純屬巧合

黃文璋教授
國立高雄大學應用數學系

1 人生何處不相逢

西元1995年8月我與內人帶著女兒至北京開會,開完會搭飛機至西安遊覽。在北京機場遇到一個中原大學的教授訪問團,也要搭同一班飛機至西安。其中有一位還是數學系的教授。那時覺得眞巧,因在台灣還不容易碰到他。抵西安後,去參觀兵馬俑。排隊買票時,遇到一位我們系上的教授,我事先並不知道他也到西安來。在西安除了參觀古蹟,有一晚我們去欣賞仿唐樂府的表演。那是一場頗具水準的演出。中場休息回頭一看,見到中央研究院統計科學研究所的一位研究員及其夫人,再度事先我並不知他們也來到西安。眞是人生何處不相逢!

浙江餘姚人的余秋雨,在"鄉關何處"一文中,也提到底下的經歷(見余秋雨(1995)) 。他十歲離開家鄉,到上海求學,自此很長一段時間未回故鄉。重新揀回故鄉是在上大學之後。先是發現王陽明亦爲餘姚人,接著黃宗羲、朱舜水等中國文化史上兩位讓他景仰的大師亦爲同鄉。進上海戲劇學院讀書後,發現全院學術威望最高的兩位教授都是餘姚人。近幾年怪事更多了。有一次他參加上海市的一個教授評審組,其中有好幾個來自各大學的評審委員都是餘姚人。他又舉了好些例子,以佐證故鄉在追蹤和包圍他,使他到處遇到餘姚人。

生活中常充滿著各式各樣令我們感到很驚訝的巧合。某回你至美國洛杉磯(Los Angeles) 的迪斯奈樂園(Disneyland) 旅遊,突然遇到一多年未見的朋友。你可能覺得很神奇,居然在同一段時間你們都來到美國,而美國這麼大,在同一天你們卻來到同一地方。

Diaconis and Mosteller (1989) 一文,及Weaver (1982) 一書的Chapter

XIII, 列出不少令人吃驚的巧合事件。Weaver (1982) 還引了Lewis Carroll 所著愛麗思夢遊仙境(Alice in Wonderland) 中的一段話:

"There is no use trying," Alice laughed, "One can't believe impossible things."

"I daresay you haven't had much practice," said the (White) Queen. "When I was your age I always did it for half an hour a day. Why, sometimes I've believed as many as six impossible things before breakfast."

在早餐前就有6件不可能的事!這是怎麼回事?事實上如果我們留意的話,經常就是會遇到我們以爲絕不可能發生的事。但爲什麼這麼許多巧合,這麼許多不可能發生的事,一再地發生呢?

Fraser (1999) 一文舉出許多她認爲巧不可言的例子。然後說,這些事例說明爲什麼正統科學家都只把巧合當成偶然現象:因爲這樣的巧合違背了已知的因果律,也違背了時空關係律。然而,也正是如此匪夷所思才引人遐想。精神病學家卡爾・榮格(Carl Jung) 指出,認爲巧合毫無意義是既忽視了巧合冥冥中自有道理,也忽視了其中的人爲因素。Fraser 又寫著,大多數現代的研究者只收集整理巧合事件,把理解巧合發生過程的工作留給物理學家,把解釋巧合意義的任務交給心理學家。

Fraser 顯然忘了除了精神病學家、物理學家及心理學家,正統科學家中,尚有機率學家的存在。那麼機率學家是如何看待巧合事件呢?

註1. Fraser (1999) 文中之Carl Jung ,即爲Diaconis and Mosteller (1989) 一文所屢提及的C.G. Jung。

2 今生緣

我們先看底下著名的生日問題(birthday problem,或稱重覆生日問題, problem of repeated birthdays)。一個房間中要有多少人,才會使其中有兩人生日相同的機會超過沒有人生日相同的機會?所謂生日相同是指出生的月份及日期皆相同。爲了簡化問題,我們不妨忽略二月份有時會有29天,因此共有365個不同的生日。又我們也假設每日出生的機會皆相

同,即1/365 。雖然有研究指出,假日之出生率較低(見Cohen (1983) 及其中所列文獻),但忽略這些因素對整個問題並無太大影響。

曾有人對一群大學生提出上述問題,結果學生所猜答案的中間值(median)爲385。大部分人的答案明顯地太高且不合理,此因只要一房間中有366人(或367人,若考慮2月29日),則必至少有2人生日相同。事實上正確的答案是只要23個人便夠了。推導並不困難。首先一房間中之 n 個人生日皆相異的機率爲

$$(1)\qquad p_n = \frac{365}{365}\frac{364}{365}\cdots\frac{365-n+1}{365} = \frac{365\cdot 364\cdots(365-n+1)}{365^n} = \prod_{i=1}^{n-1}(1-\frac{i}{365}),$$

當然 n 要不超過365,若 $n>365$,顯然 $p_n=0$。因此 n 人中,至少有2人生日相同的機率爲 $1-p_n$。又可看出 $n\le 365$ 時,p_n隨著 n 之增大而減小。對一些不同的 n,我們列出 $1-p_n$ 之近似值如下:

n	4	16	20	22	23	32	40	48	56	64
$1-p_n$	.016	.284	.411	.476	.507	.753	.891	.961	.988	.997

表1. n 人中至少有二人生日相同的機率

由表1知,一房間中只要有少少的23人,則會有2人生日相同的機率便超過1/2,也就是此時有人生日相同較沒有人生日相同更易發生。若有40人,則會有2人生日相同的機率約爲0.891。若有64人,則會有2人生日相同的機率約爲0.997,已相當接近1了。

在一場足球賽中,兩方各有11名球員,再加上1名裁判(那麼大的足球場中只有1名疲於奔命的裁判),場中共有23人,因此不難在其中找到2人生日相同。不少團體人數都在23人以上,如國中一班有40人左右,要找到2人生日相同的機率約爲0.891。也就是約有九成的國中班級,可在班上找到2人生日相同。在一團體中,2人若生日相同,可能彼此覺得眞有緣份,備感親切。我們現在知道這其實是一很容易發生的事件。

