

所期爲何

黃文璋教授

國立高雄大學應用數學系

1 期望值

丟一均勻的骰子,也就是1, 2, 3, 4, 5, 6 每個面出現的機率皆爲 $1/6$ 。令 X 表所得之點數,則 X 之期望值(expectation, expected value,或稱mean),以 $E(X)$ 表之,且

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5。$$

即3.5爲丟骰子所得點數之期望值。這是很簡易的計算,學過機率統計的人都知道期望值,通常也不認爲計算有太大困難。但期望值的“值”,往往令人感到迷惑,與我們平常對“期望”的了解並不相同。

“期待他日再相逢,共度白首”,這是鳳飛飛“流水年華”中的一句歌詞。此處所期的爲再相逢,而這是有可能達到的期待。又如唐詩三百首中,元稹遣悲懷三首中的其三,“同穴窅冥何所望,他生緣會更難期”。期、望皆出現了。這裡的期及望,也都是所欲達到的。

但丟骰子,點數之期望值爲3.5,而點數爲整數,無論怎麼丟,都得不到3.5,那怎會稱3.5爲所期望之值呢?前述詩及歌中之期望雖難達到,但總是有可能達到,而丟骰子之期望值,卻爲一絕不可能達到之值,那爲何將其視爲期望值呢?

設 X 爲一離散型的隨機變數,且可能取值 x_1, x_2, \dots ,而對應之機率分別爲 $f(x_1), f(x_2), \dots$ (f 稱爲 X 之機率密度函數(probability density function,簡稱p.d.f.))。則 X 之期望值定義爲

$$(1) \quad E(X) = x_1 f(x_1) + x_2 f(x_2) + \dots = \sum_{k=1}^{\infty} x_k f(x_k),$$

只要上述級數收斂。若 $\sum_{k=1}^{\infty} x_k f(x_k)$ 發散,則稱 X 之期望值不存在。如果 X 只取值在一有限的實數集合 $\{x_1, x_2, \dots, x_n\}$,則 X 之期望值當然存在。當 $\sum_{k=1}^{\infty} x_k f(x_k) = \infty$ (或 $-\infty$) 時(此時期望值不存在),習慣上以 $E(X) = \infty$ (或 $E(X) = -\infty$) 表之,並稱此時期望值為無限大(或負無限大)。若 X 為一連續型的隨機變數,且以 f 為 p.d.f.,則 X 之期望值定義為

$$(2) \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

只要此積分存在。

丟一骰子大約會出現多少點?買一張彩券大約會中多少錢?人的壽命大約是多少?得某種病後大約尚可活多少年?

對一隨機現象,我們常想粗略地知道其值究竟多大?期望值就是常被拿來扮演這種以一單一的值,來代表一隨機現象中之變數大小的角色。

考完試,想知道某班究竟考得好不好?怎麼評斷呢?通常的作法是算出平均分數。期望值與平均數(average,簡稱平均)的關係密切,我們稍後再說明。中國人跟美國人誰比較高?這怎麼比呢?美國人不乏矮個子,中國人也不乏高個子。算出平均身高是一種方法。平均值常被拿來當做某一群體之代表值。但也不全是如此。例如,若問中國人的網球跟日本人的網球孰強?在這種情況下,很可能是兩國各挑五位高手比賽,平均值在此毫無用途。又如某生就讀外地的大學,父親問他每月需多少錢才夠用?該生很可能不會提出各月需求的平均值,而是提出各月需求的極大值,以保證夠用。

對一大堆數據,我們常只想以一值來代表它們,這是人的習性。正如說某校校風很好,客家人能吃苦等。某校學生數千,每位學生行為不一,但長期下來,大致給人不錯的印象,校風很好的美譽遂冠上了。人們習於用一形容詞,來描述一大群個體的行為。

對一隨機變數,期望值為一代表它的值。此隨機變數所可能取的值散佈在期望值的左右,期望值像是隨機變數之分佈的一個核心一樣。其他常被拿來當做隨機變數之代表值的尚有中位數(median)及眾數(mode)。一實數 a 若滿足 $P(X \leq a) \geq 1/2$ 且 $P(X \geq a) \geq 1/2$,便稱為隨機變數 X 之一中位數,或母體中位數(population median)。中位數不一定唯一。若

$f(a)$ 為 X 之 p.d.f. f 的一極大值,則 a 稱為 X 之一眾數,眾數也不一定唯一。

為什麼會有這麼多不同的代表值,而又該採用那一個呢?事實上,視情況及需求之別,期望值、中位數及眾數各有其舞台。由於大數法則(Law of Large Numbers),凸顯期望值的重要性。又因具有許多好的性質,期望值因此最常被拿來當做隨機變數之代表值。另外,有時我們會對一分佈之中間位置有興趣,也就是在此值之左及右的的機率各約為 $1/2$ 。若有樣本 X_1, X_2, \dots, X_n , 且按小至大排列,則樣本中位數(sample median)為其位居中間大的那一個(如果 n 是奇數),或是兩個中間的值的平均(如果 n 是偶數)。如果資料中有所謂離群值(outlier),樣本中位數較不會起伏很大,此時樣本中位數較適合當做資料的中心。這是樣本中位數的優點。譬如說,有1至9及100等10個數,則其平均為14.5,而中位數為5.5。在有離群值100的情況下,以5.5來當做這些數字的一代表值,顯然是較恰當的。其他如在一支職業籃球隊中,若有一、兩位年薪是天價的球員,便可能將全隊平均年薪拉得很高,這時中位數亦較適合用來衡量全隊球員薪水高低。

對隨機變數之代表值,有一些不同的指標是很好的,可比較在採用不同的指標下,結果之差異,而也因此使統計理論變得更豐富。此正如武俠小說裡,因有各派互有高下之武功,而讓小說顯得更有趣。

2 期望值與平均

假設有1,2,1,3,2,1等6個數,則其平均為何?當然是

$$\frac{1+2+1+3+2+1}{6} = \frac{10}{6}。$$

隨後發現此6數中,有3個1, 2個2,1個3,因此6數和等於 $1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1$,而其平均可寫成

$$\frac{1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1}{6} = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6}。$$

原來此6數之平均可視為1,2,3之一加權平均數(weighted average,簡稱加權平均)。1所佔之份量為全部的 $3/6$, 2佔 $2/6$, 3佔 $1/6$ 。 $3/6$, $2/6$ 及 $1/6$

分別稱為1,2及3之權數(weight)。由此角度看,期望值便類如加權平均:對一離散型的隨機變數 X ,將 X 可能取的值與其機率相乘後,再全部加起來(見(1)式),便成為 X 的期望值。

假設你參與一項賭局:莊家丟一骰子(就當做是公正的),得到幾點就給你多少元。試問每次賭之前,你須付莊家若干元,你與莊家雙方才會皆覺得公平?

