機率應用不易

黃文璋

國立高雄大學應用數學系、統計學研究所

1　前言

統計裡常在做預測。但如量子論泰斗, 曾獲1922年諾貝爾物理獎, 丹麥的**波耳**(Niels Bohr, 1885-1962)所說:

> 預測很難, 尤其關於未來。

世上多的是事後諸葛, 而對於隨機現象做預測時, 誤差常難以避免。只是誤差之意義, 並不易為一般人所理解。統計學家對未來的預測, 因此備受挑戰。有人甚至因此不相信統計, 以為統計不過是謊言。

其實不要說隨機或誤差, 甚至連最原始的機率之意義, 都非三言兩語可說清楚。即使對統計學研究所的學生, 就算學過各種對機率的解釋, 如以相同可能性、頻率、主觀, 及公理化(機率空間)等, 以及各種較深入的機率理論, 常常也會算錯一些, 表面上看起來很簡單的機率。這其中特別是**條件機率**(conditional probability), 是一般人較不易掌握的。可以這麼說:

> 機率很難, 尤其條件機率。

機率值會變, 是機率的特性。假設生男生女之機率各為1/2。有人按

你家門鈴。此人是男是女? 如果沒有其他資訊, 你會想機率大約各$1/2$。但如果你知道按門鈴者, 是送披薩的, 那很可能會認為, 至少有0.9的機率是男生。因根據你的經驗, 送披薩的通常是男生。

這就是條件機率! 即在給定"某事件發生"之條件下(有新的資訊), 原先那一事件發生的機率, 有時會隨之而變。條件機率會不會有不變的時候? 也是有的, 若兩事件**獨立**(independent), 則給定其中之一發生, 對另一事件發生之機率, 便不會有影響。即此資訊對預測原事件發生之機率, 並沒有幫助。**獨立性**可說是機率論中一特別的概念。譬如說, 假設你的心情不受洋基隊輸贏之影響, 則若洋基隊今天贏球, 你投擲一銅板出現正面之機率, 是不會改變的。當然也不會影響你所修的那門機率論期中考及格的機率。但若銅板乃來自洋基隊, 他們若贏球, 給你A銅板; 若輸球, 給你B銅板, 則投擲銅板出現正面之機率, 將隨洋基隊輸贏而有所改變。

曾看過一篇名為"機率與文字陷阱"的文章。該文先舉下述例子。

例1. (i)有一好友有二小孩, 已知老大是女孩。試問老二亦是女孩之機率為何?

(ii) 有一好友有二小孩, 已知有一個女孩。試問另一小孩亦是女孩之機率為何?

該文於算出(i)答案是$1/2$, 及(ii)答案為$1/3$, 並做了一些說明後, 卻自己覺得有些問題。遂舉某高中的一道數學競試題目來討論。我們列為底下例2。

例2. 甲投擲兩硬幣, 並讓乙猜朝上的兩面是"相同"或"相異"。乙正準備要猜, 丙從旁經過, 說"有一個正面"。試問這時乙該猜相同或相異?

該文說出題高中所給之解為: 應猜相異, 猜對機率為$2/3$。算法與例1中的(ii)一樣。然後該文指出:

> 但再仔細想想, 今天如果丙看到的是反面, 那麼乙也要猜相異, 而且猜對的機率也是$2/3$。所以乙只要知道丙有說話, 儘管

不知道丙說什麼, 猜相異對的機率就是2/3 。那其實乙根本不需要丙幫忙, 只要他猜的時候, 假想有一個丙走過來跟他說話, 那猜相異對的機率就比較大(因為不管丙說什麼都是要猜相異) 。於是得到結論: 投擲兩硬幣, 朝上兩面相異之機率為2/3, 因為一定會有正面或反面 。講到這裡, 很明顯有錯, 因為相同及相異之機率, 從國中以來, 就教皆為1/2 。

此質疑看起來還頗有道理的 。你現在相信條件機率不容易了吧! 經過一番討論, 且對例2中的情況找人做了200次實驗, 該文宣佈例1中的(i)與(ii), 及例2, 其中的機率皆為1/2 。該文接著又給一例 。

例3. 所有有兩個小孩且有女孩的家庭中, 兩小孩皆為女孩的機率為何? 該文說解為: 1/3 。

最後該文給出下述結論:

如果題目內有知道的"知", 或是有第三者當仲介給提示或條件, 條件機率做出來都會錯 。反之, 如果題目有強調"所有的"(如例3), 那麼每一個情況發生的機率都相同, 就可以放心的用條件機率 。

搞了半天, 是題目敘述有語病 。這是文字陷阱 。這題一開始的問法, 應該是上題這種問法才對, 只是不小心敘述錯誤, 造成大問題 。我想, 大家往後解題, 應該會多注意這種情況了 。