在紅樓夢第62回,於發現寶玉、平兒、寶琴及岫煙4人生日相同後,探春笑道"倒有些意思,一年十二個月,月月有幾個生日,人多了,就這樣巧。也有三個一日的兩個一日的。大年初一也不白過,大姐姐占了去,怨不得他福大,生日比別人都占先,又是太祖太爺的生日冥壽。過了燈節,就是大太太和寶姐姐,他們娘兒兩個遇的巧。三月初一是太太的,初九是璉二哥哥,二月沒人。"襲人道"二月十二是林姑娘,怎麼沒人?只不是偺們家的。"探春笑道"你看我這個記性兒!"寶玉笑指襲人道"他和林妹妹是一日,他所以記得。"

生日問題爲下述問題之一特例。

例1. 將n個球隨機地丟進M個盒子, $n \leq M$,則沒有一盒子中有兩個以上的球之機率爲

$$P(A) = \frac{(M)_n}{M^n} = \prod_{i=1}^{n-1}(1 - \frac{i}{M}), \tag{2}$$

其中

$$(M)_n = M(M-1)\cdots(M-n+1)。 \tag{3}$$

至於$n > M$會如何?由鴿籠原理(The Pigeon-Hole Principle)知,此時至少有一盒子中有兩個以上的球之機率爲1。現設n個球,隨機地放進n個盒子,則每一盒子中恰有一球之機率,由(2)式知爲$n!/n^n$。當$n = 7$,此值約爲0.00612,比我們想像的小很多。換句話說,至少有兩球在同一盒中之機率很大,約爲0.99388。假設你住在一棟8層樓的大廈,若有7人自一樓同時進電梯,則要每一層恰有一人出電梯的機會其實很小, 1,000次中才約6次。也就是至少有一層有兩人以上一起出去的機會相當大。大家可能也曾注意到,要嘛就一天好幾通電話找你,要嘛就一整天一通也無,較少是均勻的,每天一樣多通。這其實也是差不多的道理。

類似的問題很多。假設高雄市某國中的某一班有5人畢業後進入高雄女中就讀。又設高雄女中一年級有24班,且新生是隨機地分班。則此5人中至少有2人高一同班的機率爲$1 - (24)_5/24^5$,約等於0.359。超過三分之一的機會,並不算低。高二及高三都可能重新分班, 所以若國中時同班,高

中又同班,對她們而言,很可能覺得今生眞是有緣。你現在知道了,發生的機會並不低。

高中同班要國中亦曾同班,才會感到巧合(只是同校不夠)。對年代久遠者,如大學同系,只要小學時曾同校, 便會覺得很巧。但一所小學,日後考上同一所大學5人並不罕見,而不少大學的學系還不到24個呢!何況有人小學還讀過不只一所。對年代隔愈久者,只要曾有一些薄弱的關係(如同鄉),則若遇在一起,便以爲巧妙了(如余秋雨的感覺)。年紀愈大,經歷愈多,曾有關係的人也就愈多,於是巧遇不斷了。

例2. 一團體中若人數增多,則要有更多人生日相同的機會也會增大。那至少要有多少人,才會使至少有 k 個人生日相同的機率大於1/2?此問題等價於將 n 個球隨機地丟進365個盒子,求 n 要多大,才會保證至少有一盒子中有 k 個以上的球之機率大於1/2。Levin(1981)解出此問題,他並給出表2(見Diaconis and Mosteller(1989))。

k	2	3	4	5	6	7	8	9	10	11	12	13
n	23	88	187	313	460	623	798	985	1,181	1,385	1,596	1,813

表2. 至少有 k 個人生日相同之機率大於1/2所需之人數 n

所以若一個年級有1,000個學生,則不難(機率大於1/2)在其中找到9個人生日相同。不過各位應也看出,n 並非隨著 k 而線性增大,而是成長快速:$k=2$ 時,n僅爲23; $k=3$ 時,n 便跳至88;$k=10$ 時,n更跳至1,181了。

例3. Abramson and Moser(1970)推廣生日問題,並得到要使有2人之生日差距不超過 m 天之機率大於1/2,所需之最少人數 $s(m)$(m=0即對應生日相同)的求法。他並給出表3。

表3指出,只要有14人,則其中有2人生日差距不超過1天之機會便超過一半。而只要有7人,便有一半以上的機會,其中有2人之生日差距不超過一週。雖然不是完美的吻合,但只要少少的7個人,也常可造成不小的驚訝。

m	0	1	2	3	4	5	6	7	8	9
$s(m)$	23	14	11	9	8	8	7	7	6	6

表3. 有2人生日差距不超過m天之機率大於1/2所需之最少人數$s(m)$

對於要使有2人生日相同之機率大於1/2所需之最少的人數23人,Stewart(1998)給出其中有兩對生日相同的機率約爲0.111,所以差不多9場足球比賽就會發生一次。而有三對生日相同之機率約爲0.018,差不多每55場比賽有一次。至於有3人生日相同之機率約爲0.007,小很多,此可與表2對照。

看了以上的例子,各位應已了解原來生活中所充滿的巧合,其發生的機會,往往並非我們所以爲的那麼低。處處留意皆文章,看來在早餐前,就遇到6件"不可能"的事,似乎是不難的。

另一方面,我們之所以對所遇的某些事件,感到不可思議,乃是因其發生違反我們的"常識"。若原先的常識被修正了,則對有些並非巧合的事件,便不至於再驚訝了。

3 他生緣

期待他日再相逢,共度白首。這是在鳳飛飛(一位民國六十年代的著名歌手)"流水年華"中,令人聞後惆悵不已的一句歌詞。再相逢是否容易呢?

