運氣好不行嗎？

黃文璋

國立高雄大學應用數學系

為鼓勵購物索取統一發票(底下簡稱發票), 財政部訂出發票對獎辦法, 每兩個月開獎一次。發票有8碼, 每期開出最大的特別獎1組, 獎金高達1,000萬元; 其次頭獎3組, 獎金各200萬元, 一直到最小的六獎, 獎金200元。要中大獎當然很難。六獎則是看發票的末3碼, 是否與頭獎中獎號碼末3位相同, 即通常有3組。而末3碼由000至999, 共有1,000組, 因此每張發票中六獎的機率為千分之3。中大獎很難, 故本文所說的中獎, 皆指中六獎。有時會增加幾組六獎, 即使如此, 中獎機率仍不太高。大部分的人, 並沒太多發票, 每期能中獎一張便很高興了, 但若中太多, 有時也會啓人疑竇。

民國100年2月, 報上有一則發票中獎的新聞。某夫婦在兩期(即四個月)裡, 中獎24張, 引起國稅局注意。這對夫婦抱怨"運氣好難道有罪嗎?" 針對此事件, 黃文璋(2011)一文, 便是討論這樣的好運, 究竟有多不尋常? 以說明國稅局的關切, 是否合理。

無獨有偶, 民國102年8月14日, 數家媒體, 均有一則"中12張200元發票遭逼繳回"的報導。原來有位大學生, 於前一年7、8月間, 與同學在某家網咖消費, 共得到百餘張發票, 其中中獎12張。兌獎後, 卻接到通知, 要該生到國稅局說明, 有官員甚至建議其繳回全部2,400元的獎金。這位大學生遂向媒體投訴說"運氣好也不行嗎? 國稅局根本是找麻煩！"

民國101年7、8月那期增開兩組六獎, 中獎機率提高至0.005。這家網

咖同一期所開發票, 另有13張中獎, 亦即該店在一期裡, 共中獎25張, 這也讓國稅局產生懷疑 。

報載財政部賦稅署(負責監督各地區國稅局的業務)表示, 以中獎機率來說, 該名大學生和同學要在同一家店消費3,000張發票才有可能, 因此認為是“不合常情” 。台中女中教務主任則估算, 連中12張的機率約為$2.4 \cdot 10^{-28}$; 至於一家店要連中25張, 機率約為$2.98 \cdot 10^{-58}$ 。而威力彩頭獎的中獎機率約為$4.4 \cdot 10^{-8}$, 相對高出很多 。針對在一期內開出25張中獎發票, 該網咖的一名店員說“純屬巧合吧!” 雖黃文璋(2011)一文已說明, 但類似事件不時發生, 對究竟是運氣或異常, 仍常是各說各話, 認知有很大差異 。而看似專業人士者, 所提供之說明, 常也存在若干問題, 參考價值有限 。底下便來對此事件中, 所涉及的機率, 略做討論, 盼能釐清一些概念 。

運氣好當然沒有不行, 但過度好的運氣, 讓人產生懷疑, 也屬合理 。媒體上均只說該名大學生共有“百餘張”發票, 未明講有幾張 。假設有n張, 每張中獎機率p 。每張發票中獎之機率雖可視為相同, 但各發票是否中獎, 相互間其實並不見得獨立 。因與樂透彩不同, 發票號碼不能自己挑選, 也不是隨機產生 。同一家店接連開出的發票應為連號, 故若前一張中獎, 則後一張中獎機率將可能減小 。由於只想要有個粗略的了解, 不妨就將各張發票中獎與否, 視為獨立 。因此n張中, 總共中的張數, 便可以**二項分佈**(binomial distribution), $\mathcal{B}(n,p)$來描述 。即在n張發票裡, 中k張的機率為

$$\binom{n}{k} p^k (1-p)^{n-k},$$

其中**二項係數**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k = 0, 1, 2, \cdots, n \text{ 。}$$

當$n \geq 12$, 至少中12張的機率若以A_n表之, 則

$$A_n = \sum_{k=12}^{n} \binom{n}{k} p^k (1-p)^{n-k} \text{ 。}$$

由於依報導, 只知那位大學生共有百餘張發票, 而不確定究竟有幾張, 對$p = 0.005$, 分別令$n = 100$, 150, 及200, 求出A_n的值, 藉此了解中獎至少12張的機率之大小。這些機率值均很小, 利用計算機, 可得

$$\begin{aligned} A_{100} &\doteq 1.8185 \cdot 10^{-13}, \\ A_{150} &\doteq 2.2211 \cdot 10^{-11}, \\ A_{200} &\doteq 6.2633 \cdot 10^{-10}\text{。} \end{aligned}$$