在上面這段該文作者有意思的心得中, 我們猜想"這題"乃指例1之(ii), 而"上題"指例3 。該文作者寫作的動機, 是為了釐清一些常會引起學生困擾的條件機率問題 。可惜他的文章引發的問題, 恐多於解決的問題 。我們稍後會回到他所提出的幾個例子 。

諸位看, 有時給的條件是一段文字, 如何將這段文字的內涵正確解讀, 並不見得都很容易 。若解讀錯誤, 得到的條件機率當然也就不對了 。鑒於

條件機率處處可見, 但其概念, 卻又不易為一般人所了解 。本文將對此略做討論 。

2 條件機率

曾看到下述一則新聞報導:

> 美國加州有一家庭, 爸爸媽媽和剛出生的小孩, 都在同月同日生 。$\cdots$ 。這樣巧合的機率只有0.00000751 。$\cdots$ 。夫妻兩人當初就是因生日相同, 相信緣分天註定而結婚, 沒想到第一個小孩也在同一天出生 。

上述機率是如何求出呢? 假設1年有365天, 則任意3人生日相同之機率為

$$\binom{365}{1}\frac{1}{365^3} = \frac{1}{365^2} = 7.506 \cdot 10^{-6} \text{ 。}$$

這當然與"某家族中有3人生日相同"之機率不同, 也與"某學校中有3位學生生日相同"的機率不同 。但對這對自認緣分天註定的夫妻而言, 他們的第一個小孩生日與他們同一天之機率, 卻為1/365, 並不眞那麼小 。他們夫妻生日相同, 是一既成的事實, 可視為一給定的條件 。在此條件下, 要求第3人(他們的第一個小孩)生日與他們同一天的機率 。這與任挑選的3人, 生日同一天, 情況不同 。

提醒初學者:求機率時, 務必要弄清楚究竟在求什麼事件之機率 。在給不同的條件下, 機率值可能會因此不同 。

我們在樣本空間上定義機率 。有時得到一些資訊, 則根據所獲得的資訊, 樣本空間可能有所改變, 因而機率空間也就隨之而變 。得到的新機率, 就是所謂**條件機率** 。

在數學裡不會有這種情況 。給定某數是2, 它就一直是2 。不變是數學的特性 。但在討論機率時, 某事件的機率, 是有可能因情況不同而變 。這本來是不奇怪的, 但因大部分的人受數學的薰陶較久, 習慣數學中

處理"不變"的問題, 所以在學習機率時, 看到機率值居然會改變, 有時便不易理解。

定義1. 設A, B為樣本空間Ω中二事件, 且$P(B) > 0$。則在給定B發生下, A發生之**條件機率**, 以$P(A|B)$表之, 定義為

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{。} \tag{1}$$

在條件機率的定義中, B成為新的樣本空間: $P(B|B) = 1$。也就是原先的樣本空間Ω修正為B。所有事件發生之機率, 都要先將其針對與B的關係做修正。舉幾個特例來看。假設$P(B) > 0$。若A與B為**互斥**(disjoint)事件(即$A \cap B = \emptyset$), 則知道B發生, A必不發生, 所以$P(A|B)$應為0。因$P(A \cap B) = 0$, 故(1)式的確給出$P(A|B) = 0$; 若$P(A)$亦為正, 則此時亦有$P(B|A) = 0$。另外, 若$B \subset A$, 因$P(A \cap B) = P(B)$, 故$P(A|B) = 1$。這當然是正確的。因$B \subset A$, 故若知道B發生, 則A就一定發生。最後, 若$A \subset B$, 則因$P(A \cap B) = P(A)$, 故$P(A|B) = P(A)/P(B)$。這當然也是對的。因在給定B之下, B成為新的樣本空間, 而A包含於B, A發生的可能性, 就是A在B中所佔的"分量", 即$P(A)/P(B)$。

先給二例。

例4. 玩梭哈時, 要拿到4條很不容易。52張撲克牌, 隨機地發5張, 其中有4張點數相同之機率為

$$\frac{13 \cdot 48}{\binom{52}{5}} = \frac{624}{2{,}598{,}960} \doteq 0.00024,$$

的確很小。但若發了3張牌, 皆拿到K, 則此時會拿到4條之機率為何?

解. 令B表已發的3張牌皆為K的事件, A表拿到4條的事件。則

$$A \cap B = \{\text{首3張皆為}K, \text{第4、5張中有1張}K, \text{1張非}K\},$$

且

$$P(A \cap B) = \frac{\binom{4}{3}}{\binom{52}{3}} \cdot \frac{48}{\binom{49}{2}}, \ P(B) = \frac{\binom{4}{3}}{\binom{52}{3}} \text{。}$$

因此

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{48}{\binom{49}{2}} = \frac{2}{49} \text{ 。}$$

機率顯然提高很多。

例5. 投擲一公正的銅板兩次。求兩次投擲皆得正面之機率, 給定(i)第1次得到正面, (ii)兩次投擲至少有1正面。

解.首先樣本空間

$$\Omega = \{(\text{正},\text{正}), (\text{正},\text{反}), (\text{反},\text{正}), (\text{反},\text{反})\},$$