只賭一次不易感覺怎樣才是公平。設賭局可一直持續進行,且以 X_1 表第一次所得點數, X_2 表第二次所得點數, \dots , X_n 表第 n 次所得點數。則由大數法則知, n 次的平均 $(X_1 + X_2 + \dots + X_n)/n$,隨著 n 之增大,會以某種方式接近 $E(X_1) = 3.5$ 。換句話說,賭 n 次後,你的平均所得,只要 n 很大,會以某種方式接近3.5。因此,雙方大概皆會同意,每次須付3.5元是合理的。

對丟骰子而言,雖然沒有一次會得到3.5,但若丟很多次,所得點數之平均,會“接近”3.5。故期望值可視為乃是許多次觀測值平均的期望,而非一次觀測值的期望。也就是觀測一現象許多次後,其平均為何?我們會“期望”該平均“大約是”其期望值。期望值與平均再度連上關係。也就是在大樣本(large sample)下,期望值與樣本平均(sample mean) $\sum_{i=1}^n X_i/n$ 會“很接近”,即

$$(3) \quad E(X) \doteq \frac{X_1 + X_2 + \dots + X_n}{n}。$$

只是此處我們只很含混地說很接近,尚不擬說明其意義。

就是因有(3)式,期望值遂有平均的味道(大數法則也因此又被稱做平均數定律(Law of Average)),於是期望值便常被用來當做隨機變數之代表值。而只要樣本數夠大,樣本平均也常被拿來當做期望值之一估計值(estimate)。我們常要估計一些未知的量。譬如說,到底銅板出現正面的機率 p 為何?這是一個天才曉得的值,我們人呢?只能估計其值。好的估計法,通常要用到統計,我們稍後再談。

期望值的這種加權平均的概念在物理或工程上亦常出現。例如,設有一長度為 $b - a$ 之直的鐵棒。將鐵棒之左端點置於 x 軸之 a 點,右端點之位置在 b 。又令 $\rho(x)$ 表鐵棒之密度函數。因此 $\int_a^b \rho(x)dx$ 表此鐵棒之總

重,而

$$\frac{\int_a^b x\rho(x)dx}{\int_a^b \rho(x)dx}$$

表鐵棒在 x 軸之重心位置。對一連續型的隨機變數,以 $f(x)$ 為其 p.d.f., 因 $\int_{-\infty}^{\infty} f(x)dx = 1$ (即對應上式之分母為 1), 所以重心、期望值及加權平均, 是類似的概念。

最後對一 p.d.f. 為 f 之隨機變數 X , 亦可考慮其一變數代換 $g(X)$ 之期望值, 且有

$$(4) \quad E(g(X)) = \begin{cases} \sum_x g(x)f(x) & , \text{若 } X \text{ 為離散型,} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & , \text{若 } X \text{ 為連續型。} \end{cases}$$

至於若有二隨機變數 X, Y , 以 $f(x, y)$ 為其聯合 p.d.f., 則

$$(5) E(g(X, Y)) = \begin{cases} \sum_{x,y} g(x, y)f(x, y) & , \text{若 } (X, Y) \text{ 為離散型,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy & , \text{若 } (X, Y) \text{ 為連續型。} \end{cases}$$

3 變異數及其他

上節提到期望值為隨機變數之一代表值。令 X 表丟一公正骰子所可能得到的值, 則 $E(X) = 3.5$ 。雖名之為代表值, 但一丟後, 1, 2, 3, 4, 5, 6 皆可能得到, 偏差 (deviation) 是存在的。偏差究竟有多大呢? $X - E(X)$ 為其偏差, 此值有正有負, 但其期望值

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0。$$

也就是平均而言, 偏差之期望值為 0。此亦為期望值之一性質。但我們往往是對偏差之絕對值較感興趣。此正如假設某人射飛鏢, 有時上偏有時下偏, 他總不能自鳴得意地說平均會命中紅心。

$|X - E(X)|$ 表偏差之絕對值。此量當然還是一隨機變數。注意 $E(X)$ 則已非隨機了。此偏差之絕對值又如何衡量其大小呢? 對了! 取期望值, 你

應想得到了。也就是看偏差之絕對值平均而言有多大。問題是,如果你微積分學得還不錯的話,應知道有關絕對值函數之求和或積分是較麻煩的。對於

$$E(|X - E(X)|),$$

要討論何時 $X \geq E(X)$, 何時 $X < E(X)$ 。因此退而求其次,我們考慮偏差平方之期望值,並以 $\text{Var}(X)$ 表之,即

$$(6) \quad \text{Var}(X) = E((X - E(X))^2)。$$

$\text{Var}(X)$ 稱為 X 之變異數(variance)。由於先將偏差平方,像是距離的平方,因此若將變異數開方,將使其回到像一距離函數。變異數之正的平方根,稱為 X 之標準離差(standard deviation),或稱標準差或均方差。變異、離差,由字眼上便已可了解其含義。

我們平常說,雖不中亦不遠矣。也就是說,即使未命中,誤差也不太大。但不遠到底是多遠?標準離差的功能就是對誤差的大小,提供一參考值。如果說期望值是描述一隨機變數之分佈的“集中處”,變異數便是描述分佈關於期望值之偏差程度。