不難看出, 對固定的p, A_n隨著n的增大而增大。我們所給皆為至少中獎12張的機率, 也就是中獎12張及更極端的事件之機率。因如果中獎12張都已令人起疑, 則中獎張數為$13, 14, \cdots$, 也都會令人懷疑。因此求出中獎數會落在此區域(即至少12)之機率, 在**無罪推定**(也就是沒有不法)之假設下, 看此機率是否小到足以讓人覺得這是一**顯著**(或說不尋常、異常)事件, 國稅局該不該去追究。

我們先看$A_{100} \doteq 1.8185 \cdot 10^{-13}$到底多小? 不妨以投擲銅板來說明。假設有人投擲一銅板, 連續40次皆得正面。他宣稱該銅板為公正, 你是相信他運氣好, 或有特異功能, 還是對該銅板的公正性產生懷疑? 事實上此事件(即公正銅板連續擲出40次正面)發生之機率等於0.5^{40}, 約為$9.0949 \cdot 10^{-13}$, 而A_{100}比此值還小, 不到它的5分之1。又A_{150} 比投擲一公正銅板, 連續出現35次正面之機率(約為$2.9104 \cdot 10^{-11}$)還小。至於A_{200}, 則比連續30次投擲皆得正面之機率(約為$9.3132 \cdot 10^{-10}$)還小。公正銅板連續投擲出30次正面究竟有多稀奇? 即使全台灣2,300萬人, 每人皆執行此實驗一回(即連續投擲一公正銅板30次), 平均也約只能看到$0.021(\doteq 2.3 \cdot 10^7 \cdot 9.3132 \cdot 10^{-10})$次有人成功(即連擲出30次正面)。至於35次及40次, 當然更不易看到。所以這位大學生, 你這樣的好運氣, 眞是太不尋常了。國稅局職責所在, 想查明詳情, 實在不能算找你麻煩。

連中獎12張的機率, 即0.005^{12}, 的確如那位中學教務主任所提供, 約為$2.4414 \cdot 10^{-28}$; 連中25張的機率, 即0.005^{25}, 也的確約為$2.9802 \cdot 10^{-58}$(我們皆算至小數第4位)。只是這兩個機率值, 小則小矣, 對了解本問題, 卻毫無幫助。該大學生有一百多張發票, 人們所關心的, 乃是"其中"(至少)有

某12張中獎的機率到底有多小? 怎會在乎是否連中呢? 同理, 人們關心的是, 該網咖在一期(至少)中獎25張的機率, 這與連中25張有何關連?

其實不少人都有上述"連中"的迷思, 底下我們藉威力彩來說明。威力彩的投注, 是自01至38的號碼中任選6碼, 且自01至08的8碼中任選1碼。每期開出一組6加1的號碼, 做為頭獎號碼。因此中頭獎的機率為

$$P = \frac{1}{\binom{38}{6} \cdot 8} = \frac{1}{22{,}085{,}448} \doteq 4.527868305 \cdot 10^{-8},$$

即約1億分之4.5的機率。那位教務主任以$4.4 \cdot 10^{-8}$當做P之近似值, 誤差未免太大, 至少該說$4.5 \cdot 10^{-8}$。現假設有人第二度中威力彩頭獎, 雖屬三生有幸事件, 但若有媒體報導這種事發生之機率為$P^2 \doteq 2.050159139 \cdot 10^{-15}$, 且補充說明因每週開兩期, 一年104期, 因此平均約

$$\frac{1}{104 \cdot P^2} \doteq 4.69 \cdot 10^{12}(\text{年}),$$

即約每4.69兆年才會發生一次, 你覺得這樣講正確嗎? 曾有人估計地球誕生至今約50億年。4.69兆年, 是地球年齡的938倍呢! 這麼稀罕的事居然發生? 只是即使僅在美國, 便已產生好幾位二度中樂透彩的幸運兒, 因此真得要4.69兆年那麼久才會發生一次嗎?

實際上, P^2為這一期買一張威力彩彩券, 下一期又買一張, 兩張皆中頭獎之機率。對一特定的人, 若一期買好幾張, 且經常買(不少人是如此), 則在他一生裡, 會中兩次威力彩頭獎之機率, 乃比P^2大很多。而全台灣有人在一生裡中兩次, 其發生之機率便又更大許多了。至於全世界, 人口是台灣的3百多倍, 若每隔一陣子, 便有人第二度中樂透彩頭獎, 可說相當合理。

再來看國稅局對該網咖的懷疑, 即一期所開出之發票, 中獎25張, 是否有道理? 一般網咖的營運情況我們並不了解, 若情況良好, 則在兩個月內開出的發票不該太少, 而只要數量夠大, 要中至少25張便很容易。假設該網咖在前述7、8月那期中, 共開出5,000張(即平均一天約開出81張)發票, 則其中至少中獎25張之機率約為0.5268, 並不算小; 若一期中開出10,000張

發票, 則至少中獎25張之機率, 約為0.9999, 已近乎1了。網咖店員對該店一期中獎25張, 認為是"純屬巧合", 算是合理的評斷。

另外, 賦稅署說"要在同一家店消費3,000張發票才有可能(12張中獎)", 這講法正確嗎?