且

$$P(\omega) = 1/4, \forall\, \omega \in \Omega \text{ 。}$$

令

$$\begin{aligned}
A &= \{(\text{正},\text{正})\},\\
B &= \{(\text{正},\text{正}), (\text{正},\text{反})\},\\
C &= \{(\text{正},\text{正}), (\text{正},\text{反}), (\text{反},\text{正})\} \text{ 。}
\end{aligned}$$

本例即求條件機率(i)$P(A|B)$, 及(ii)$P(A|C)$。因

$$A \cap B = A \cap C = \{(\text{正},\text{正})\},$$

故

$$P(A \cap B) = P(A \cap C) = 1/4 \text{ 。}$$

又

$$P(B) = 2/4,\ P(C) = 3/4 \text{ 。}$$

故

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{1/4}{2/4}=\frac{1}{2},$$
$$P(A|C)=\frac{P(A\cap C)}{P(C)}=\frac{1/4}{3/4}=\frac{1}{3}\text{。}$$

在分別給定第1次得到正面, 及兩次投擲至少有1正面之條件下, 兩次皆得正面之機率, 分別是$1/2$及$1/3$。很多初學者對第二個條件機率不是$1/2$而是$1/3$常感到困惑。他們認為在給定兩次投擲至少有1正面下, 導致有兩個可能的情況:

(i)兩次皆為正面, 及(ii)1正面1反面。

因此所求之機率應為$1/2$。他們誤以為相同可能性到處適用。殊不知原先樣本空間Ω中的4個元素, 的確有相同的可能性; 一旦給定至少有1正面, 等價於告知兩次投擲的結果不可能是(反,反), 因此只剩3個相同可能性的結果(正,正), (正,反), (反,正)。而其中只有1個結果是兩次皆為正面。故所求之機率為$1/3$。

在機率中處處可見條件機率。一方面是的確常會遇到求在給定某條件下之機率; 另一方面, 某些機率值, 雖原先並非以條件機率的形式出現, 有時卻可經由條件機率求得。底下陸續會說明。

由(1)式得

(2) $$P(A\cap B)=P(A|B)P(B)\text{。}$$

故若知道$P(A|B)$及$P(B)$, 則可得到$P(A\cap B)$。當然只要$P(A)>0$, 便亦有

(3) $$P(A\cap B)=P(B|A)P(A)\text{。}$$

結合(2)與(3)式, 得

(4) $$P(A|B)=\frac{P(B|A)P(A)}{P(B)}\text{。}$$

此後若不特別聲明, 當提到上式, 就隱含假設$P(A)$及$P(B)$皆為正。

(4)式為底下**貝氏定理**(Bayes' rule)之一特例, 這是英國牧師**貝氏**(Thomas Bayes, 1702-1761)首先提出而得名。不過也有人認為法國的大數學家**拉普拉斯**(Pierre-Simon Laplace, 1749-1827), 才是第一位明確給出此定理者, 所以應稱為**拉普拉斯公式**(Laplace's formula)。

在給下定理前, 我們先介紹**分割**(partition)。對一樣本空間Ω, 事件$A_1, A_2, \cdots$, 若滿足**兩兩互斥**, 即$A_i \cap A_j = \emptyset$, $\forall i \neq j$, 且$\bigcup_{i=1}^{\infty} A_i = \Omega$, 便稱為$\Omega$之一**分割**。當然也可以有有限的分割$A_1, \cdots, A_n$, $n \geq 2$。

定理1. 設$A_1, A_2, \cdots$為樣本空間之一分割。則對任一事件B,

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)。\tag{5}$$

在(5)式中, 若有某一$P(A_i) = 0$, 則雖此時$P(B|A_i)$沒定義, 但只要將$P(B|A_i)P(A_i)$定義為0, 則(5)式仍成立。

定理2. 貝氏定理. 設$A_1, A_2, \cdots$為樣本空間之一分割。則對任意$i \geq 1$, 及事件B, 只要$P(B) > 0$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}。\tag{6}$$

例6. 有甲、乙、丙三囚犯, 國王宣佈以抽籤決定釋放其中一位, 處決另兩位。他告訴獄卒那一位將被釋放, 但要求獄卒不可先透露。甲於要求獄卒透露那一位會被釋放遭到拒絕後, 改問獄卒"乙及丙中, 那一位會被處決?"獄卒經過一番思考, 遂(誠實地)告訴甲, "乙會遭處決"。他認為這樣做並未違反國王的規定, 原因為:

> 乙、丙二人, 至少有一會遭處決, 這是大家都知道的, 因此他並未提供甲任何有關甲是否會被釋放的有用資訊。

甲聽到獄卒說乙會被處決後很高興。原先他有1/3的機率遭釋放, 現因只剩他與丙了, 所以他會被釋放的機率提高至1/2。

究竟獄卒與甲的分析, 何者正確?