例如,設有兩組學生,每組各有三人。考完試後兩組平均成績皆為60分。但第一組之成績為61,60,59,第二組之成績為90,60,30。第一組學生的程度接近,第二組學生之程度,則有很大差異。光由平均成績之大小,是無法了解兩組學生程度之差異情況。這也是對隨機變數,為何除了期望值,我們尚需要給出其變異數。給了此二值,對其“中心”及“散佈範圍”,才能有一較清晰的概念。我們從必然的世界走入隨機的世界,卻又想以較簡單的方式,來看每一隨機現象。期望值及變異數就是如此因應而生。

(6)式粗略地告訴我們:若變異數較小,則隨機變數會取值在期望值附近之機率較大,反之則會偏離期望值較遠的機率不小。由於 $(X - E(X))^2 \geq 0$,故變異數之最小可能的值為0,而此只有當 $X = E(X)$ 之機率為1才會發生。此時 X 其實是一退化的隨機變數,可稱之為常數了。除此情況之外,變異數均為正。

統計學家常會對隨機變數會落在離其期望值不超過1個,2個或3個標準差之機率感到興趣,可參考底下的例4。

如果是產品的品質管制,當然希望變異數愈小愈好。例如,若 X 表某型螺絲釘的直徑,則 X 之變異數最好小至0,而期望值最好等於所宣稱的規格。但對某一科目學生的考試成績,其變異數便要夠大,否則該科之鑑別率便不高。題目太簡單或太難,都可能使成績的變異數很低。又如對教育政策而言,限制太多,管理嚴格,學生的學習可能多半能達到一定的效果,但學生往往也不易自由發揮,亦即學生的表現,變異會較小;而較開放的政策,有些學生的學習效果可能會很差,但天資聰穎的學生,卻較能自由地學習,而成就非凡,在此環境下,學生表現的變異會較大。兩種政策各有其優缺點,不同性向的學生適合不同的環境。有些管理者喜歡掌握一切,希望事情的發展是他們所預期的,不要有太大的變異。多少創意因此被封殺,多少天才因此被埋沒。如果管理者眼光正確還好,如果是庸才,結局就很糟糕。

對一隨機變數 X ,常也以 μ_X (或 μ)表其期望值,以 σ_X^2 (或 σ^2)表其變異數。又

$$\begin{aligned}\text{Var}(X) &= E(X - \mu)^2 \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2,\end{aligned}$$

換句話說

$$(7) \quad \text{Var}(X) = E(X^2) - E^2(X)。$$

此處用到下述期望值的兩個性質:

$$(8) \quad E(\alpha) = \alpha,$$

$$(9) \quad E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y),$$

其中 α, β 為二實常數。

變異數亦有下述簡單的性質：

$$(10) \quad \text{Var}(aX + b) = a^2 \text{Var}(X),$$

其中 a, b 為二實常數。(8)-(10)除了藉由數學來證明,讀者不妨試由期望值及變異數的意義,來說明為何成立。

對一隨機變數 X , 假設我們想以一常數 a 估計它。這個估計好不好呢? 再度我們考慮誤差的平方 $(X - a)^2$ 。此值當然愈小愈好, 但因其為一隨機變數, 於是自然地取其期望值。也就是 $E(X - a)^2$ 要愈小愈好。我們有下述定理。

定理1. 設隨機變數 X 滿足 $E(X^2) < \infty$ 。則對 $a \in R$, $E(X - a)^2$ 之極小值發生在 $a = E(X)$ 。

證明. 由

$$E(X - a)^2 = E(X^2) - 2aE(X) + a^2,$$

視此為一 a 之二次式, 則易得極小值發生在 $a = E(X)$ 。

上定理指出, 若要以一常數來猜一隨機變數 X 的觀測值, 則不妨以期望值 $E(X)$ 來猜。此時誤差平方之期望值最小, 且此期望值恰為 X 之變異數。故 $E(X)$ 可視為隨機變數 X 之“最佳”常數估計值, 而 $\text{Var}(X)$ 表此估計之精確度。譬如說, 要猜丟一公正的骰子會得多少點, 則 3.5 是一“最佳”的選擇。

期望值及變異數尚有不少性質, 可查黃文璋(1994)一書。我們列出下述關於獨立隨機變數之結果。

定理2. 設 X_1, X_2, \dots, X_n 為獨立的隨機變數。則只要 $E(X_i)$ 皆存在, $\forall i = 1, 2, \dots, n$,

$$(11) \quad E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i);$$

只要 $\text{Var}(X_i)$ 皆存在, $\forall i = 1, 2, \dots, n$,

$$(12) \quad \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)。$$

(12)式給出獨立的隨機變數之和的變異數等於變異數之和。直觀上這是對的。至於和的期望值,由(9)式,不須隨機變數獨立的假設,便會等於期望值之和,即

$$(13) \quad E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i),$$

只要 $E(X_i)$ 皆存在, $\forall i = 1, 2, \dots, n$ 。除了型式簡單,及在理論上有用之外,(13)式使求隨機變數之和的期望值,變成非常方便。不須知道 X_1, X_2, \dots, X_n 之聯合分佈,而只要知道 X_1, X_2, \dots, X_n 之邊際分佈(見(5)式),計算上簡潔很多。

二隨機變數之和的變異數,有下述表示法:

$$(14) \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2E((X - E(X))(Y - E(Y))),$$

$\text{Var}(X + Y)$ 與 $\text{Var}(X) + \text{Var}(Y)$ 不一定何者較大。如若 $Y = X$,則

$$\text{Var}(X + Y) = \text{Var}(2X) = 4\text{Var}(X) \geq 2\text{Var}(X) = \text{Var}(X) + \text{Var}(Y),$$

且等號成立若且唯若 $\text{Var}(X) = 0$,即 X 為一常數。又若 $Y = -X$,則

$$\text{Var}(X + Y) = \text{Var}(0) = 0 \leq 2\text{Var}(X) = \text{Var}(X) + \text{Var}(Y)。$$

我們以 $\text{Cov}(X, Y)$ 表 X, Y 之共變異數(covariance),其定義為

$$(15) \quad \begin{aligned} \text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y)。 \end{aligned}$$

