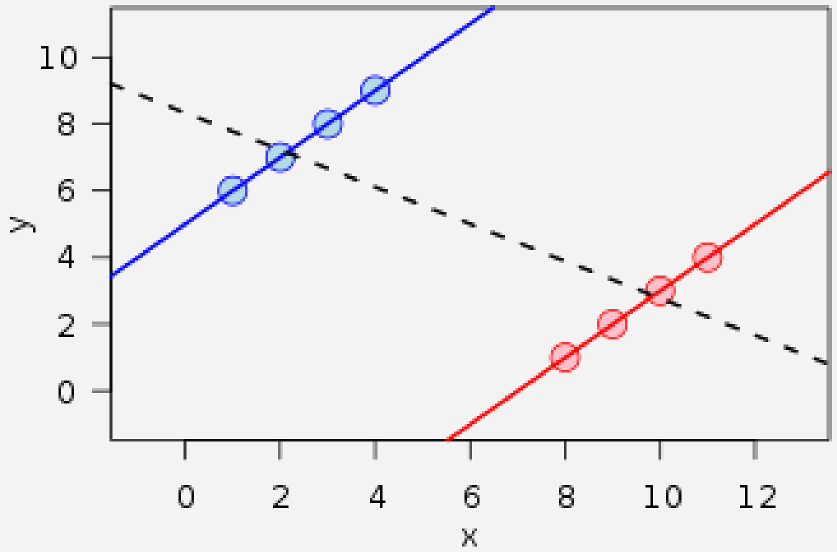


隨機的弦
（方法三：隨機中點）

關鍵在於「隨機」選擇弦的方法。若選定了隨機選擇的方法，問題自然也就會有良好定義的解答。既然不存在一個唯一的選擇方法，那麼也就不存在一個唯一的解答。這三種解答分別對應不同的選擇方法，若沒有更進一步的資訊，也沒有理由認為其中的一個解答會比另一個解答更好。

辛普森悖論（Simpson's paradox）

辛普森悖論是概率和統計中的一種現象，其中趨勢出現在幾組數據中，但當這些組被合併後趨勢消失或反轉。這個結果在社會科學和醫學科學統計中經常遇到，當頻率數據被不恰當地給出因果解釋時尤其成問題。當干擾變量和因果關係在統計建模中得到適當處理時，這個悖論就可以得到解決。辛普森悖論已被用來說明統計誤用可能產生的誤導性結果。



定量數據的辛普森悖論：兩個獨立的小組出現正的趨勢，而當小組合併時出現負的趨勢

辛普森悖論

一所美國高校的兩個學院，分別是法學院和商學院。
新學期招生，人們懷疑這兩個學院有性別歧視。

性別	錄取	拒收	總數	錄取比例
男生	8	45	53	15.1%
女生	51	101	152	33.6%
合計	59	146	205	

性別	錄取	拒收	總數	錄取比例
男生	201	50	251	80.1%
女生	92	9	101	91.1%
合計	293	59	352	

根據兩個表格來看，女生在兩個學院都被優先錄取。

→ 女生的錄取比率較高

辛普森悖論

一所美國高校的兩個學院，分別是法學院和商學院。
新學期招生，人們懷疑這兩個學院有性別歧視。

性別	錄取	拒收	總數	錄取比例
男生	209	95	304	68.8%
女生	143	110	253	56.5%
合計	352	205	557	

將兩學院的數據匯總 → 女生的錄取比率反而比男生低

這個例子說明，簡單的將分組數據相加匯總，並不能反映真實情況

辛普森悖論 分析

1. 兩個分組的錄取率相差很大，從表格可看出法學院錄取率很低，而商學院卻很高。而同時兩種性別的申請者分布比重相反。女性申請者的大部分分布在法學院，相反，男性申請者大部分分布於商學院。結果在數量上來說，拒收率高的法學院拒收了很多的女生，男生雖然有更高的拒收率，但被拒收的數量卻相對不算多。而錄取率很高的商學院錄取了很多男生，使得最後加總的時候，男生在數量上反而占優。
2. 有潛在因素影響著錄取情況。比如說，性別並非是錄取率高低的唯一因素，甚至可能是毫無影響的。至於在學院中出現的比率差，可能是隨機事件。又或者是其他因素作用（如入學成績），卻剛好出現這種錄取比例，使人誤認為這是由性別差異而造成的。

為了避免辛普森悖論的出現，需斟酌各分組的權重並乘以一定的係數，消除以分組數據基數差異而造成的影響。同時必需清楚了解情況，以綜合考慮是否存在造成此悖論的潛在因素。

伯克森悖論（Berkson's paradox）

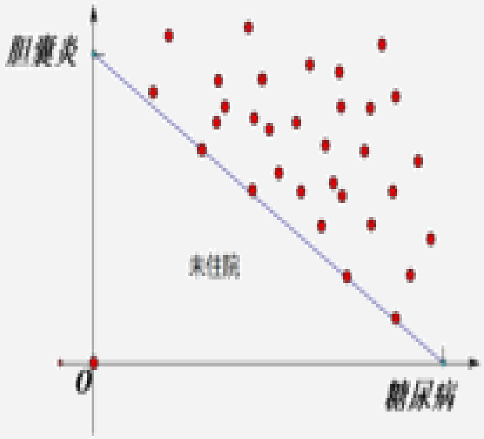
指人們的直覺觀察與實際上真實的條件概率和嚴謹的統計結果不相符，也就是說人們所發現的**看似兩個相關的因素實際上根本無關**。當人們**統計的樣本存在偏差**時，就會出現這種情況由美國醫生和統計學家約瑟夫·伯克森提出。