我們先考慮一個與上節所討論的生日問題略微不同的問題。在一房間中,除你以外,要有多少人,方能使其中會有人生日與你相同的機會超過沒有人生日與你相同的機會?要不要猜猜看? 364/2+1=183似乎是一合理的猜測。此因一年中與你生日相異的日數(仍將一年當作有365天)有364天,因此若有超過此日數之半的人在房間中,似乎便較可能有人生日與你相同。正確的答案卻是253人。

作法仍類如解生日問題。先求出生日皆與你生日相異的機率,再用1減去此機率。設房間中除你之外有n個人,因每個人生日與你生日相異的機率爲364/365,故至少有一人生日與你生日相同的機率爲

$$q_n = 1 - (\frac{364}{365})^n \text{。} \tag{4}$$

若要使 $q_n > 1/2$,經由取對數,可解出最小的 n 爲253。比我們直觀上認爲的值大很多。

靈巧的讀者可能已看出253剛好等於生日問題中的23人所產生的對數,即23人中有 $\binom{23}{2} = 253$ 對生日。這似乎又是一個巧合。

對此結果你有什麼體會呢?要在一班上找到有兩人生日相同很容易,但要找到有人與"你"生日相同,則相當不容易。再以丟球爲例。你依序將球隨機地丟進 M 個盒子。本來你很高興才丟了幾個,便有兩球落在同一盒中。但若要你再讓一球進入已有兩球的那個盒子,你可能便很納悶爲何丟了許多卻再也丟不進了。

一足球場上的23人,其中共有253對生日,若有一對相同我們便很驚訝,至於其餘252對生日若都不同,我們可能並不留意。因爲不相同,所以未能引起我們的注意。我們一向只在意一致或吻合的事,以前言中之巧遇爲例,算一算我們其實認識不少人,若再加上與我們有關係的(如校友、同鄉、弟弟的朋友等),那更是不得了。而旅遊時,一般人所去的通常是幾個點,如機場、名勝古蹟、大飯店等,要遇到熟人,仔細想一想,可能已不是那麼不容易了。而巧遇事件總是令我們多年後都難忘。甚至我們連對朋友碰過的巧遇都印象深刻。至於你自己某次旅歐21天中,一個熟人都沒遇到的事當然早就拋到九霄雲外了。沒遇到熟人的事怎會值得記住?所以長期下來,留在我們腦海中的,盡是一些巧遇事件。

有一次我對女兒說"妳怎麼每次上完後不是喊很冷就是喊很熱。"她答"如果不冷也不熱我就不說了。"的確也是,我只記得她"有喊"的時候,由此便產生她"每次"都如何如何的印象。這似乎常是我們印象產生的模式。

我們常說"眾裏尋他千百度,暮然回首,那人卻在,燈火闌珊處"(這是辛棄疾"青玉案"詞中的一句,被王國維拿來做爲古今之成大事業、大學問者,必經過三種境界的第三境)也是類似的情景。我們回首千百次,什麼都沒看到時,當然不會有特殊的感覺,但若見到"那人",可能便感動莫名。

白居易(772-846)及元稹(779-831),此二位感情很好,且共創元和體,提倡通俗化詩歌的名詩人,各有一句詩可拿來做爲我們所討論的巧合問題之

註腳(見邱燮友註譯(1996)新譯唐詩三百首第208及207首,兩首詩接連在一起只是巧合。又元和體並非元白體之誤。此名之由來,乃因其風行於唐德宗貞元至唐文宗太和的三十年間。元與元稹的姓相同也只是巧合):一夜鄉心五處同;他生緣會更難期。對於與你鄉心同的今生緣,若想有他生緣,一千多年前的詩人早就告訴我們:太難了。

4 近似公式

在前兩節中,我們以生日爲例,來說明有些事件之發生爲眞正巧合,而有些我們視爲巧合的事件,卻其實毫不足爲其奇。之所以用並不聳動的生日當例子,乃是因其中涉及的機率較易計算。讀者不難將其想法應用到其他情況。

底下對一些巧合事件,我們給出近似公式,以增進實用性。

問題1.標準生日問題。

首先在例1中,當 M 大些,而 n 與 $M^{2/3}$ 比算是小的,則沒有一盒中有兩個以上的球之機率有下述近似關係:

$$\prod_{i=1}^{n-1}(1-\frac{i}{M}) \doteq e^{-n^2/2M} \text{。} \tag{5}$$

此處用到因 $\log(1-i/M) \doteq -i/M$,故

$$\sum_{i=1}^{n-1}\log(1-i/M) \doteq -\sum_{i=1}^{n-1}\frac{i}{M} = -\frac{n(n-1)}{2M} \doteq -\frac{n^2}{2M} \text{。} \tag{6}$$

要令至少有一盒中有兩個以上的球之機率爲 p 所需之 n,可由下式解出其近似值:

$$1-p = e^{-n^2/2M} \text{。} \tag{7}$$

當 $p=1/2$時,解出

$$n \doteq 1.2\sqrt{M} \text{。} \tag{8}$$

事實上,若要精確些,應是 $n \doteq 1.1774\sqrt{M}$,不過因在(6)式中,我們以 n^2 取代 $n(n-1)$,所以這裡取大一點的 n。我們不過是要藉一簡潔的公式,使得對不同的 M,能了解n 約要多大才行。當 $p=0.95$ 時,由(7)式解出

$$n \doteq 2.5\sqrt{M}, \tag{9}$$

再度應是 $n \doteq 2.4477\sqrt{M}$。當 $M=365$,且 $p=1/2$ 時,由(8)得 $n \doteq 22.9$,故取 $n=23$;當 $p=0.95$ 時,由(9)得 $n \doteq 47.76$,故取 $n=48$。

問題2.多類巧合問題。

生日問題有下述推廣。假設有一群人相遇在一起,互相談論及介紹自己。這時不同的巧合便可能出現了,包含生日相同、高中校友、同鄉、同一年出生、太太同姓、父親在同一機構上班等。問至少會有某一項巧合之機率爲何?