有時以 σ_{XY} 表 $\text{Cov}(X, Y)$ 。易見 $\text{Cov}(X, X) = \text{Var}(X)$,所以共變異數為變異數之推廣。若 $\text{Cov}(X, Y) = 0$,則稱 X 與 Y 為無相關(uncorrelated)。由定理2知,二獨立的隨機變數必為無相關(因此時 $E(XY) = E(X)E(Y)$)。其逆不真,見稍後的例7中。

設 X, Y 為二變異數皆存在且不為0之隨機變數,則常以相關係數(correlation coefficient) $\rho(X, Y)$ 來描述其相依程度。定義為

$$(16) \quad \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}。$$

可看出 $\rho(X, Y) = 0$, 若且唯若 X 與 Y 為無相關。相關係數之意義,我們留至第6節再說明。

對 $\forall n \geq 1$, 若 X_1, X_2, \dots, X_n 之變異數皆存在,則可得到

$$(17) \quad \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j)。$$

對每一正整數,隨機變數 X 之 m 次動差(moment)之定義為 $E(X^m)$ (不一定存在,但高次動差若存在,則較低次動差亦存在)。期望值便是一次動差。你可以說這是人類的習性,優游於由繁至簡及由簡至繁的樂趣中。對隨機變數,於給出一次動差及二次動差(變異數是基於二次動差)後,對其分佈有了基本的了解,但尚嫌不足,繼續考慮高次動差。因一隨機變數之動差,對描述其分佈有很大的幫助,知道愈多次動差,對其分佈便愈了解。有些分佈也可由其所有動差而唯一決定,見黃文璋(1994)第四章定理2.5及p.225。

最後對一隨機變數 X ,若只取非負整數值,且 $P(X = k) = a_k, k = 0, 1, \dots$,我們常考慮其母函數

$$(18) \quad g(s) = E(s^X) = \sum_{k=0}^{\infty} a_k s^k, |s| \leq 1。$$

若 X 為非負的隨機變數,則往往考慮其拉普拉斯轉換(Laplace transformation)

$$(19) \quad \phi(s) = E(e^{-sX}), s \geq 0。$$

若 X 為連續型,有p.d.f. f ,則

$$(20) \quad \phi(s) = \int_0^{\infty} e^{-sx} f(x) dx;$$

若 X 為離散型,取值在 $\{\lambda_0, \lambda_1, \dots\}$,則

$$(21) \quad \phi(s) = \sum_{k=0}^{\infty} p_k e^{-s\lambda_k},$$

其中 $p_k = P(X = \lambda_k)$ 。

這些轉換都具有一以貫之的功能,能以一扼要的方式,來包含關於隨機變數 X 之所有資訊。由這些轉換也可得到 X 之各次動差。至於對一取正亦取負值的隨機變數,則可考慮其特徵函數(characteristic function)

$$(22) \quad \psi(t) = E(e^{itX}), \quad t \in R,$$

其中 $i = \sqrt{-1}$ 。動差母函數(moment generating function)也是一常被考慮的轉換,其定義為

$$(23) \quad M(t) = E(e^{tX}),$$

$\forall t \in R$,只要此期望值存在。只是對某些隨機變數,有可能對 $\forall t \neq 0$, $M(t)$ 皆不存在。

4 例子

我們給幾個簡單的例子。

例1.設 X 有 $P(\lambda)$ 分佈,即 X 之p.d.f.為

$$(24) \quad f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots。$$

則 X 之母函數為

$$\begin{aligned} g(s) &= E(s^X) = \sum_{x=0}^{\infty} f(x) s^x = \sum_{x=0}^{\infty} \frac{e^{-\lambda} (\lambda s)^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda s)^x}{x!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}, \quad |s| \leq 1。 \end{aligned}$$

此處用到對 $\forall a \in R$,

$$\sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a。$$

又可求出

$$(25) \quad E(X) = \text{Var}(X) = \lambda。$$

例2. 設 X 有 $\text{Ber}(\theta)$ 分佈, 即 X 之 p.d.f. 為

$$(26) \quad f(x) = \theta^x(1 - \theta)^{1-x}, x = 0, 1。$$

則 X 之母函數為 $g(s) = E(s^X) = 1 - \theta + \theta s, |s| \leq 1$ 。又可求出

$$(27) \quad E(X) = \theta, \text{Var}(X) = \theta(1 - \theta)。$$

例3. 設 X 有 $\mathcal{B}(n, \theta)$ 分佈, 即 X 之 p.d.f. 為

$$(28) \quad f(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, x = 0, 1, \dots, n,$$

$0 < \theta < 1$ 。則 X 之母函數為

$$\begin{aligned} g(s) &= E(s^X) = \sum_{x=0}^n s^x \binom{n}{x} \theta^x (1 - \theta)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (\theta s)^x (1 - \theta)^{n-x} \\ &= (1 - \theta + \theta s)^n, |s| \leq 1。 \end{aligned}$$

又可求出(可直接求,或先寫成 $X = X_1 + X_2 + \dots + X_n$, 其中 $X_i, i = 1, 2, \dots, n$, 為相互獨立, 且每一皆有 $\text{Ber}(\theta)$ 分佈, 因此 $E(X_i) = \theta, \text{Var}(X_i) = \theta(1 - \theta)$, 再利用(12)及(13))得

$$(29) \quad E(X) = n\theta, \text{Var}(X) = n\theta(1 - \theta)。$$

例4. 設隨機變數 X 有 $\Gamma(\alpha, \beta)$ 分佈, 即 X 之 p.d.f. 為

$$(30) \quad f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha) \beta^\alpha}, x > 0。$$

在此 gamma 函數(gamma function) $\Gamma(u)$ 之定義為

$$(31) \quad \Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx, u > 0。$$

則對 $\forall k \geq 1$,

$$\begin{aligned} (32) \quad E(X^k) &= \int_0^{\infty} x^k f(x) dx \\ &= \frac{\beta^k}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha+k-1} e^{-y} dy \\ &= \frac{\beta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}, \end{aligned}$$

