學術小品

大家都知道大學有「大學五件事」：讀書、上莊、住 hall、兼職、出 pool。這五件事會與我們所學的知識息息相關嗎？

<u>讀書篇：Buffon’s Needle Problem</u>

在許多基礎概率學（例如 STAT2901/STAT2601）的材料裡面提到著名的 Buffon’s Needle Problem，其問題如下：

「一個無限大平面上有許多相距為 D 的平行直線，當一根長度為 L（L<D）的針自由跌落時，求這根針穿過任意一根線的概率。」

Notes 上提到，在針的中點離任何一根線的最短長度（設為 X）和針的偏轉角度（設為θ）皆服從均勻分佈：

$$X \sim U\left(0, \frac{D}{2}\right), \qquad \Theta \sim U(0, \pi)$$

以上兩個隨機變數 (random variable) 的聯合概率密度函數 (joint probability density function) 為

$$f_{X,\Theta}(x,\theta) = \begin{cases} \dfrac{2}{\pi D}, & \text{if } 0 < x < \dfrac{D}{2}, 0 < \theta < \pi \\ 0, & \text{otherwise} \end{cases}$$

所求的概率可由雙重積分 (double integration) 計出：

$$\Pr(\text{穿過一根線}) = \Pr\left(X < \frac{L}{2}\sin\theta\right)$$

$$= \int_0^{\pi} \int_0^{\frac{L}{2}\sin\theta} \frac{2}{\pi D}\, dx\, d\theta$$

$$= \frac{2L}{\pi D}$$

認真溫書的 B 同學當然不會放過任何一個細枝末節。他想：「倘若 D 是一定長度的，但是針是無限長的，這時用 2L/πD 算出來的概率遠遠大於 1，是不可能

成立的。因此，關於不同的 L、D 關係，一定有多於 1 種情況。」

現在請你幫忙解決 B 同學的問題，到底 L、D 呈現什麼關係時就不能用 2L/πD 的結論了？在這種情況下又應該如何運算這個概率呢？

思路：我們應該考慮到以上 2L/πD 的推論有一個前提，即是對 X 的積分上限為$(L\sin\theta)/2$。這是由於針的長度L足夠短，$\frac{L}{2} < \frac{D}{2}$（即 $L < D$），不需要考慮有其他上限的情況。實際上，如果針足夠長，當$D/2 < (L\sin\theta)/2$時，X 的範圍應該受到 D / 2 的限制。積分公式應改為

$$\Pr(\text{穿過一根線}) = \int_0^{\frac{\pi}{2}}\int_0^{m(\theta)} \frac{4}{\pi D}\,dx\,d\theta$$

$$\text{其中}m(\theta) = \min\left\{\frac{L\sin\theta}{2}, \frac{D}{2}\right\}$$

本問題在本學會的 2018 年 Newsletter@STAT I 也出現過，詳情可見：
https://drive.google.com/drive/folders/13vq5Fxo1FrwpY3Z18ip99iP8DeRhyUMe

上莊篇：Bertrand’s Ballot

「贊成 14 票，反對 11 票，棄權 0 票，共計 25 票。由於贊成票數多於或等於反對的票數，恭喜 X 同學成功上莊！」

而在一旁的計票員則在思考一個問題：選票是他一張接著一張地打開並統計公布的，並沒有按照什麼特定的順序。「現在已知贊成多於反對的票數，但是中間票數統計的順序與過程卻是未知的，**其實在已打開點算的票當中贊成的票數是不是一直都領先反對的票數呢？**」

推廣來說，他寫下：「如果贊成票為 A，反對票為 B，票數的編排則為 A 與 B 的任意序列，已知贊成 p 票，反對 q 票（p>q），試問在點票過程中任何時候贊成票數都多於反對的票數的概率是多少？」

(如：AAABABBABABAB 則符合上面的情況而 AAABABBABBABB 則不符合)

思路：可以從簡單的情況入手：我們先假設事先知道贊成票（A）為 4 票，反對票（B）為 1 票:

順序	贊成票（A）一路領先嗎?
AAAAB	是
AAABA	是
AABAA	是
ABAAA	否
BAAAA	否

P(A 一路領先)=3/5

當贊成與反對的票數比例改變時，P 也會跟著變化。

你能發現什麼規律嗎？

詳情可見：

http://episte.math.ntu.edu.tw/articles/mm/mm_04_4_03/page2.html

兼職篇：Secretary Problem

在某公司人事部內，人力資源主管正在為如何選拔來兼職的大學生發愁：

「我們公司這次有 27 位面試者，不過只有一個工作名額，那當然要選最優秀的大學生啊。但是公司又規定，每面試完一個大學生，就要立刻告訴他的去留，不能等全部人都面試完畢之後才統一公布結果。這樣就算我可以當場給出每位面試者的面試分數，卻不能知道他是不是最優秀的那一個，我該怎麼辦啊！」