解. 令A, B, C分別表甲 、乙 、丙三人會被釋放的事件 。如果我們考慮的結果是誰會被釋放, 則樣本空間$\Omega = A \cup B \cup C$ 。由假設

$$P(A) = P(B) = P(C) = 1/3 \text{ 。}$$

令K表獄卒說"乙會被處決"的事件 。必須要了解, 若乙 、丙皆會被處決, 獄卒其實是自乙 、丙中, 任挑一位(即各1/2的機率, 我們隱含做了此假設)告訴甲誰會被處決; 若乙將被釋放, 獄卒只能告訴甲, "丙會被處決"; 若丙將被釋放, 獄卒只能告訴甲, "乙會被處決" 。我們想求$P(A|K)$ 。

首先由定理1:

$$\begin{aligned} P(K) &= P(K|A)P(A) + P(K|B)P(B) + P(K|C)P(C) \\ &= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2} \text{ 。} \end{aligned}$$

因此

$$\begin{aligned} P(A|K) &= \frac{P(\text{獄卒說乙會被處決, 且甲被釋放})}{P(K)} \\ &= \frac{1/3 \cdot 1/2}{1/2} = \frac{1}{3} \text{ 。} \end{aligned}$$

換句話說, 在獄卒告訴甲, "乙會被處決後", 甲被釋放的機率(即$P(A|K)$)仍維持為$1/3$ 。此一資訊, 對甲可說是沒用的 。

讀者可能會好奇, 那獄卒所提供的資訊是否便毫無用處呢? 那倒未必 。若丙偷聽到獄卒與甲的對話, 則便知他被釋放的機率(即$P(C|K)$)提高至$2/3$ 。而若乙偷聽到獄卒與甲的對話, 則便知沒有活命的機會了(即$P(B|K) = 0$) 。這樣說好了, 乙 、丙二人中, 有一人被釋放之機率為$2/3$ 。若給定乙被處決, 則丙便獨自擁有全部被釋放之機率, 即$2/3$ 。至於甲, 被釋放之機率並未改變, 還是$1/3$ 。而三人被釋放之條件機率和,

$$P(A|K) + P(B|K) + P(C|K) = 1/3 + 0 + 2/3,$$

仍是1 。

最後, K的機率為$1/2$, 直觀上是對的, 這點讓各位自行想一想。

上例再度顯示, "相同可能性"並非到處適用。又對條件機率, 必須要謹慎處理, 否則極易犯錯。上例最早是Tierney(1991)所提出,有時以不同的型式出現。如著名的**汽車與山羊問題**(Car-Goat Problem), 即為一例。此問題亦曾出現在2008年一部很賣座的電影**決勝21點**(21)中。有三扇門, 其中有一扇門後有汽車, 另兩扇門後各只有一頭山羊。能得到汽車當然是比較好的。你選定一扇門後, 主持人打開另兩扇門中的一扇, 發現門後是山羊, 問你是不是要更改選擇。如上例中的討論, 若主持人事先知道汽車在那一扇門後, 則換是較好的選擇; 但若主持人事先不知汽車在那一扇門後, 則可能打開一扇門後有汽車, 此時遊戲結束; 而若打開的那扇門後是山羊, 則換或不換, 會得到汽車的可能性相同, 即機率皆為$1/2$。

例7. 衛生局至高雄大學免費檢驗某疾病。假設檢驗的結果有正、負兩種反應。如果呈正反應, 便表示可能有病, 須至醫院做進一步檢驗; 如果呈負反應, 則衛生局便認為沒有問題。衛生局宣稱檢驗之可靠度為90%, 且平均每5,000人中, 有一人患此病。基於上述資訊, 你是否願意接受此檢驗?

解.題意顯示, 檢驗並非百分之百可靠, 但醫學上通常也沒有完全精確的檢驗。可靠度90%的意義為, 若無病, 檢驗會呈負反應之機率為0.9; 若有病, 則檢驗會呈正反應之機率亦為0.9。但我們該知道的, 其實是當檢驗呈正反應下, 的確有病的機率, 及當檢驗呈負反應下, 的確無病之機率。

以"正"表檢驗呈正反應, "負"表檢驗呈負反應。則

$$\begin{aligned} P(\text{有病}|\text{正}) &= \frac{P(\text{正}|\text{有病})P(\text{有病})}{P(\text{正}|\text{有病})P(\text{有病})+P(\text{正}|\text{無病})P(\text{無病})} \\ &= \frac{0.9\cdot\frac{1}{5{,}000}}{0.9\cdot\frac{1}{5{,}000}+0.1\cdot\frac{4{,}999}{5{,}000}} \\ &= \frac{9}{5{,}008} \doteq 0.001797\text{。} \end{aligned}$$

即當檢驗呈正反應, 會有病的機率, 才約0.001797, 不到$1/500$, 與所謂90%可靠度實在差太遠。看到此結果, 你可能不太想接受檢驗了, 否則一旦呈

正反應, 要到醫院受罪。有趣的是, 當檢驗呈負反應下, 的確無病的機率倒是很接近1:

$$P(\text{無病}|\text{負}) = \frac{44{,}991}{44{,}992} \doteq 0.999977773 \text{。}$$

難道檢驗只對呈負反應可靠? 似乎不該如此。那原因何在?