其中利用到變數代換令 $y = x/\beta$ 。由(32)得

$$(33) \quad E(X) = \alpha\beta, \quad E(X^2) = \alpha(\alpha+1)\beta^2,$$

因此

$$(34) \quad \text{Var}(X) = E(X^2) - E^2(X) = \alpha\beta^2。$$

又 X 之拉普拉斯轉換為

$$\begin{aligned} (35) \quad \phi(s) &= \int_0^{\infty} e^{-sx} f(x) dx = \int_0^{\infty} \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-sx} x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} x^{\alpha-1} e^{-(s+1/\beta)x} dx = \frac{1}{(1+\beta s)^\alpha}, \quad s \geq 0。 \end{aligned}$$

註1. Gamma函數滿足

$$(36) \quad \begin{aligned} \Gamma(\alpha+1) &= \alpha\Gamma(\alpha), \quad \alpha > 0, \\ \Gamma(1) &= 1, \Gamma(n+1) = n!, \quad n \text{ 爲一非負整數。} \end{aligned}$$

例5. 設隨機變數 X 有 $\mathcal{N}(0,1)$ (標準常態(standard normal))分佈, 即 X 之 p.d.f. 為

$$(37) \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty。$$

則由微積分裡的結果知, 對 $\forall k \geq 1$, $\int_{-\infty}^{\infty} x^k e^{-x^2/2} dx$ 收斂。故 $E(X^k)$ 存在, $\forall k \geq 1$ 。又利用分部積分(integration by parts), 可得

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx = \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{k-2} e^{-x^2/2} dx, \quad \forall k \geq 2。$$

即

$$E(X^k) = (k-1)E(X^{k-2}), \forall k \geq 2。$$

又由於 $E(X) = 0$ (因 f 為一偶函數 (even function), 即 $f(x) = f(-x), \forall x \in R$), 且 $E(X^0) = E(1) = 1$, 故

$$(38) \quad E(X^{2k}) = \prod_{i=1}^k (2i-1), \forall k \geq 1,$$

且

$$(39) \quad E(X^{2k+1}) = 0, \forall k \geq 0。$$

由此得

$$\text{Var}(X) = E(X^2) - E^2(X) = 1。$$

設 Y 有參數 μ 及 σ^2 之常態分佈, 以 $\mathcal{N}(\mu, \sigma^2)$ 表之, 即其 p.d.f. 為

$$(40) \quad h(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)}, \quad y \in R,$$

其中 $\mu \in R, \sigma > 0$ 。則利用變數代換可得 $(Y - \mu)/\sigma$ 有 $\mathcal{N}(0, 1)$ 分佈。由此即得 $E(Y) = \mu, \text{Var}(Y) = \sigma^2$ (見習題)。再由 $\mathcal{N}(0, 1)$ 分佈之機率值表, 可得

$$P(-\sigma < Y - \mu < \sigma) \doteq 0.6827,$$

$$P(-2\sigma < Y - \mu < 2\sigma) \doteq 0.9545,$$

$$P(-3\sigma < Y - \mu < 3\sigma) \doteq 0.9973,$$

也就是 Y 會落在離期望值不超過 1 個, 2 個及 3 個標準差之區間的機率, 分別約為 0.6826, 0.9544 及 0.9973。亦即會超出期望值 3 個標準差是極不可能的。

最後 Y 之動差母函數為

$$(41) \quad \begin{aligned} M(t) &= \int_{-\infty}^{\infty} e^{ty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/(2\sigma^2)} dy, \\ &= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu-\sigma^2 t)/(2\sigma^2)} dy, \\ &= e^{\mu t + \sigma^2 t^2/2}, \quad -\infty < t < \infty。 \end{aligned}$$

例6.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,每個信封恰放進一份薪資單。令 N 表放對的信封數,求 $E(N)$ 及 $\text{Var}(N)$ 。

解.首先 N 可表示為 n 個指示函數(indicator function)之和:

$$N = I_{A_1} + I_{A_2} + \cdots + I_{A_n}。$$

$I_{A_i} = 1$ 或 0 就視事件 A_i 是否發生,其中 A_i 表第 i 個信封放的是正確的薪資單之事件。注意 A_1, A_2, \cdots, A_n 既不互斥亦不獨立。但 $P(A_i) = 1/n$, 因此 $E(I_{A_i}) = P(A_i) = 1/n, \forall i = 1, 2, \cdots, n$ 。由(13)得

$$E(N) = \sum_{i=1}^n E(I_{A_i}) = n \cdot 1/n = 1。$$

其次

$$E(N^2) = \sum_{i=1}^n E(I_{A_i}^2) + 2 \sum_{i < j} E(I_{A_i} I_{A_j})。$$

又 $E(I_{A_i}^2) = E(I_{A_i}) = 1/n$, 且對 $i \neq j$,

$$E(I_{A_i} I_{A_j}) = P(A_i \cap A_j) = \frac{1}{n(n-1)}。$$

故

$$E(N^2) = n \cdot \frac{1}{n} + 2 \binom{n}{2} \frac{1}{n(n-1)} = 2。$$

因此

$$\text{Var}(N) = E(N^2) - E^2(N) = 2 - 1 = 1。$$

故在“亂放”的情況下,平均會放對一封,標準差亦不大,也是1。

期望值並非對每一隨機變數皆存在,見下例。

例7.設 X 有 $C(0, 1)$ 分佈,即 X 之 p.d.f. 為

$$(42) \quad f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in R。$$

因 $\int_0^\infty xf(x)dx = \infty$, 故 $\int_{-\infty}^\infty xf(x)dx$ 不存在,即 $E(X)$ 不存在。讀者若對瑕積分(improper integral)不熟,請查一般微積分的書。

如果繪 f 及

$$g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, x \in R,$$

之圖形,可能會覺得二圖形樣子有點像,均對稱於 y 軸,且在曲線下與 x 軸間之面積均為1。但分別乘上 x 後,在 $[0, \infty)$ 之積分,前者為 ∞ ,後者仍為有限(且值為 $1/\sqrt{2\pi}$)。