爲了簡化問題,假設這些不同的類之發生與否爲獨立事件,且設有 k 類,每類分別有$M_1, M_2, \cdots, M_k$ 項,又設每類中的每項之發生可能性皆相同。例如第一類中,每項之發生可能性皆爲 $1/M_1$,餘類推。現設有 n 個人,則仿(5)式,沒有一類中,有兩項以上發生的機率爲:

$$\prod_{j=1}^{k}\prod_{i=1}^{n-1}(1-\frac{i}{M_j}) \doteq e^{-\sum_{j=1}^{k} n^2/2M_j}\text{。} \tag{10}$$

故要使至少有一類中有兩項以上發生的機率至少爲 p,n 須滿足(近似值)

$$n \geq (-2\log(1-p)/\sum_{j=1}^{k} M_j^{-1})^{1/2}\text{。} \tag{11}$$

如前,常以1.2取代 $(-2\log(1/2))^{1/2}$。例如,$p=1/2$,且 $M_1=M_2=\cdots=M_k=M$, 則 $n \geq 1.2(M/k)^{1/2}$(若 $p=0.95$,則以2.5取代1.2)。又如設 $k=4$,且$p=1/2$,且取 $M_1=365$(表生日),$M_2=30$(表高中數),$M_3=20$(表出生年,一群朋友年紀差距通常不超過20歲),$M_4=35$(表35省)。則由(11),

$$n \geq 1.2(\frac{1}{365}+\frac{1}{30}+\frac{1}{20}+\frac{1}{35})^{-1/2} \doteq 3.54,$$

故只要4人便足夠。換句話說,只要有4個人,則其中有人生日相同、或是高中校友、或同年出生、或同鄉,其機會爲一半以上。要攀拉關係眞是太容易了。

問題3.多重巧合問題。

對於表2至少 k 個人生日相同之機率大於1/2所需人數 n, Diaconis and Mosteller(1989) 以一 k 之函數來求 n 之近似值。當 k 不太大時(譬如說小於20),還算接近。即

$$n \doteq f(k) = 47(k-1.5)^{3/2} \text{。} \tag{12}$$

例如,$k=5$ 時,得 $f(5)=307.7$,實際之 n 值(見表2)爲313;$k=10$ 時,$f(10)=1{,}164.7$,實際之 n 值爲1,181。雖有些誤差,但至少對不同的 k,可快速地掌握 n 約需要有多大。

Diaconis and Mosteller(1989)還對 M(1類中之項數,如例1中之 M)較大時(但 k 並未隨著變大),給出其中至少有 k 個爲同一項之機率爲 p,所需之n(球數或人數等)之近似值,可由解下述方程式得到:

$$ne^{-n/(Mk)}(1-n/(M(k+1)))^{-k^{-1}} = (-M^{k-1}k!\log(1-p))^{k^{-1}} \text{。} \tag{13}$$

例如,某人常對他的妻子及女兒,與他生日的日期相同(譬如說都是5號)感到很奇妙。事實上將 $M=30$(每月的日數),$k=3$ 及 $p=1/2$ 代入(13)式,解出 $n \doteq 18$(讀者不妨將 $n=18$,$M=30$,$k=3$,$p=1/2$ 代入(13)式,左側得約15.5575,右側得約15.5264)。換句話說,在每18個人中,有3個人生日的日期相同,其機率便超過二分之一。

問題4.生日接近問題。

對於例3中,要使有2人之生日差距不超過 m 天之機率大於1/2,所需之人數 n, Diaconis and Mosteller (1989) 亦給出 n 之近似值。設項數爲 M,$p=1/2$,則

$$n = s(m) \doteq 1.2(M/(2m+1))^{1/2} \text{。} \tag{14}$$

例如 $M=365$,$m=1$,代入(14)式,得 $n \doteq 13.2$;若 $m=9$,得 $n \doteq 5.2$。由表3,實際之 n 值分別爲14及6,相當精確。

5 其他巧合事件

1.新字現象。

相信不少人都有下述經驗,當你第一次聽到或讀到某一新的字(或新的詞),過不久便又再度聽到或讀到該新字,甚至那一段時間該新字常在各種地方出現。類似的例子很多。例如,你有位朋友買了一輛新車,性能不錯,讓你也有些心動。之後,你便常在路上看到該款車,你很奇怪爲什麼過去卻都未曾注意到。我們以新字爲例,來解釋這種巧合。

首先可能是因你的閱讀習慣改變了,你常讀此字較易出現的報章雜誌,當然便較易讀到該新字。其次可能是因此字令你印象深刻,在閱讀時你不禁特別留意該字,聽到該字時,也不由豎起耳朵。通常我們在閱讀時,眼睛是以掃描的方式,大部分的詞句是一掃即過,因此讀完一段文章,只粗略知道該段的意思。但現在可能一遇到該字,便停頓一下,自然讓我們覺得該字變成常出現了。另一個原因,可稱之爲環境的改變。例如,"戒急用忍"四字,以前並不常出現,僅見之於二月河(1994)p.53:雍正皇帝被他那幾位弟弟氣得手腳冰涼,正要發作,猛抬頭看見康熙皇帝賜給自己的條幅,一筆楷書端正寫著四個字"戒急用忍",氣遂消了。不過自李登輝總統提出來後,許多人便常引用,於是此四字就到處出現了,成爲一流行用語。

這種新字現象,其實並非屬於巧合的一種。但通常我們習而不察,也誤以爲是巧合,而每每拿來印證巧合的確處處在。

2.隨機取樣。

某校的一年級有五百位學生,每逢要出公差,該校便隨機抽取(將學生編號1至500)。一學期下來,共抽了50位,其中有4位被抽中兩次,但並沒有被抽中超過兩次者。該校有位數學老師懷疑這眞的是隨機抽取的嗎?

事實上,隨著母體數 M 的增大,自該母體中,隨機地抽取(每次取後放回)10% 的樣本,愈來愈不容易所抽中的皆相異。

由於一般的桌上型計算機無法計算 $(500)_{50}$,我們考慮較小的母體數 M,且自其中取 10% 的樣本。例如,設 $M=20$,則樣本數爲2。易見從20個相異事物中抽取兩個(每次抽取後放回),皆相異之機率爲 $(20)_2/20^2=$

0.95。故會有重覆之機率爲 $1-0.95=0.05$。一般而言,若樣本數爲 n,母體數爲 $10n$,且每次取出後放回,則樣本中會有重複之機率 r_n 爲

$$r_n = 1 - \frac{(10n)_n}{(10n)^n} = 1 - \prod_{i=1}^{n-1}(1-\frac{i}{10n})。 \tag{15}$$