直觀上看, 由於檢驗有10%的錯誤沒病卻呈正反應, 而在每5,000人中, 有病的很少(平均才1人), 因此在5,000人中, 約有500個正反應, 但其中才約1人有病。$1/500 = 0.002$, 與所求出的0.001797就接近了。那檢驗不就沒什麼用? 也不盡然, 本來任何1人有病的機率為$1/5{,}000 = 0.0002$, 一旦檢驗呈正反應, 有病的機率升為0.001797, 約成為8.985倍, 增加了不少。至於任何一個人被認為沒病之機率原先為$4{,}999/5{,}000 = 0.9998$, 本來就很接近1, 一旦檢驗呈負反應, 只是略微升高而已。

在患病比率為1/5,000, 於不同檢驗可靠度下, 表1給出當檢驗呈正反應時, 有病之機率。可看出即使檢驗可靠度高達99.99%, 當檢驗呈正反應, 有病之機率也才約2/3。主要是此為罕見疾病之故。切記條件機率$P(\text{正}|\text{有病})$與$P(\text{有病}|\text{正})$是完全不同的。因此千萬不要被那些宣稱高可靠度的檢驗所誤導。特別是對罕見疾病, 更要注意條件機率。本例也顯示, 何以醫學上奇蹟偶而總會出現。有些被醫生判定無救者, 最後卻安然出院。看到這種現象, 讀者應同意, 醫生也該學點機率, 尤其是條件機率。

檢驗可靠度	90%	95%	99%	99.9%	99.99%
$P(\text{有病}\|\text{正})$	0.001797	0.003786	0.019419	0.166556	0.666689

表1. 患病比率1/5,000, 於不同檢驗可靠度下之$P(\text{有病}|\text{正})$

最後, 當檢驗可靠度為90%, 於不同患病比率下, 我們亦給出$P(\text{有病}|\text{正})$於表2。此表顯示, 患病比率愈高, $P(\text{有病}|\text{正})$也隨之提高。當患病比率達到平均每2人中有1人, 此機率為0.9, 與檢驗可靠度90%相同; 當患病比率

高達平均每5人中有4人, 此機率則高達$36/37 \doteq 0.97297$ 。

患病比率	1/5,000	1/5,00	1/50	1/5	1/2	4/5
P(有病|正)	9/5,008	9/508	9/58	9/13	9/10	36/37

表2. 檢驗可靠度90%, 於不同患病比率下之P(有病|正)

下例將有助於釐清前言中那3個例子 。

例8. 有一對夫妻剛搬進某社區, 大家只知他們有兩個小孩, 並不知性別 。某日社區一管理員, 見到此家之媽媽帶著家中一小孩在玩耍 。若該小孩是女孩, 求此家兩小孩皆為女孩之機率 。

解.先定義下述事件:

G_1 : 老大是女孩,

G_2 : 老二是女孩,

G :媽媽帶著的小孩是女孩 。

次將女孩改為男孩, 類似地定義B_1及B_2 。本例即要求$P(G_1 \cap G_2|G)$ 。依定義

$$P(G_1 \cap G_2|G) = \frac{P(G_1 \cap G_2 \cap G)}{P(G)} = \frac{P(G_1 \cap G_2)}{P(G)}, \tag{7}$$

此處用到明顯的事實$G_1 \cap G_2 \subset G$ 。利用定理1, 得

$$\begin{aligned} P(G) &= P(G|G_1 \cap G_2)P(G_1 \cap G_2) + P(G|G_1 \cap B_2)P(G_1 \cap B_2) \\ &\quad + P(G|B_1 \cap G_2)P(B_1 \cap G_2) + P(G|B_1 \cap B_2)P(B_1 \cap B_2) \\ &= \frac{1}{4} + \frac{1}{4}P(G|G_1 \cap B_2) + \frac{1}{4}P(G|B_1 \cap G_2) 。\end{aligned} \tag{8}$$

在上式中用到

$$P(G|G_1 \cap G_2) = 1,\ P(G|B_1 \cap B_2) = 0,$$

且

$$P(G_1 \cap G_2) = P(G_1 \cap B_2) = P(B_1 \cap G_2) = P(B_1 \cap B_2) = \frac{1}{4},$$

這是因假設生男生女的機率均為$1/2$。將(7)式分母的$P(G)$以(8)代入, 得

$$
(9)\quad P(G_1 \cap G_2|G) = \frac{1/4}{1/4 + P(G|G_1 \cap B_2)/4 + P(G|B_1 \cap G_2)/4} = \frac{1}{1 + P(G|G_1 \cap B_2) + P(G|B_1 \cap G_2)} 。
$$