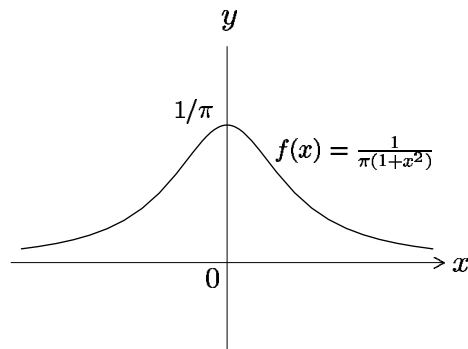


圖1. $C(0,1)$ 分佈之圖形

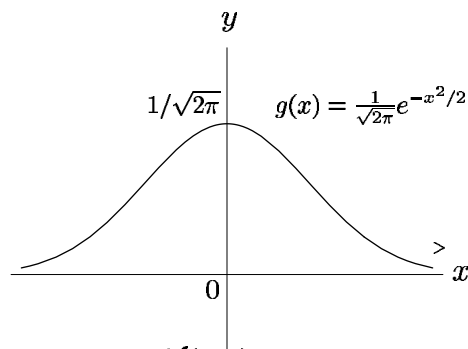


圖2. $\mathcal{N}(0,1)$ 分佈之圖形

這是一有趣的現象。一瑕積分之收斂與否,主要是看其尾部(tail),尾部接近0的速度夠快就收斂。乘上 x 後,當 x 很大時,會大幅度地拉抬 f 及 g 之值,但 g 下降的速度太快,拉抬效果不大,因此 $\int_0^\infty xg(x)dx$ 仍然為一有限值。但對 f 就產生效果了, $\int_0^\infty xf(x)dx$ 成爲 ∞ 。事實上對 $\forall k \geq 1$, $\int_0^\infty x^k g(x)dx < \infty$ (此亦可由(36)及(37)式看出)。也就是 $e^{-x^2/2}$ 下降到0的速度實在太快,以任意大的 x^k 都拉它不起。

$C(0,1)$ 分佈為一特別的柯西分佈(Cauchy distribution),此分佈當然是為了紀念著名的法國大數學家柯西(Cauchy, 1789-1857)。若 X 有 $C(\theta, a)$ 分佈,則其p.d.f.為

$$(43) \quad h(x) = \frac{a}{\pi(a^2 + (x - \theta)^2)}, \quad x \in R,$$

其中 $\theta \in R, a > 0$ 。柯西分佈在工程及物理的某些領域中用途不小(見Olkin et al.(1994)p.473)。由於尾部與常態分佈行為上的迥異,在某些統計題材中,常被拿來做為與常態分佈相對照。附帶一提,若 U, V 為二獨立且皆有 $\mathcal{N}(0,1)$ 分佈之隨機變數,則 $X = U/V$ 有柯西分佈。

例8.獨立地投擲一公正的骰子二次,令 X_1 及 X_2 分別表第一次及第二次所得之點數。又令 Z 表兩次投擲中,得到點數6之次數, Y 為事件 A 之指示函數,其中 A 表 $X_1 + X_2$ 為奇數之事件。表1給出 Y, Z 之聯合p.d.f. $f(y, z)$,及 Y, Z 之邊際p.d.f.'s $g(y)$ 與 $h(z)$ 。

		z			$g(y)$
		0	1	2	
y	0	$\frac{13}{36}$	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{2}$
	1	$\frac{12}{36}$	$\frac{6}{36}$	0	$\frac{1}{2}$
	$h(z)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

表1.

由表1立即看出 Y 與 Z 不獨立,但因

$$E(Y) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$

$$E(Z) = 0 \cdot \frac{25}{36} + 1 \cdot \frac{10}{36} + 2 \cdot \frac{1}{36} = \frac{1}{3},$$

$$E(YZ) = \frac{6}{36} = \frac{1}{6},$$

故

$$\text{Cov}(Y, Z) = E(YZ) - E(Y)E(Z) = \frac{1}{6} - \frac{1}{2} \cdot \frac{1}{3} = 0,$$

即 Y 與 Z 為無相關。

此為無相關但不獨立之例。

例9.某餅乾公司為了促銷其產品,在每盒餅乾中均隨機地放一張風景卡片,且設總共有 N 張不同的卡片。

我們先求每次若只買一盒,欲收集齊全 N 張卡片所需買的盒數之期望值。可看出當已有 n 張不同卡片時, $0 \leq n < N$, 欲得一張尚未有的卡片,所需買的盒數 X_n 有自1開始的幾何分佈,且參數為 $(N - n)/N$, 因此期望值為 $N/(N - n)$ (見底下註2)。故收集全部 N 張卡片所需買的盒數 $X_0 + X_1 + \cdots + X_{N-1}$ 之期望值為

$$E(X_0 + X_1 + \cdots + X_{N-1}) = \sum_{n=0}^{N-1} \frac{N}{N-n} = N(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}).$$

由微積分裡的結果知,當 $N \rightarrow \infty$ 時,

$$(44) \quad 1 + \frac{1}{2} + \cdots + \frac{1}{N} - \log(N) \rightarrow \gamma,$$

其中 γ 為歐拉常數(Euler's constant, Euler(1707-1783)為有史以來最多產的數學家,其生平介紹見黃文璋(1999)第十一章),其值約為0.5772。因此 N 很大時,所需買的盒數之期望值可以 $N(\log N + \gamma)$ 來估計。不過要注意 $N(1 + \frac{1}{2} + \cdots + \frac{1}{N})$ 與 $N(\log N + \gamma)$ 之差可能會很大。

其次,若一次買很多盒,譬如說 T 盒,則其中會包含多少張不同的卡片呢?我們仍求其期望值。為了方便,我們引進隨機變數 Y_1, Y_2, \cdots, Y_N , 其中 $Y_i = 1$, 若第 i 張卡片在這 T 張裡, 否則 $Y_i = 0$ 。則

$$P(Y_i = 0) = (1 - \frac{1}{N})^T = 1 - P(Y_i = 1).$$

因此

$$E(Y_i) = 1 - (1 - \frac{1}{N})^T。$$

故 T 張中共包含不同卡片張數的期望值為

$$E(\sum_{i=1}^T Y_i) = N(1 - (1 - \frac{1}{N})^T)。$$

明顯地上述期望值小於 N 。而之前我們推導出,若一次只買一盒,則當 N 很大時,欲收集全部 N 張卡片,所需盒數之期望值約為 $N \log N + N\gamma$ 。由此結果,引導我們考慮 $T = N \log N$ 。當然此 T 值並非整數,不過我們只是想了解這時 $N(1 - (1 - 1/N)^T)$ 的大小。