對一些不同的 n,我們列出此機率值如表4(取自Nixon(1998/9))。

由表4看出 n 愈大,取樣重覆之機率 r_n 愈大。事實上,利用Apostol(1974) Theorem 8.55(設 $a_n \geq 0$,則 $n \to \infty$ 時,$\prod_{i=1}^{n}(1-a_i)$ 趨近至一不爲0之值,若且唯若 $n \to \infty$ 時,$\sum_{i=1}^{n} a_i$ 收斂), 因 $n \to \infty$ 時, $\sum_{i=1}^{n-1} i/10n = (n-1)/20 \to \infty$,故

$$\lim_{n\to\infty}\prod_{i=1}^{n-1}(1-\frac{i}{10n}) = 0,$$

即 $n \to \infty$ 時,$r_n \to 1$ 。

隨著母體數之增加,自其中取 10% 的樣本,愈來愈容易重覆(重覆之機率趨近至1)。所以日後對禍不單行(或福不單行),不用再太怨嘆(或太感激上帝)了,原來是這麼容易發生。

母體數	樣本數	取樣重覆之機率
30	3	0.098
40	4	0.143
50	5	0.186
60	6	0.227
70	7	0.226
100	10	0.372
200	20	0.626
300	30	0.777
380	38	0.853

表4. 自數量爲 $10n$ 之母體中隨機取 n 個樣本(每次取出後放回),會有重覆之機率

上述重覆取樣的容易讓我們驚訝,可能是與下述事件混淆。

分別丟一公正銅板二次,十次及一百次,何者較易得到正反面數相同?不少人以爲丟的次數愈多愈容易。事實上,恰好相反,愈不容易。

設丟一公正銅板 $2n$ 次,則會得到 n 個正面及 n 個反面之機率爲

$$s_n = \binom{2n}{n}(\frac{1}{2})^{2n} 。 \tag{16}$$

當 $n = 1$ 時,$s_1 = 1/2$。當 $n = 5$ 時,s_5=252/1,024,約爲 $1/4 < 1/2$。利用Stirling(1696-1770,蘇格蘭著名的數學家)公式(Stirling formula):

$$n! \sim \sqrt{2\pi}n^{n+1/2}e^{-n}, \tag{17}$$

其中符號"$\sim$"表 $n \to \infty$ 時,兩側之商趨近至1,即所謂近似相等(asymptoticallyequal),可得

$$\binom{2n}{n}(\frac{1}{2})^{2n} = \frac{(2n)!}{(n!)^2 2^{2n}} \sim \frac{\sqrt{2\pi}(2n)^{2n+1/2}e^{-2n}}{(\sqrt{2\pi}n^{n+1/2}e^{-n})^2 2^{2n}} = \frac{1}{\sqrt{\pi n}} 。 \tag{18}$$

由(18)知投擲一公正銅板100次,恰出現50個正面之機率,約爲 $(50\pi)^{-1/2} \doteq 0.0798$,遠小於1/4。

有些初學者以爲丟銅板的次數(假設爲偶數)愈多,愈容易出現各一半正面及反面。現在你知道其實愈不容易。這也牽涉到大數法則(Law of Large Numbers)及中央極限定理(Central Limit Theorem)的區別。有些事件是很容易發生的(如 n 很大時, r_n 很接近1),有些事件是很難發生的(如 n 很大時, s_n 很接近0),由於對機率大小未能正確估算,使我們常做出錯誤的判斷,也分不清何者才是眞正的巧合。

3.非巧合事件。

有些看似巧合的事件,背後其實有其原因(因此並非巧合事件)。由於知識的不足,使我們將很多不是巧合的事件,皆視爲巧合事件。我們先以首位數字現象(first significant digit phenomenon) 爲例。

假若你有位朋友提議就下述事項與你打賭。隨意翻閱某百科全書,任選一頁統計數字,若首位是1,2,3,4的多,則你付他一元,若首位是5,6,7,8,9的多,則他付你1元。你一聽當然對你有利,因9個數字(首位不會是0)你佔了5個。結果翻了幾頁,你都輸了。運氣實在太不好了,你可能會這樣想。

以簡明大英百科全書爲例(台灣中華書局印行,1989年版),附錄有各種統計數字。在面積和人口項,按字母排列,第一頁共有46個國家或地區,面積最小的爲5.6平方哩,最大的爲3,696,100平方哩。其中首位爲1,2,3,4的共有35個(佔 $35/46 \doteq 0.761$),首位爲5,6,7,8,9的共有11個。若改爲平方公里呢,首位爲1,2,3,4的共有31個(佔 $31/46 \doteq 0.674$)。而最近普查的人口,首位爲1,2,3,4的共有32個(佔 $32/46 \doteq 0.696$)。在運輸項的第二頁,有45個國家或地區。在公路長度欄,首位爲1,2,3,4的,以哩計有32個(佔 $32/45 \doteq 0.711$),以公里計有31個(佔 $31/45 \doteq 0.689$)。

這並非偶然,並且是一很早就被注意到的現象,就稱爲首位數字現象。在諸如物理上、化學上、經濟調查等數據,首位數字小於或等於 n 之出現頻率,常約爲 $\log_{10}(n+1)$, $n = 1, 2, \cdots, 9$。Pinkham(1961)也證出此結果。利用此結果,首位數字爲1,2,3,4的約佔 $\log_{10} 5 \doteq 0.699$。檢視我們由百科全書所查出的一些數據,算是很精準的。連費氏數列(Fibonacci sequence) 的首位亦有此現象(見黃文璋(1999)第十五章)。由Pinkham (1961)的結果知,首位爲 n 之出現頻率約爲

$$\log_{10}(n+1) - \log_{10}(n) = \log_{10}(1 + 1/n), \tag{19}$$

隨著 n 之增大而遞減。當 $n = 1$ 時,頻率約爲 $\log_{10} 2 \doteq 0.3010$, $n = 9$ 時,頻率約爲 0.0458,明顯地小很多。

再看一例。在第2及第3節,兩個不同的生日問題中,各有一 $253 = \binom{23}{2}$ 的數字出現,這究竟是不是巧合呢?

