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

97年10月15日

◆統計裡常在做預測，但如波耳(Niels Bohr, 1922年諾貝爾物理獎)所說：

預測很難，尤其關於未來。

- 對隨機現象做預測時，誤差常難以避免。
- 誤差之意義，卻不易為一般人所理解。
- 有人因此不相信統計，以為統計只是謊言。

不是謊言，那

統計是什麼？

◆隨機、誤差，及機率之意義，均不易說清楚。

◆特別是條件機率：

機率很難，尤其條件機率。

◆機率值會變，是機率的特性。

有人按門鈴，是男是女？

- 在給定某事件發生(有新的資訊)，原先那一事件發生的機率，有時會隨之而變。

- 何時不變？

  獨立！

- 條件機率看來似乎不難。
- 但實際應用時，並沒那麼容易。
- 當然數學要應用也不容易。

在一篇名為機率與文字陷阱的文章

例1.

(i)好友有二小孩，已知老大是女孩。

問：老二亦是女孩之機率？

(ii)好友有二小孩，已知有一個女孩。

問：另一小孩亦是女孩之機率？

例2.（某高中數學競試）

甲投擲兩硬幣，乙則猜朝上的面相同或相異。乙正準備要猜，丙從旁經過，說：

有一正面。

問：乙該猜相同或相異？

- 例2與例1-(ii)情況相同否？

該文指出：

- 但再仔細想想，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，儘管乙不知丙說什麼，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，那猜相異猜對機率就比較大(因為不管丙說什麼都要猜相異)。
- 結論：投擲兩硬幣，朝上兩面相異之機率是2/3，因為一定會有正面或反面。

- 疑問：明顯有錯，因國中以來就教相同、相異機率皆為1/2。

- 最後，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？

該文之結論

- 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。

說文解字

- 有時給的條件是以一段文字，描述一情境。如何將這段文字的內涵正確解讀，並不見得很容易。
- 若解讀錯誤，得到的條件機率當然也就不對。

條件機率

一則新聞

美國加州有一家庭，爸媽和剛出生的小孩，生日相同。…。機率只有

0.00000751。

兩人當初因生日相同，相信緣份天註定而結婚，沒想到第一個小孩也在同一天出生。

◆這對夫妻的第一個小孩，生日與他們同一天之機率，應為

$$\frac{1}{365},$$

並沒那麼小。

- 在樣本空間上定義機率。有時得到一些資訊，則樣本空間可能有所改變，機率空間也就隨之而變：

條件機率。

- 在數學裡不會如此。
- 給定某數是2，它就一直是2。
- 不變是數學特性。

◆定義1.

設A、B為樣本空間Ω中二事件，且$P(B)>0$。

給定B發生，A發生之條件機率：

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

◆例4.

玩梭哈要拿到4條之機率為

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

但若發了3張牌，皆拿到K，則此時會拿到4條之機率就大很多。

◆例5.

投擲1公正的銅板兩次。

求兩次皆得正面之機率，

給定

(i) 第1次得到正面，

(ii)兩次中至少有1正面。

分割

- 對一樣本空間Ω，事件$A_1, A_2, \ldots,$ 若滿足兩兩互斥，即

$$A_i \cap A_j = \varnothing, \forall i \neq j, \text{ 且} \bigcup_{i=1}^{\infty} A_i = \Omega,$$

便稱為Ω之一分割(partition) 。

- 有限分割 $A_1, A_2, \cdots, A_n, n \geq 2$ 。

◆定理1.

設$A_1, A_2, \ldots$為樣本空間之一分割。

則對任一事件B，

$$P(B)=\sum_{i=1}^{\infty} P(B \mid A_i)P(A_i) \text{ 。}$$

◆定理2. 貝氏定理

設$A_1, A_2, \ldots$為樣本空間之一分割。

則對$i \geq 1$，及事件B，只要$P(B)>0$，

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

- 例6.

甲、乙、丙三囚犯，國王宣佈以抽籤決定釋放一位，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。

甲請獄卒透露那一位被釋放遭拒後，改問獄卒：

乙及丙中，那一位會被處決？

獄卒經一番思考，遂(誠實地)告訴甲：

乙會遭處決。

他認為這樣做並未違反國王規定：

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，現在因只剩他與丙了，所以機率提高至1/2。

- 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。

K表獄卒說乙會被處決的事件。

樣本空間

$$\Omega=A\cup B\cup C。$$

由假設

$$P(A)=P(B)=P(C)=1/3。$$

想求$P(A|K)$。

$$P(K)=P(K\mid A)P(A)+P(K\mid B)P(B)+P(K\mid C)P(C)$$

$$=\frac{1}{2}\cdot\frac{1}{3}+0\cdot\frac{1}{3}+1\cdot\frac{1}{3}$$

$$=\frac{1}{2}\text{。}$$

$$\Rightarrow P(A\mid K)=\frac{P(\text{獄卒說乙會被處決，且甲被釋放})}{P(K)}$$

$$=\frac{1/3\cdot 1/2}{1/2}$$

$$=\frac{1}{3},\ \text{機率沒有變大！}$$

獄卒所提供的資訊是否無用？

- 若丙偷聽到獄卒與甲的對話，則知他會被釋放的機率提高至2/3。
- 若乙偷聽到獄卒與甲的對話，則知他沒有活命機會。
- 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2/3。