令

$$y = (1 - \frac{1}{N})^{N \log N}。$$

則

$$\begin{aligned} \log y &= N \log N \cdot \log(1 - \frac{1}{N}) \\ &= N \log N \cdot (-\frac{1}{N} - \frac{1}{2N^2} - \frac{1}{3N^3} - \dots) \\ &= -\log N \cdot (1 + \frac{1}{2N} + \frac{1}{3N^2} + \dots) \\ &< -\log N。 \end{aligned}$$

故得 $y < 1/N$ 。因此若一次買 $[N \log N] + 1$ 盒,其中 $[x]$ 表小於或等於 x 之最大整數,則所獲得不同卡片張數之期望值

$$\begin{aligned} &N(1 - (1 - \frac{1}{N})^{[N \log N] + 1}) \\ &> N(1 - (1 - \frac{1}{N})^{N \log N}) > N(1 - \frac{1}{N}) = N - 1。 \end{aligned}$$

由於最多只能得到 N 張不同的卡片,故得此時期望值已算是很接近 N 了。

註2. 參數 θ 之幾何分佈 (geometric distribution) 的 p.d.f. 為

$$f(x) = \theta(1 - \theta)^x, \quad x = 0, 1, \dots,$$

以 $Ge(\theta)$ 表之,其中 $0 < \theta < 1$ 。自1開始之幾何分佈之p.d.f.則為

$$g(x) = \theta(1 - \theta)^x, \quad x = 1, 2, \dots。$$

前者之期望值為 $(1 - \theta)/\theta$,後者為 $1/\theta$ (見習題)。

期望值與平均關係密切,那樣本平均之期望值及變異數又是如何呢?

例10.設 X_1, X_2, \dots, X_n 為一組獨立且有共同分佈(以i.i.d.表之)之隨機變數,且設 $E(X_1) = \mu$, $\text{Var}(X_1) = \sigma^2$ 存在。以 $\bar{X}_n = \sum_{i=1}^n X_i/n$ 表樣本平均,則

$$E(\bar{X}_n) = \frac{1}{n}E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}nE(X_i) = \mu,$$

且

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n},$$

此處用到(10)及(12)。也就是對一組變異數存在之i.i.d.的樣本,其樣本平均之期望值不變,但變異數會變小。即 \bar{X}_n 仍在 μ 之附近變動(因 μ 為 \bar{X}_n 之期望值),但隨著 n 之增大,此變動愈來愈小。

另外,一般以

$$(45) \quad S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

表 X_1, X_2, \dots, X_n 之樣本變異數(sample variance)。為什麼除以 $n-1$ 而不是除以 n 呢(不過有時為了簡便只除以 n)?因如此一來, S_n^2 的期望值剛好就是 σ^2 。即

$$(46) \quad E(S_n^2) = \sigma^2。$$

這只要利用

$$(47) \quad \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2,$$

便可得證,我們留在習題讓各位自行完成。

5 期望值引發之謬誤

由於人們對小的數往往不易理解究竟小到什麼程度,於是有時改以大的數或換算成時間來說明。

例如,要形容丟一銅板連得10個正面之不容易,可以說“假設某人每天丟一銅板10次,則差不多要3年,才可能得到10個正面。這裡便是用到連得10個正面之機率為 $1/1,024$,而3年共約有1,095天, $1,095$ 乘上 $1/1,024$ 約等於1。

其實不只對小機率事件,對很大的數,有時也會換以時間來描述。如美國微軟(Microsoft)公司總裁比爾蓋茲(Bill Gates)富可敵國,其所擁有財產達880億美元。人們用各種方式來說明880億美元究竟有多大:如果每秒鐘賺2,500美元,則一年可賺788.4億美元;我國一般公務員,月薪若以6萬台幣計,每年發薪13.5個月,1美元以換32元台幣計,則不吃不喝要約350萬年才有相當的財富。

大部分的人滿意於這樣的解釋。機率與時間(次數)遂經常被互換,謬誤因此而生。

民國83年6月21日,聯合報第3版有下述三段報導。

1.綠基會:百座核電廠二十年內出現機率四成五爐心熔毀機率 美國解密資料

【記者林如森/台北報導】環保團體綠色消費者基金會昨天公布一份該基金會從美國得來的解密資料指出,美國一百座核電廠在二十年內出現的爐心熔毀機率累積值可達百分之四十五,若加上不確定因素,其機率則在百分之九十九至百分之六間,這個機率遠比核能界所宣稱的十萬爐年一次高很多(一個反應爐運轉一年為一爐年),...

2.台電:不是百分之四五 而是零點四五次 核四廠十萬年才會發生一次

【記者李文娟/台北報導】台電公司昨天指出,綠色消費者基金會說美國一百座核電廠未來二十年內出現爐心熔毀的機率值可達百分之四十五是不對的,應該說“累積發生爐心熔毀次數是零點四五次”而且就算發生了,影響也不大。

台電說,這種算法是假設每一個反應器的爐心熔毀機率是每一萬年發年二點二五次,然後推估一百座核電機組連續運轉二十年,累積發生爐心熔毀次數為零點四五次,以此類推,如果一百座核電機組連續運轉四十年,累積發生爐心熔毀次數為零點九次,也就是說,一百座核電機組連續運轉四十年,爐心熔毀一次也不會發生。...