由微積中的結果知,當 i 與365相比很小,則下述近似成立:

$$(\frac{364}{365})^i = (1 - \frac{1}{365})^i \doteq 1 - \frac{i}{365}。 \tag{20}$$

利用(20),對 m 個人裡,至少有二人生日相同的機率 $1 - p_m$(見(1)式),可得下述近似:

$$1 - p_m \;=\; 1 - \prod_{i=1}^{m-1} (1 - \frac{i}{365})$$

$$
\begin{aligned}
&\doteq 1-\prod_{i=1}^{m-1}(\frac{364}{365})^{i} \\
&= 1-(\frac{364}{365})^{1+2+\cdots+(m-1)} \\
&= 1-(\frac{364}{365})^{m(m-1)/2} \text{。}
\end{aligned}
$$

另外, (4)式給出 n 個人中,至少有一人生日與你相同的機率 q_n:

$$
q_n = 1-(\frac{364}{365})^{n} \text{。}
$$

因此,當 $m(m-1)/2$ 與 n 很接近時,二機率 $1-p_m$ 與 q_n 近似。反過來也是對的。

現由於23是滿足 $1-p_m > 1/2$ 之最小的正整數 m, 253是滿足 $q_n > 1/2$ 之最小的正整數 n, $1-p_{23}$ 與 q_{253} 當然很接近。因此 $\binom{23}{2} = 23(23-1)/2$ 與253很接近(事實上此處爲相等),並不是一件奇怪的事。

再舉一非巧合的事件。一些研究指出(見Phillips(1972)),有些人可"延遲"其死亡日期,以等待某件他們認爲重大的事。如生日,特別是名人生日,爲一顯著的例子。另外,美國每逢總統大選前,死亡率也常降低(企盼投票)。John Adams (1735-1826) 及Thomas Jefferson (1743-1826),分別爲美國第二位(1797-1801)及第三位(1801-1809)總統。他們皆死於西元1826年7月4日,恰是美國通過獨立宣言(Declaration of Independence)的五十年後。這兩位美國開國元勳皆在美國獨立紀念日那天去世,可能並非巧合。我們引用Jefferson 的醫生所記載Jefferson 最後所講的話做爲佐證:

About seven o'clock of the evening of that day, he (Jefferson) awoke, and seeing my staying at his bedside exclaimed, "Oh Doctor, are you still there?"in a voice however, that was husky and indistinct. He then asked, "Is it the Fourth?"to which I replied, "It soon will be."These were the last words I heard him utter.

4.有趣事件。

打橋牌時,每人分到13張牌,每人會拿到的牌,總共有

$$\binom{52}{13} = 635,013,559,600 \tag{21}$$

種可能性。因此會拿到任何一組13張牌的機率,皆爲635,013,559,600分之一,可說非常地小。任何一手牌,雖然出現的機會均極小,幾乎是不可能,但出現的機會均相同。因此任何一手牌都不該被視爲驚訝的事件(surprising event)。所以當你拿到一手13張皆爲黑桃,你只能說,雖然此手牌出現的機會幾乎是不可能,爲一稀有事件(rare event),但它並非一驚訝的事件,而是一有趣的事件(interesting event)。要知拿到13張黑桃的機會,與拿到紅桃6、J,黑桃4、6、A,方塊3、5、Q、K,梅花5、7 、J 、Q的機會是一樣的。但你大概不會到處告訴人家說"眞神奇!我居然拿到紅桃6、J、⋯"。提醒各位,不要將有趣的事件視爲驚訝的事件。

每個人感到有趣的事件並不相同,每次發牌,你總會拿到某13張牌,它們出現的機會均相同,並不須對某一種組合的出現感到驚訝,雖然任何一組合均爲一稀有事件。一驚訝事件必爲一稀有事件,但卻並非所有稀有事件均爲驚訝事件。有那些才屬驚訝事件呢?前面已指出,丟一公正銅板100次,出現50個正面之機率約爲0.08。但出現95個以上的正面之機率呢?由中央極限定理,令 X 表出現的正面數,則

$$\begin{aligned} P(X \geq 95) &= P\left(\frac{X-50}{\sqrt{100 \cdot 0.5 \cdot 0.5}} \geq \frac{95-50}{\sqrt{100 \cdot 0.5 \cdot 0.5}}\right) \\ &\doteq P(Z \geq 9) \leq P(Z \geq 4) \doteq 0.00003, \end{aligned} \tag{22}$$

其中 Z 有 $\mathcal{N}(0,1)$(標準常態)分佈。之所以只求出 $P(Z \geq 4)$,乃是因一般常態分佈的數值表沒有給出 $z > 4$ 時之 $P(Z \geq z)$。求 $P(Z \geq 4)$ 對應 $P(X \geq 70)$。要出現"70個以上"的正面之機率便已較出現50個正面的機率小很多了,更不要說要出現95個以上的正面了,其機率是微乎其微的。所以對丟100個銅板而言,出現95個以上的正面之事件,不但是一稀有事件,也是一驚訝事件,當然對大部分的人來說,亦爲一有趣事件。底下我們引一段二月河(1997)pp.54-57的故事。

清朝施琅於準備攻打台灣前,拿出一把銅錢,向將士們說“一百枚康熙銅錢,擲向台灣海域圖,倘若我軍出師順利,當有九十五枚以上的字面朝上!”一語既出,將台上下將士們無不失色。結果一擲之下,有九十九枚是“康熙”字面朝上。一霎時,將台上下轟動了,有人直呼“天心厭鄭”,武將們全身的血都在奔湧。李光地心思靈動,陡地一轉念:莫非有九十五枚銅錢是特鑄的兩面字兒?不禁莞爾一笑,卻跟著高呼“萬萬歲”。原來施琅離京前,康熙特賜了一百枚兩面字兒的銅錢,且叫他如此操作。只施琅怕有精明人起疑,特在裡頭換取了五枚,倒使眾人信得更其紮實。

註2.上段中“倘若 $\cdots$ 朝上”一句,此說法並不妥當,讀者可否看出?又熟知歷史掌故的讀者,或許知道上述故事原先流傳的版本,主角爲宋朝名將狄青。在水滸傳的楔子裡, 宋仁宗出生後,晝夜哭啼不止。後來有一老叟在他耳邊低低說了“文有文曲,武有武曲,”八個字,他便不哭了。原來玉皇大帝派了兩座星辰下來輔佐他。文曲星爲包拯,武曲星便是狄青。狄青在征討濃智高前, “取百錢自持之,且與神約果大捷,則投此期盡錢面也”。在萬人注視下,他“揮手倏一擲,則百錢盡面矣”。這是出自宋朝蔡絛撰的“鐵圍山叢談”一書第二卷。其改寫版見之於不少給兒童看的書中,或見陳雨明(1997)pp.224-225。如今二月河所寫的故事,有可能是他(或康熙)模仿來的。不過他也做了一些修改,如改爲丟爲後有一字面朝下, 使眾人信得更紮實。