故欲求之$P(G_1 \cap G_2|G)$為何, 與$P(G|G_1 \cap B_2)$及$P(G|B_1 \cap G_2)$有關。

底下先看幾個特別的情況。

情況(i)。假設不論兩小孩之性別為何, 若只帶一小孩出門, 媽媽帶老大出門之機率為一定值p(因此帶出門的是老二的機率為$1-p$)。實際上媽媽帶老大出門的機率, 可與兩小孩之性別有關, 我們稍後將討論。即假設

$$
P(G|G_1 \cap B_2) = p,\ P(G|B_1 \cap G_2) = 1 - p 。
$$

代入(9)式, 得

$$
P(G_1 \cap G_2|G) = \frac{1}{1 + p + (1-p)} = \frac{1}{2} 。
$$

即原問題之答案為$1/2$, 與p無關。

情況(ii)。假設當兩小孩之性別不同, 則媽媽帶女兒出門之機率為一定值q, 不論她是老大或老二。即假設

$$
P(G|G_1 \cap B_2) = P(G|B_1 \cap G_2) = q 。
$$

代入(9)式, 得

$$
(10)\qquad P(G_1 \cap G_2|G) = \frac{1}{1 + q + q} = \frac{1}{1 + 2q} 。
$$

因此, 這時兩小孩皆為女孩之條件機率與q有關。例如, 設$q = 1$, 即若有兒子及女兒, 媽媽一定帶女兒出門, 則$P(G_1 \cap G_2|G) = 1/3$。此其實即為例5的(ii)。因“看到該家媽媽帶女兒出門”, 等價於“該家至少有一女兒”。次設$q = 1/2$, 即若有兒子及女兒, 媽媽會帶女兒或兒子出門之機率各半, 則$P(G_1 \cap G_2|G) = 1/2$。又當$q = 0$時, $P(G_1 \cap G_2|G) = 1$。注意在(10)式中之機率$P(G_1 \cap G_2|G)$, 為一q之漸減函數, 且

$$
1/3 \le P(G_1 \cap G_2|G) \le 1 。
$$

我們發現, 除非有其他資訊, 否則看到該家媽媽帶著一個女孩, 並不能解讀為, 此資訊等價於"該家至少有一女兒"。這可能與不少人想的不同。又情況(i)及情況(ii), 並非樣本空間之一分割, 甚至兩情況也不互斥。

由上討論知, 依題意所給條件, 本問題並無法解出。雖然我們已較原題意, 多做了一個生男生女機會相等的假設。但並不夠, 必須要有額外的假設, 否則無法給出原問題的解。為什麼會這樣呢?

我們在求機率時, 常不太在意機率空間。大部分的時候也相安無事, 能得到正確的答案。但有時遇到較細膩的情況, 就得將機率空間弄清楚。事實上, 在本問題裡, 樣本空間Ω中有8個元素, 包含所有型如(s_1, s_2, i)的樣本, 其中s_1為老大之性別, s_2為老二之性別, 而i為所見到媽媽帶著的小孩之排序(老大或老二)。欲知$\forall \omega \in \Omega$的機率, 光給$(s_1, s_2)$之機率不夠, 還須做額外的假設。譬如說, 須知"給定媽媽帶著的小孩之性別下, 該小孩之排序"的條件機率。

在生男生女的機率均為1/2的假設下, Ω中8個元素的機率為:

$$
\begin{aligned}
P(\{(\text{女},\text{女},\mathrm{I})\}) = \frac{p_1}{4},\ P(\{(\text{女},\text{女},\mathrm{II})\}) = \frac{1-p_1}{4},\\
P(\{(\text{女},\text{男},\mathrm{I})\}) = \frac{p_2}{4},\ P(\{(\text{女},\text{男},\mathrm{II})\}) = \frac{1-p_2}{4},\\
P(\{(\text{男},\text{女},\mathrm{I})\}) = \frac{p_3}{4},\ P(\{(\text{男},\text{女},\mathrm{II})\}) = \frac{1-p_3}{4},\\
P(\{(\text{男},\text{男},\mathrm{I})\}) = \frac{p_4}{4},\ P(\{(\text{男},\text{男},\mathrm{II})\}) = \frac{1-p_4}{4}\text{。}
\end{aligned}
$$

其中I, II分別表媽媽帶著的小孩為老大或老二之事件。附帶一提, p_1即為給定兩小孩皆為女孩$(G_1 \cap G_2)$之下, I發生(媽媽帶著的小孩為老大)之機率, p_2, p_3及p_4的意義可依此類推。因

$$G \cap G_1 \cap B_2 = \{(\text{女},\text{男},\mathrm{I})\}, G \cap B_1 \cap G_2 = \{(\text{男},\text{女},\mathrm{II})\},$$

故

$$P(G|G_1 \cap B_2) = \frac{P(\{(\text{女},\text{男},\mathrm{I})\})}{P(\{(\text{女},\text{男})\})} = \frac{p_2/4}{1/4} = p_2,$$