路過的統計部同事剛好聽到了主管的抱怨。他思考了一會兒，留下了一張紙條：

「根據最優停止理論（optimal stopping theory）我們可以借助 1/e 幫忙選擇面試者。27 * 1/e 大概是 10。因此應該從第 11 個人開始，先拒絕頭 10 為面試者，一旦有比前面的面試者都優秀的學生就立即錄取他吧！」

看到這個 1/e，主管既是納悶又是驚喜：到底是不是這麼神奇呢？

思路：這個問題的最優解 (optimal solution)是找到一個最優的停止規則。面試官會拒絕頭 r - 1 個應聘者（假設他們中的最佳人選為應聘者 M），然後選出第一個比 M 好的應聘者。可見最優策略包含於這個系列的策略中。（如果 M 在所有 n 個應聘者中也是最好的一個，那麼這個策略將選不出任何人選）對於任意的截斷值 r，最佳人選被選中的概率是：

$$
\begin{aligned}
P(r) &= \sum_{i=1}^{n} P(\text{applicant } i \text{ is selected} | \text{applicant } i \text{ is the best}) \times P(\text{applicant } i \text{ is the best}) \\
&= \left(\sum_{i=1}^{r-1} 0 \times \frac{1}{n}\right) + \left(\sum_{i=r}^{n} P\left(\begin{array}{l}\text{the best of the first } i-1 \text{ applicants} \\ \text{is in the first } r-1 \text{ applicants}\end{array} \middle| \text{applicant } i \text{ is the best}\right) \times \frac{1}{n}\right) \\
&= \sum_{i=r}^{n} \frac{r-1}{i-1} \times \frac{1}{n} \quad = \quad \frac{r-1}{n} \sum_{i=r}^{n} \frac{1}{i-1}.
\end{aligned}
$$

詳情可見：

https://en.wikipedia.org/wiki/Secretary_problem

住 hall 篇：Airplane Seating Problem

現在有 100 位準大一新生參加 hall 的迎新活動。每人住一間房間，各有自己的房間號，新生輪流入住。

可惜第一位新生粗心大意，房卡不見了。情急之下，他隨便選了一間房間入住。之後的新生採取以下的入住方法：

1. 敲自己房間的門
2. a) 倘若沒人入住，入住
   b) 倘若有人已經入住了，隨便選一個房間入住

於是有一些新生入住了自己的房間，有一些新生沒有，至於誰入住了誰的房間好像完全是隨機的。那麼問題來了，第 100 位新生有多大的概率入住他自己的房間？是 1/100？還是 0？又或者是……1/2？

思路：我們可以從兩個房間的情況入手：假設第一個人走對房間，第二個人也可以走對房間；如果第一個人走錯，第二個人也會走錯。因此第二個人走對房間的概率將會是 50%。

當總共有三個房間的時候，如果第一個人走對房間，即轉化為兩個房間的情況；那麼當第一個人走錯房間，是否也能轉化為兩個房間的情況？倘若房間數繼續增加，還能不能通過這種方法推斷呢？

本題源自飛機選座位問題，詳情可見
https://www.teamten.com/lawrence/puzzles/airplane_seating.html

出 pool 篇：Stable Marriage Problem

本學術小品之出 pool 篇源自 Stable Marriage Problem，那麼理論和現實之間會有什麼差距呢？

Stable Marriage Problem 描述的情景如下：

若男生 A、B、C 與女生 D、E、F 要配成三對，他們各自對於三位異性有自己的排名，如下圖所示：

	A	B	C
D	(3，1)	(2，3)	(1，2)
E	(1，2)	(2，1)	(3，1)
F	(1，3)	(3，2)	(2，3)

例如 B 與 D 交界的（2，3）表示 D 把 B 排在第二位，B 把 D 排在第三位。

請問如何配對才能使三對情侶的關係更加穩定（使每個人對自己情侶心中的排名都儘量高）？

Deferred-Acceptance Algorithm 闡明了一種非常好的配對方法：

第一步：男生向最鍾意的女生表白，女生選擇最喜歡的那個，拒絕其他人的要求。

第二步：所有仍單身的男生從沒有拒絕他的女生中選擇自己最喜歡的那位。所有女生從表白者與現在暫時的男朋友中選擇一個她最喜歡的，拒絕其他人。

這兩步一直循環到所有人都有配對的人為止。

根據 Deferred-Acceptance Algorithm 配對的過程應該如下：

第一輪：

A 向 D 表白；B、C 向 E 表白

D 接受 A；E 接受 B 拒絕 C

第二輪：（C、F 沒有情侶）

C 向 D 表白

D 接受 C 拒絕 A

第三輪：（A、F 沒有情侶）

A 向 E 表白

E 接受 A 拒絕 B

第四輪：（B、F 沒有情侶）

B 向 F 表白

F 接受 B

到此所有人都有情侶了，A↔️E ； B↔️F ； C↔️D，配對結束。

這個會不會是最優的方法？你有其他的想法嗎？

此外，點此可見本學會過去的有趣的學術小品。

https://drive.google.com/file/d/1FmNc68u-bDXnxjiFlk1qDRu8bLuRSoXV/view?usp=sharing