註3.令 $\phi(x)$ 及 $\Phi(x)$ 分別表 $\mathcal{N}(0,1)$ 分佈之機率密度函數及分佈函數,即

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ \Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du,\ x\in R\text{。}$$

則

$$\phi(x)/(x+x^{-1})<1-\Phi(x)<\phi(x)/x,\ x>0, \tag{23}$$

且 $x\to\infty$ 時,

$$1-\Phi(x)\sim\phi(x)/x\text{。} \tag{24}$$

證明並不難,可參考黃文璋(1994)p.304。利用(23)即得

$$P(Z \geq 9) = 1 - \Phi(9) < \phi(9)/9 = \frac{1}{9\sqrt{2\pi}} e^{-81/2} \doteq 1.028 \cdot 10^{-18},$$

可說非常地小(二月河應是不曉得此機率值,他只知道很小)。所以丟一公正銅板100次,要得95個以上的正面,幾乎是不可能(得到100個正面之機率則爲 $0.5^{100} \doteq 7.888 \cdot 10^{-31}$)。又同理求出 $P(Z \geq 4) < 0.000037$,而查表所得爲0.00003。當無法由查表得到$\Phi(x)$(或$1-\Phi(x)$),便可藉助(23)及(24)得到近似值。

6 當稀有碰到大數

機率裡有所謂確實大的數之法則(The Law of Truly Large Numbers)。一件事雖發生的機率很低,但若樣本數夠大,便有可能發生了。以橋牌裡拿到13張黑桃的事件而言,雖其發生的機率很低(僅約 $1.57 \cdot 10^{-12}$),但想想每天全世界有多少人在玩橋牌,每次玩多少副牌,一年下來13張黑桃的出現並不足爲奇。又如若某件巧合,在每個人身上每天會發生的機率爲百萬分之一,台灣有兩千萬以上的人口,因此每天共發生的次數之期望值便超過20次,一年便超過七百次。所以當稀有事件碰到大樣本,其發生便有可能處處可見了。此道理看似簡單,但眞正了解的人可能並不多。見下例。

例4. 民國83年10月15日聯合報繽紛版有幅漫畫,其中文字說明爲"根據指紋學家的研究,六百四十億個指紋才會出現一次相同的指紋;若以現在世界五十億人口,每個人有十個指紋來計算,世界上的指紋總數是五百億個,所以還是找不到指紋相同的人"。

我們對指紋的辨識毫不了解,猜想其含義爲指紋有些特徵,若將任兩指紋拿來比對,會被判定相同之機率爲 $640 \cdot 10^8$ 分之一。只是難道小拇指也可拿來與大拇指比對嗎?我們姑且就相信此報導好了。依其報導,任兩指紋不吻合之機率爲

$$1 - \frac{1}{640 \cdot 10^8},$$

而500億個指紋,任取兩個來比對,共有

$$k = \binom{500 \cdot 10^8}{2} \doteq 1.2499 \cdot 10^{21}$$

對。因此兩兩指紋比對,皆不相同的機率爲

$$(1 - \frac{1}{640 \cdot 10^8})^k \text{。}$$

而會找到一對相同之機率爲1減去上述機率值。利用微積分裡的結果:當 $|a_n|$ 很小,而 b_n 很大,則

$$(1 + a_n)^{b_n} \doteq e^{a_n b_n}, \tag{25}$$

只要 $\{a_n b_n\}$ 不會趨近至 ∞,否則此時 $(1 + a_n)^{b_n}$ 的值會很大。利用(25)即得

$$(1 - \frac{1}{640 \cdot 10^8})^k \doteq e^{-1.95312 \cdot 10^{10}},$$

此值微乎其微幾乎是零了(以桌上型計算機可得 $e^{-226} \doteq 7.07 \cdot 10^{-99}$,再高的次方就得不到了)。因此會找到一對指紋相同之機率極接近1。一件幾乎確定會發生的事件,居然被以爲不會發生,相差眞是不可以道里計。這是比將事件發生所需實驗次數之期望值,視爲所需次數更離譜者。

我們亦可求出,若將500億個指紋,分別拿來跟某一特定的指紋比對,則會有相同的機率爲

$$1 - (1 - \frac{1}{640 \cdot 10^8})^{500 \cdot 10^8} \doteq 1 - e^{-50/64} \doteq 0.5422,$$

大於1/2,也是極易發生的。

有些門禁是以指紋辨識(這是電影裡常有的情節),預存合法可進入的指紋,雖每一指紋會被判定相同的機率很低($640 \cdot 10^8$ 分之一),但若指紋數很多(500億個),則會有被誤判相同的時候就不難發生了(機率約0.5422)。附帶一提,根據民國88年10月13日中國時報第13版之報導,聯合國人口統計學家推算,全球人口在10月12日達到六十億。他們是根據生育率、死亡率及人口流動率推算出來的。

註4.在台灣究竟可不可以做國民卡的爭論中,新新聞週報第606期(1998年10月18日－10月24日) 專訪中華國民卡團隊總經理張公僕。他指出"不是你本人的指紋卻被誤認爲是你的之機率(錯誤接受率),現在的數據是機率必須小於十萬分之一,是相當安全的。而是你本人卻被誤認爲不是你的(錯誤拒絕率),目前要求你按兩次而被拒絕的機率必須要小於千分之一。"可見現今指紋的辨識儀器,並沒有那麼精準。另外,依據報導(88年1月8日聯合報第49版,記者黃裕元),星友科技宣稱其所開發出之FG-60指紋辨識器,符合FBI國際標準,且其錯誤接受率及錯誤拒絕率,與前述數據近似。

例5.美國紐約時報(New York Times)曾在第一版報導(西元1986年2月14日),一位名叫做Adams (Evelyn Marie Adams)的女士第二度贏得紐澤西(New Jersey)州的樂透獎(lottery)。她第一次獲三百九十萬美元,第二次獲一百五十萬美元。紐約時報說這是十七兆分之一(1 in 17 trillion)的機會。