且

$$P(G|B_1\cap G_2)=\frac{P(\{(\text{男},\text{女},\mathrm{II})\})}{P(\{(\text{男},\text{女})\})}=\frac{(1-p_3)/4}{1/4}=1-p_3\text{ 。}$$

將上二式代入(9)式, 得

$$P(G_1\cap G_2|G)=\frac{1}{1+p_2+1-p_3}=\frac{1}{2+p_2-p_3}\text{ 。}\tag{11}$$

即要知道p_2-p_3之值, 才能得到所欲求之機率$P(G_1\cap G_2|G)$ 。特別地, 當$p_2=p_3=p$, 則

$$P(G_1\cap G_2|G)=\frac{1}{2},$$

與之前的情況(i)所得吻合; 當$p_2=q$, $p_3=1-q$, 則

$$P(G_1\cap G_2|G)=\frac{1}{1+2q},\tag{12}$$

與之前情況(ii)所得吻合 。即前述情況(i)及(ii), 有p有q, 看起來似乎包含很多可能, 其實均只為本一般情況之特例 。

在上例中, 尚可有其他情境 。如

(i)管理員問那位媽媽"你有沒有女兒? "媽媽答"有";

(ii)管理員問那位媽媽"你老大是女兒嗎? "媽媽答"是";

(iii)管理員見到那位媽媽帶兩個小孩及一條狗在玩耍, 其中有一女兒站著, 另一小孩跪在地下, 但被狗遮住, 看不出性別 。

各位可分別對此三情況, 求此家兩小孩皆為女孩之機率 。

最後回到一開始那三個例子 。

對於例1, (i)之答案為1/2, 應很容易理解 。至於(ii), 若無其他資訊, 則假設"有一個女孩", 等價於"家中至少有一女孩", 仿例5之(ii), 可得另一小孩亦是女孩之機率, 的確為1/3 。

對於例2, 若假設當丙看到兩硬幣有1正面及1反面朝上, 便各有1/2的機率說"有一個正面"及"有一個反面", 則此對應例8中的情況(ii), 且$q=1/2$,

此時朝上的兩面相同及相異的機率皆為$1/2$。也就是在此情況下, 丙所提供的資訊是沒用的。但若$q \neq 1/2$, 則丙所提供的資訊就會有用了。這時可類似如例8的討論。

最後對於例3, 仍對應例5之(ii), 兩小孩皆為女孩之機率的確為$1/3$。

要注意的是, 在例2中, 丙說"有一個正面"(假設若有一正面一反面, 則有$1/2$之機率丙說有一個正面), 及問丙"有沒有正面? "丙答"有", 此二事件是不一樣的。前者之機率為$1/2$, 後者為$3/4$。

3 結語

2008年美國總統大選, 民主黨於8月底舉行全國代表大會, 決定正副總統候選人。接著在9月4日, 美國共和黨的全國代表大會上, 阿拉斯加州的州長裴林(Sarah Palin), 被提名為共和黨的副總統候選人。原先共和黨總統候選人馬侃(John McCain)的民意支持度, 落後民主黨的總統候選人歐巴馬(Barack Obama)。於提名斐林後, 馬侃人氣迅速竄升, 聲勢立漲, 在幾份不同的民調中, 均勝過歐巴馬, 共和黨陣營當然很興奮。但一位長期研究美國大選的專家, 維吉尼亞大學(University of Virginia)政治學者薩巴托(Larry Sabato), 根據1960年以來的資料, 指出全代會後民調結果與大選結果相符者, 只有一半, "跟丟銅板預測差不多(You could flip a coin and be about as predictive)"。又說"大會回憶褪色之迅速, 令人意外(It is really suprising how quickly convention memories fade)"。

民意如流水, 對政治人物無情, 是偉大國家的象徵。固然不用因全代會後民調領先而過度高興。但對共和黨而言, 是否全代會後隨即做的民調, 不論領先或落後, 於當年11月的總統大選, 其提名人當選或落選之機率相同, 也就是皆為$1/2$? 如果眞是如此, 那全代會後所做之民調, 就確實是沒用了。民調無用, 統計工作者可能會有點沮喪。但統計學者針對此問題, 有沒有可以著墨處? 還是聽了那位政治學者之分析後, 便只能閉嘴?