人們有時會以期望值來反映事件是否"可能"發生。由於中獎機率為0.005, 因此2,400張發票中, 中獎張數之期望值為

$$2,400 \cdot 0.005 = 12 \text{ 。}$$

有人便認為若有2,400張以上的發票, 便"可能"中獎12張, 而若發票數少於2,400張, 便"不太可能"中獎12張。此概念並不正確。要知即使僅有12張發票, 也"可能"全中, 雖機率如前所述很小, 僅約$2.4414 \cdot 10^{-28}$。不需3,000張, 若有2,000張發票, 其中至少12張中獎之機率, 約為0.3029, 已算夠大了, 因此發生不難。若有2,400張發票, 至少中獎12張之機率約為0.5387, 大於2分之1, 可說發生很容易。至於若有3,000張發票, 此機率將大到約0.8159。可以這麼講, 大學生、國稅局(及賦稅署)、教務主任(他擁有博士學位), 及店員等四方, 觀念比較清楚的, 應該是那名店員。眾大小自以為是的專家, 都要靠邊站。

我們再舉一例來說明, 在有些情況下, 期望值並不見得能輕易轉化成描述事件發生可能性之大小。假設一袋子中有A, B兩種牌, A牌有999張, B牌只有1張, 共有1,000張牌。又設A牌上寫0, B牌上寫1,000。今自袋中隨機地取一張牌, 顯然取中A牌之機率為0.999, 取中B牌之機率為0.001。則所得點數之期望值為

$$0.999 \cdot 0 + 0.001 \cdot 1,000 = 1 \text{ 。}$$

期望值雖為1, 但會取中1點以上之機率為0.001, 相當小。反之, 會取中1點以下之機率為0.999, 相當大。至於取中1點的機率則為0, 根本不會發生。

最後, n很大時, 二項分佈$\mathcal{B}(n,p)$之機率值並不太好求。但當n很大, p很小, 而np適當地大(有些書說$np \leq 7$), 便可以參數為np的Poisson分佈, 即$\mathcal{P}(np)$分佈, 來近似, 這是所謂**Poisson近似**。對$n = 100, \cdots, 200$,

$p = 0.005$, 圖1給出A_n之值, 及以$\mathcal{P}(np)$分佈所得之近似值, 以為比較。注意縱軸一單位長表機率值10^{-10}。因np太小, Poisson近似, 尚不算很精準。至於當n較大時, 雖也可用常態分佈來近似二項分佈$\mathcal{B}(n,p)$, 但一般要np及$n(1-p)$皆大於5才適用(即誤差比較小)。此處由於$p = 0.005$, 對n由100至200, $np \leq 1$, 故若以常態分佈來近似, 誤差將會很大。例如, $A_{150} \doteq 2.2211 \cdot 10^{-11}$, 若以Poisson分佈來近似, 得$3.3141 \cdot 10^{-11}$, 但若以常態分佈來近似, 得$4.5547 \cdot 10^{-39}$, 誤差可說極大。

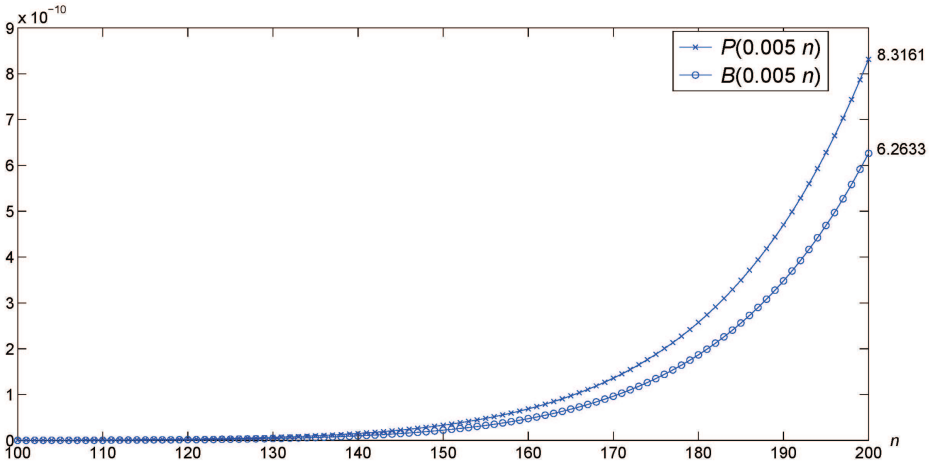

圖1. A_n之實際值及以Poisson分佈所得之近似值

一事件只要發生之機率為正, 不論多小, 便都有可能發生。小機率事件由於不容易發生, 因此一旦見到, 讓人特別留意, 乃屬合理。利用機率, 雖無法證明事情的眞偽, 但可提供決策者參考, 以進一步探究。

參考文獻

1. 黃文璋(2011). 運氣好有罪嗎? 科學人, 115期(2011年9月號): 30。