伯克森悖論（Berkson‘s paradox）

他研究了一個醫院中患有糖尿病的病人和患有膽囊炎的病人，結果發現患有糖尿病的人群中，同時患膽囊炎人數較少；而沒有糖尿病的人群中，患膽囊炎的人數比例較高。這似乎說明患有糖尿病可以保護病人不受到膽囊炎的折磨，但是從醫學上講無法證明糖尿病能對膽囊炎起到任何保護作用。

假設這個醫院只治療兩種疾病：糖尿病和膽囊炎。畫一個平面直角座標系：橫座標表示他患有糖尿病的嚴重程度，縱軸表示患有膽囊炎的嚴重程度，再把每一個人按照兩種疾病的輕重畫在座標系中。

如果對全體人員進行統計，就會發現糖尿病和膽囊炎並沒有相關性。但是如果只對醫院中的患者進行統計，就會出問題。如果病人的糖尿病或者膽囊炎問題比較輕，病人就不需要住院，所以不會被統計到。來到醫院的病人要麼是糖尿病，要麼是膽囊炎，要麼二者兼有。所以，我們需要把圖像左下方的點都去掉，他們不在我們統計的範圍內，我們只會統計到這條線右上方的點。**故發現糖尿病和膽囊炎呈現出負相關了**(如右圖)。不患有糖尿病的人，更有可能患有膽囊炎，而患有糖尿病的人，膽囊炎的比例就會下降了。



伯克森悖論

帥哥都是渣男？

學霸總不學習？

寒門出貴子?

基於對受到限制樣本的觀察，常常成為日常生活和學術研究中獲得錯誤結論的根源

參考資料

生活中的機率問題

黃文璋（2010）。 機率應用不易 。 數學傳播， 34卷 1期， p14-28。
https://web.math.sinica.edu.tw/math_media/d341/34102.pdf
看TED學數學 -從機率的角度看生活中的神奇巧合！ —授課橋 。
https://teach-range.com/%E5%A4%9A%E5%85%83%E6%95%99%E6%A1%88/390
機率中賭博輸光問題分析—中華科技大學 。
http://www.cust.edu.tw/mathmet/graph/randomwalk.htm

貝氏定理

蘇俊鴻（2014）。 貝葉斯和貝氏定理 (1)（Thomas Bayes and Bayes’ Theorem(1)）
科學 Online高瞻自然科學教學資源平台
https://highscope.ch.ntu.edu.tw/wordpress/?p=58765
蘇俊鴻（2014）。貝葉斯和貝氏定理 (2) (Thomas Bayes and Bayes Theorem(2))
科學 Online高瞻自然科學教學資源平台
https://highscope.ch.ntu.edu.tw/wordpress/?p=58766
陳昱成（2013）。 貝氏定理的應用 。 科學教育月刊 ，第 357期 。
https://www.sec.ntnu.edu.tw/uploads/asset/data/62564060381784d09345bb7f/02-101022-(%E7%9F%A5%E8%AD%98)%E8%B2%9D%E6%B0%8F%E5%AE%9A%E7%90%86%E7%9A%84%E6%87%89%E7%94%A8%EF%BC%88%E6%9C%88%E5%88%8A%EF%BC%89.pdf

機率悖論

悖論－維基百科 。資料檢索日期： 2022年 12月 8日 。
https://zh.wikipedia.org/zh-tw/%E6%82%96%E8%AE%BA
說謊者悖論－維基百科 。資料檢索日期： 2022年 12月 8日 。
https://zh.wikipedia.org/zh-tw/%E8%AA%AA%E8%AC%8A%E8%80%85%E6%82%96%E8%AB%96
理髮師悖論－維基百科 。資料檢索日期： 2022年 12月 8日 。
https://zh.wikipedia.org/zh-tw/%E7%90%86%E5%8F%91%E5%B8%88%E6%82%96%E8%AE%BA
考你的眼力能看清楚這些神奇悖論圖嗎？－隨意窩。
https://blog.xuite.net/lee54816/01/14246629
數據分析不踩雷必讀：讓人懷疑人生的七大悖論－IT 邦幫忙。
https://ithelp.ithome.com.tw/articles/10230554
蒙提霍爾問題(三門問題)—維基百科。資料檢索日期：2022 年12 月8 日。
https://zh.m.wikipedia.org/zh-tw/%E8%92%99%E6%8F%90%E9%9C%8D%E7%88%BE%E5%95%8F%E9%A1%8C
伯特蘭箱子悖論，悖論：破解科學史上最複雜的9 大謎團－泛科學。
https://pansci.asia/archives/38866
伯特蘭悖論－維基百科。資料檢索日期：2022 年12 月8 日。
https://zh.wikipedia.org/wiki/%E4%BC%AF%E7%89%B9%E8%98%AD%E6%82%96%E8%AB%96
辛普森悖論－維基百科。資料檢索日期：2022 年12 月8 日。
https://zh.wikipedia.org/zh-tw/%E8%BE%9B%E6%99%AE%E6%A3%AE%E6%82%96%E8%AE%BA
伯克森悖論－百度百科。資料檢索日期：2022 年12 月8 日。
https://baike.baidu.hk/item/%E4%BC%AF%E5%85%8B%E6%A3%AE%E6%82%96%E8%AB%96/19132222

謝謝大家的觀賞

Thank you!