依薩巴托的分析, 可假設

$$P(\text{當選}|\text{領先}) = P(\text{落選}|\text{領先}) = 1/2, \tag{13}$$

其中“領先”表在兩黨全國代表大會, 已決定正副總統候選人後, 在對兩組候選人所立即做的民調, 共和黨領先; “當選”表在當年總統大選時共和黨獲勝 。類似地, 可定義“落後”及“落選” 。在(13)式之假設下, 我們想知道

$$P(當選|落後) = P(落選|落後), \tag{14}$$

是否成立? 如果(14)式成立(即(14)式左 、右二側之機率皆為1/2), 則全代會後的民調領先或落後, 共和黨便可不必在意了 。甚至此民調根本就是多餘的 。

令

$$P(當選|落後) = a, \tag{15}$$

則

$$P(落選|落後) = 1 - a \text{ 。} \tag{16}$$

但由(13)式, 並無法決定a值 。我們再令

$$P(當選) = r,\ P(領先) = s, \tag{17}$$

其中$0 < r,\ s < 1$ 。仍由定理1, 得

$$P(當選) = P(當選|領先)P(領先) + P(當選|落後)P(落後) \text{ 。} \tag{18}$$

再由(13) 、(15)及(17)式, 且利用

$$P(落後) = 1 - P(領先) = 1 - s,$$

(18)式可改寫為

$$r = \frac{1}{2}s + a(1 - s) \text{ 。} \tag{19}$$

即

$$a = \frac{r - s/2}{1 - s} \text{ 。} \tag{20}$$

若$r=s=1/2$, 則$a=1/2$, 且(14)式成立。也就是若過去的資料顯示, 兩黨全代會後做的民調, 及選舉結果, 兩黨表現眞的一樣(即民調領先, 及當選機率, 皆為1/2), 則全代會後的民調領先或落後就眞的對當選與否, 沒有影響了。至於若$r=0.48$, $s=0.5$, 則$a=0.46<1/2$; 若$r=0.52$, $s=0.6$, 則$a=0.55>1/2$。a之值乃與r及s有關。

所以共和黨不必因聽了"專家"的話, 就誤以為全代會後的民調結果, 對大選時誰當選沒有影響。民主黨也是一樣。當然11月大選,如大家所知是歐巴馬當選。此時之前說誰當選的機率為何,便都沒用了。條件機率之功能再度顯現(給定的條件是歐巴馬當選)。

總之, 在應用機率, 特別是處理條件機率時, 須得很謹愼,否則極易犯錯。最後我們給幾道習題讓大家練習。

1. 假設有四個盒子, 其中恰有一盒裝有獎品, 且設主持人知道獎品在那一盒。某君任選一盒, 然後主持人打開其餘三盒中之一盒, 並發現其中並無獎品。此時若該君不改變選擇, 則獲獎之機率為何? 若該君改自其餘兩盒中任選一盒, 則獲獎之機率為何?

2. 考慮下述賭局。莊家 A 向顧客們展示三張牌,J,Q,K 各一張。洗牌後, 將一張牌放在一盒子內, 另兩張面朝下放在桌上。有一助理 B, 翻了桌上兩張牌, 然後拿起一張, 並展示給顧客們看, 假設是 Q。此時顧客可以開始賭在盒內的牌是否為 K。賭法如下:賭盒內的牌為 K 的, 要買張8元的票, 若對了,可得18元; 賭盒內的牌不為 K 的, 票每張11元, 若對了, 亦可得18元。有些顧客想, 盒內的牌會是 K 的機率為1/3, 因此他們買非 K 的票, 心想有 2/3 的機率會得到18元, 平均可得12元, 比票價11元還多1元。另有一些顧客想, 只剩兩張牌, K 和 J 各有 1/2 的機率, 因此他們買 K 的票, 心想有 1/2 的機率會得到18元, 平均可得9元,比票價8元還多1元。另一顧客 C, 他偷偷地要求 B 透露一些內幕。B 告訴他: A 的洗牌絕無問題。而 B 所接受的指示是, 翻看桌上的兩張牌,若都不是 K, 則隨便拿一張給顧客看。但若桌上的兩張牌中有一張是 K, 則 B 必須拿非 K 的那張

給顧客看。此因票是事先印製的, 若讓顧客知道 K 已不在盒內, 則此賭局便無意義了。試問 C 該如何賭?

3. 電影**越戰獵鹿人**(The Deer Hunter)裡, 有一描述虐待戰俘的方法。在一可裝6發子彈的左輪手槍(revolver)裡, 只放一顆子彈, 隨機地一轉後, 要二戰俘輪流用手槍向自己的頭部發射, 直到一名戰俘中槍, 另一名戰俘才逃過一劫。這就是所謂**俄羅斯輪盤**(Russian roulette)的遊戲。試問

   (i)先發射者是否較不利?

   (ii)若改為放兩顆子彈,結果有何不同?

   (iii)若改為每次發射前,均須將彈匣隨機地一轉,則結果有何不同?

參考文獻

1. 黄文璋(2003). 隨機思考論(第五章瞻前顧後)。華泰文化事業股份有限公司, 台北。

2. Hille, J.W. (1978/79). A Bayesian look at the jury system. *Mathematical Spectrum* 11, 45-47.

3. Stewart, I. (1996). The interrogator's fallacy. *Scientific American* 275, No.3, 134-136A.

4. Tierney, J. (1991). Behind Monty Hall's doors: puzzle, debate, and answer? *New York Times*, July 21, 1991.