- 至於甲，被釋放之機率不會改變，還是1/3。
- 而三人被釋放之條件機率和：

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

◆此例有時候以不同的型式出現，如汽車與山羊問題(Car-Goat Problem)。

◆在電影決勝21點(21)中曾出現。

◆例7.

衛生局至高雄大學免費檢驗某疾病。假設檢驗結果有正、負兩種反應。

呈正反應，表可能有病，須至醫院進一步檢驗；

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。

你是否願意接受檢驗？

解.

可靠度90 %的意義：

若無病，檢驗會呈負反應之機率為0.9；若有病，檢驗會正負反應之機率亦為0.9。

想知道：

檢驗呈正反應下，有病的機率，及檢驗呈負反應下，無病之機率。

以正表檢驗呈正反應，負表檢驗呈負反應。

- 則

$$P(\text{有病}|\text{正})=\frac{P(\text{正}|\text{有病})P(\text{有病})}{P(\text{正}|\text{有病})P(\text{有病})+P(\text{正}|\text{無病})P(\text{無病})}$$

$$=\frac{0.9\cdot 1/5000}{0.9\cdot 1/5000+0.1\cdot 4999/5000}$$

$$=\frac{9}{5008}$$

$$\sim 0.001797$$ 。

- 當檢驗呈負反應下，的確無病的機率：

$$P(\text{無病}|\text{負})=\frac{44991}{44992}\sim 0.999977773\approx 1\text{。}$$

檢驗可靠度	90%	95%	99%	99.90%	99.99%
P(有病|正)	0.001797	0.003786	0.019419	0.166556	0.666689

表1. 患病比率1/5000，於不同檢驗可靠度下之P(有病|正)

患病比率	1/5000	1/500	1/50	1/5	1/2	4/5
P(有病|正)	9/5008	9/508	9/58	5/13	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)

◆例8.

有一對夫妻剛搬進某社區，大家只知道他們有兩個小孩，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.

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。

將女孩改為男孩，類似地定義B_1及B_2。

要求$P(G_1 \cap G_2|G)$。

$$P(G_1 \bigcap G_2 \mid G) = \frac{P(G_1 \bigcap G_2 \bigcap G)}{P(G)} = \frac{P(G_1 \bigcap G_2)}{P(G)},$$

$$\begin{aligned} P(G) &= P(G \mid G_1 \bigcap G_2)P(G_1 \bigcap G_2) + P(G \mid G_1 \bigcap B_2)P(G_1 \bigcap B_2) \\ &\quad + P(G \mid B_1 \bigcap G_2)P(B_1 \bigcap G_2) + P(G \mid B_1 \bigcap B_2)P(B_1 \bigcap B_2) \\ &= \frac{1}{4} + \frac{1}{4}P(G \mid G_1 \bigcap B_2) + \frac{1}{4}P(G \mid B_1 \bigcap G_2) \quad \text{。} \end{aligned}$$

用到

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

且

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

$$\Rightarrow P(G_1\cap G_2\mid G)=\frac{1/4}{1/4+P(G\mid G_1\cap B_2)/4+P(G\mid B_1\cap G_2)/4}$$

$$=\frac{1}{1+P(G\mid G_1\cap B_2)+P(G\mid B_1\cap G_2)}\text{。}$$

此機率與$P(G|G_1\cap B_2)$ 及$P(G|B_1\cap G_2)$有關！

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，媽媽會帶老大出門之機率為一定值p。

即設

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

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

⇒原問題之答案為1/2，與p無關。

<u>情況(ii).</u> 假設當兩小孩之性別不同，則媽媽會帶女兒出門之機率為一定值q。

即設

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

$$\Rightarrow P(G_1 \cap G_2 \mid G) = \frac{1}{1+q+q} = \frac{1}{1+2q},$$

與q有關。

特例

- 設q=1。即若有兒子及女兒，媽媽一定帶女兒出門 $\Rightarrow P(G_1 \cap G_2|G)=1/3$。

  此即例5-(ii)。

- 因

  看到該家媽媽帶女兒出門$\Leftrightarrow$該家至少有一女兒。

- 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 $\Rightarrow P(G_1 \cap G_2|G)=1/2$。