十七兆分之一其實是另一問題的答案,若你有兩次各買一張紐澤西州的樂透獎券,則兩次皆獲獎的機率便約爲十七兆分之一(Adams女士其實常買彩券,且每次買好幾張)。

事實上此問題從某個角度看,乃是類似生日問題及指紋問題。人們眞正想知道的是:在美國千萬買樂透獎券的人裡,有人一生中,曾贏兩次的機會。不要忘了不少人每次都買許多張。

美國普渡大學(Purdue University)統計系的兩位教授Stephen Samuels及George McCabe,於前述報導刊登後兩週,亦在紐約時報登出他們的投書。他們指出雖然Adams女士在一生中要贏兩次樂透獎的機會很小,但在全美中有人贏得兩次,可說是一件必然的事(practically a sure thing)。他們估計7年內有人贏得兩次的機會超過二分之一。他們也估計四個月(Adams女士贏兩次之間距)內有人贏兩次之機會超過三十分之一。他們的預測並不離譜,於西元1998年5月, Robert Humphries 第二度贏得賓夕凡尼亞(Pennsylvania)州樂透獎(合計有六百八十萬美元)。

千萬不要低估大數的威力。

習　題

1. 自一副標準的52張樸克牌中,隨機地取5張牌,求其中至少有一張牌爲8點之機率,並求至小數4位之近似值。

2. 設1,000張彩券中有5張有獎,某人買10張彩券,求其中至少有一張中獎之機率,並求至小數4位之近似值。

3. 要利用機器將 n 個人的薪資單放進 n 封貼有名字的信封,由於作業錯誤,變成隨機地將薪資單放進信封,每個信封恰放進一個。求至少有一個信封放的是正確的薪資單之機率。註:此稱爲Montmort問題(Montmort's Problem)。

4. 令 A 表投擲4個公正的骰子一次,其中至少有一個1點的事件,B 表投擲2個公正的骰子24次,其中至少有一次得到兩個1點的事件。問何者機率較大? 註:此即著名的de Méré詭論(paradox)。在西元1654年,de Méré向Pascal請教此問題之解。

5. 問一團體中,除你之外,要有多少人,才會使有人生日與你相同的機率大於0.9?

6. (i).求夫妻二人生日相同之機率(假設一年有365天);
(ii).求最少要有多少對夫妻,才會使其中至少有一對夫妻生日相同之機率大於1/2。

7. 假設每月皆有30天,求一三口之家生日之日期皆相同之機率。

8. 假設美國大學的百支籃球隊中,有30隻球隊每場比賽獲勝之機率均超過0.8。則在某一球季,有球隊自開打後,連勝20場之機率,至少爲若干?

9. 設全世界每人每天各丟一公正銅板100次。試求平均約要多久,才會出現有人某天得到95個以上的正面。

10. 試指出註2中提及的"倘若 $\cdots$ 朝上"一句,此說法的問題。

11. 在例4中,利用二項分佈趨近至Poisson分佈,估計若將500億個指紋,分別拿來跟某一特定的指紋比對,會找到有(i)1個,或(ii)2個以上指紋相同的機率。

12. 假設任兩指紋相同的機率爲 $5 \cdot 10^{10}$ 分之一。全世界設有60億人,每人各有10個指紋,求會找到有一指紋與你任一指紋相同之機率,並求其近似值。

13. 假設指紋之錯誤接受率爲十萬分之一。則在兩千萬人中,若皆辨識大拇指,將約有多少人會被判定與你指紋相同?

14. 舉出一件你曾遇過之很巧合的事情,並嘗試說明其發生的原因。

15. 不少人都常有下述經驗:要找某件物品,卻偏找不到,過了一段時間,都已經不找了,該物卻突然出現了。正是"踏破鐵鞋無覓處,得來全不費工夫"。試釋斯事。

參考文獻

1. 二月河(1994).雍正皇帝—雕弓天狼<上>。巴比倫出版社,台北。

2. 二月河(1997).康熙大帝—玉宇呈祥<下>。巴比倫出版社,台北。

3. 余秋雨(1995).山居筆記。爾雅出版社有限公司,台北。

4. 邱燮友註譯(1996).新譯唐詩三百首,修訂九版。三民書局股份有限公司,台北。

5. 陳雨明(1997).歷代人物外傳。漢欣文化事業有限公司,台北縣。

6. 黃文璋(1994).機率論講義。全華科技圖書有限公司,台北。

7. 黃文璋(1999).數學欣賞。華泰文化事業股份有限公司,台北。

8. Abramson, M. and Moser, W.O.J.(1970). More birthday surprise. *American Mathematical Monthly* **77**, 856-858.

9. Apostol, T.M.(1974). *Mathematical Analysis*, 2nd ed. Addison-Wesley , Reading, Massachusetts.

10. Cohen, A.(1983). Seasonal daily effect on the number of births in Israel. *Journal of the Royal Statistical Society Series C: Applied Statistics* **32**, 228-235.

11. Diaconis, P. and Mosteller, F.(1989). Methods for studying coincidences. *Journal of the American Statistical Association* **84**, 853-861.

12. Fraser, S.(1999).眞是巧!哪能眞是那麼巧?一定有點道理在。讀者文摘,1999年三月號,13-16。

13. Nixon, C.(1998/9). Mathematics in the classroom. *Mathematical Spectrum* **31**, 17-18.

14. Phillips, D.P.(1972). Deathday and birthday: an unexpected connection. In *Statistics:A Guide to the Unknown* (Edited by J.M. Tanur and by J. Mosteller, W.H. Kruskal, R.F. Link, R.S. Pieters and G.R. Rising). Holden-Day, Inc., San Francisco.

15. Pinkham, R.S.(1961). On the distribution of first significant digits. *Annals of Mathematical Statistics* **32**, 1223-1230.

16. Stewart, I.(1998). What a coincidence. *Scientific American* **278**, No.6, 76-77.

17. Weaver, W.(1982). *Lady Luck: The Theory of Probability.* Dover Publications, Inc., New York.