- q=0 $\Rightarrow P(G_1 \cap G_2|G)=1$。

- 除非有其他資訊，否則

看到該家媽媽帶著一個女兒，

不等價於

該家至少有一女兒。

- 注意：情況(i)及情況(ii)並非樣本空間之一分割，兩情況也不互斥。
- 結論：由所給條件，無法解出本問題。
- 為什麼如此？

- 樣本空間Ω中共有8元素，包含所有型如(s_1, s_2, i)的樣本，

  s_1：老大性別，

  s_2：老二性別，

  i：見到媽媽帶著的小孩之排序。

- 欲知$P(\omega), \forall \omega \in \Omega$，光給$(s_1, s_2)$之機率不夠。

◆設生男生女的機率均為1/2，Ω中各元素的機率：

$$P(\{女,女,I\})=\frac{p_1}{4},\ P(\{女,女,II\})=\frac{1-p_1}{4},$$

$$P(\{女,男,I\})=\frac{p_2}{4},\ P(\{女,男,II\})=\frac{1-p_2}{4},$$

$$P(\{男,女,I\})=\frac{p_3}{4},\ P(\{男,女,II\})=\frac{1-p_3}{4},$$

$$P(\{男,男,I\})=\frac{p_4}{4},\ P(\{男,男,II\})=\frac{1-p_4}{4},$$

I, II分別表媽媽帶著的小孩為老大或老二。

p_1為給定兩個小孩皆為女孩下，I發生之機率，p_2，p_3，及p_4依此類推。

因

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

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

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

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

$$\Rightarrow P(G_1 \cap G_2 \mid G) = \frac{1}{1+p_2+1-p_3} = \frac{1}{2+p_2-p_3}。$$

知道$p_2 - p_3$，才能得$P(G_1 \cap G_2|G)$。

- 當$p_2 = p_3 = p$

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

與情況(i)吻合。

- 當$p_2 = q$，$p_3 = 1-q$

$$\Rightarrow P(G_1 \cap G_2 \mid G) = \frac{1}{1+2q}。$$

與情況(ii)吻合。

- 情況(i)及(ii)，均只為本一般情況之特例。

尚可有其他情境

(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小孩及一條狗在社區公園玩，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。

- 分別對三情況， 求兩小孩皆為女孩之機率。

回到一開始那三個例子

- 例1-(i)之答案顯然為1/2。
- 例1-(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-(ii)，可得另一小孩亦是女孩之機率，確為1/3。

- 對於例2，若假設當丙看到兩硬幣有1正面及1反面向上，便分別有1/2的機率說有一個正面及有一個反面，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。即在這種情況下，丙所提供的資訊沒用。
- 但若$q \neq 1/2$，則丙所提供的資訊就有用。可類似如例8的討論。

- 在例3，仍對應例5-(ii)，兩小孩皆為女孩之機率的確為1/3。

- 注意：例2中，

  丙說有正面(假設若有一正面一反面，則有1/2之機率丙說有正面)

  及

  問丙有沒有正面？丙答有。

  此二事件不同。

- 前者之機率為1/2，後者為3/4。

結語

2008年美國總統大選

- 9月4日，美國共和黨的全國代表大會上，阿拉斯加州長裴林(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)。
- 提名裴林後，馬侃人氣迅速竄升，聲勢立漲，在幾份民調中，均勝過歐巴馬。

- 維吉尼亞大學政治學者薩巴托(Larry Sabato)，根據1960年以來的分析，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.

- 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？
- 民調無用？只需丟銅板？

◆依薩巴托的分析，設

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

◆其中

領先表兩黨已決定正副總統候選人後，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。

$$P(\text{當選}|\text{落後})=P(\text{落選}|\text{落後})\,?\qquad (**)$$

- 若成立，則全代會後的民調領先或落後，共和黨便可不必在意。甚至此民調根本就是多餘的。

- 令

$$P(\text{當選}|\text{落後}) = a,$$

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

- 由(*)並無法決定a。再令

$$P(\text{當選}) = r\text{，}P(\text{領先}) = s\text{。}$$

$$\Rightarrow P(\text{當選}) = P(\text{當選}|\text{領先})P(\text{領先})$$

$$+ P(\text{當選}|\text{落後})P(\text{落後})\text{。}$$

$$\Rightarrow r = \frac{1}{2}s + a(1-s),$$

$$\Rightarrow a = \frac{r - s/2}{1-s}。$$

- $r = s = 1/2 \ \Rightarrow a = 1/2 \Rightarrow$ (**)成立。
- $r = 0.48$，$s = 0.5 \Rightarrow a = 0.46 < 1/2$；
- $r = 0.52$，$s = 0.6 \Rightarrow a = 0.55 > 1/2$。
- a值乃與r及s有關！

機率應用不易！