3.數據需要科學解釋

【記者李彥甫/特稿】綠色消費者基金會引用美國核能管制委員會的報告指出,未來一百座核電廠運轉二十年將有百分之四十五的爐心熔毀率。這個驚人的數據正說明,核四問題一定要用正確的科學方法解釋,而不要輕易掉入正反雙方的統計陷阱。

百分之四十五的爐心熔毀率是假設每一爐年(每一反應爐運轉一年)發生的機率是萬分之三,乘上一百座核電廠運轉二十年所得,但從機率理論來看,能否直接乘這些數值,要看母體的大小,換句話說,以美國核管會報告的算法,全世界數百座核電廠運轉數十年的爐心熔毀將超過百分之百,這是不合邏輯的。

綠色基金會以台北市一個月的車禍率為十萬分之一為例,並非每輛車會出十萬分之一的車禍,而是台北市三百萬輛車有三十輛會在一個月內出事。時間與母體群大小在這種風險機率計算中非常重要,一般來說只有大母體才能使用,例如車禍就是大母體計算很好的例子,但核電廠數目屈指可數,似乎不宜適用大母體計算。

但這件事也給社會大眾一個警惕,台電宣稱爐心熔毀率每爐年只有十萬分之一,其實是不完整的資訊,台灣有六部核電機組,也都運轉了好幾年,發年爐心熔毀的機率絕不只十萬分之一,真實的數目為何,大概只有專家才算得出來,台電有責任告訴社會大眾事實的真相,科學是不容一筆帶過的輕描淡寫。

這三段報導都是不知所云。在第一段中,還有機率在百分之九十九至百分之六間?這麼大的區間,乾脆說在百分之百至0間好了。還說此機率比十萬爐年一次高很多。依其方式,西瓜與甘蔗也可以比大小了。至於那位

跳出來將兩方各打50大板的記者,使我們覺得孔子所講的“知之為知之,不知為不知,是知也”,是多麼不易做到。他既沒有“隨機”的概念(說“三百萬輛車有三十輛會在一個月內出事”),又胡亂堆砌一些“時間”及“母體群大小”等字眼。

至於第二段,為台電的反駁。台電專業性實不應懷疑,只是仍是令人遺憾。不知是聘的人不夠好,還是進台電後,人會變鈍了。

依台電的說法,我們猜測他們是假設一反應器每年會發生爐心熔毀的機率為 $2.25/10,000$ 。假設每年及每台反應器之運轉為獨立,則100個反應器連續運轉40年,皆未發生爐心熔毀之機率為

$$\left(1 - \frac{2.25}{10,000}\right)^{4,000}。$$

利用微積分裡的結果:當 $|a_n|$ 很小,而 b_n 很大,則

$$(48) \quad (1 + a_n)^{b_n} \doteq e^{a_n b_n},$$

只要 $\{a_n b_n\}$ 不會趨近於至 ∞ 。由(51)即得

$$\left(1 - \frac{2.25}{10,000}\right)^{4,000} \doteq e^{-0.9} \doteq 0.4066。$$

亦即至少會發生一次爐心熔毀的機率約為

$$1 - 0.4066 = 0.5934。$$

台電(以及一般人)基本上是犯了類如下述的錯誤:對一出現正面機率為0.1的銅板,丟9次不會出現正面,要丟10次才會出現1個正面。這是不少人的看法。西元1998年,世界杯足球賽在巴黎舉行。美國隊於分組預賽時,輸給伊朗。美國足球協會主席羅森伯格認為這支美國隊比上屆還強,理由是“和伊朗比十場會贏九場,但偏偏這是第十場”(見87年6月23日中國時報第7版)。看多這種口語式的講法後,一般人對機率與次數,便常不正確的互相轉換。

事實上,令 X 表得到第1個正面所需丟之次數,則 X 有參數0.1,自1開始之幾何分佈,因此 $E(X) = 1/0.1 = 10$ 。10是期望值,而非真的需要丟10次

才會得到一個正面。運氣好丟第一次便得到正面,運氣不好,丟1,000次也得不到一個正面。所需次數的期望值與真正所需的次數不要混為一談。本節一開始提及的每日丟銅板10次的例子也是類似。10個皆為正面之機率為 $1/1,024$,每日丟一次,故得到10個正面所需之日數,亦為自1開始,參數為 $1/1,024$ 之幾何分佈,因此期望日數為1,024。在這類例子中,由於幾何分佈的關係,成功機率的倒數,便為期望次數。

現今 Y 表發生爐心熔毀所需之爐年,則 Y 有參數 $2.25/10,000$,自1開始之幾何分佈。因此

$$E(Y) = \frac{10,000}{2.25} \doteq 4444.4。$$

期望值的確大於4,000,當然還是不表示經4,000爐年仍不會發生爐心熔毀。

這類誤將期望時間(或次數)當做所需時間的例子不少。另外,有些人以時間來表示機率,卻流於文字及數字遊戲,連要想替其找個合理的解釋都很困難。例如,在王天戈(1985),為了說明核能電廠之安全,將核能工業與其他工業或天然災害比較,他引用表2而得到核能電廠“事實上可能”造成的災害要比一般人“以為可能”造成的災害要小得多。

表2中之事故,有些是實際上會發生,且曾發生的,有些則是從不曾發生的。如除隕星衝擊(這是沒人知道是否曾發生過的事)外,那些要一千年以上才發生一次的事。曾發生的事故,以大家較熟悉的飛機失事及地震為例,所列之幾年一次,遠低於實際的發生頻率。光是中華航空公司在西元1994年(名古屋)及1998年(大園),就各發生一次死亡人數均超過200人的空難。地震方面,西元1990年以來,十年間世界上死亡人數超過百人的地震就已至少有15次,超過千人則有十次之多,即使台灣,自西元1900年以來,百年間死亡超過百人的地震也有6次之多(見新新聞週報655期,1999年9月23日-9月29日)。至於不曾發生的事故,所謂幾年一次,實難了解其含義。之前我們將丟銅板連得10個正面,換成要丟3年,也須講清楚是每天丟一次,否則光說每3年發生一次,只會令人莫名其妙。而表2中之事故的幾年一次,一方面沒有說明其代表的意義,另一方面,也沒有交代那些數字是如何產生的。顯

然該文作者也不擬讓人明白。

事故種類	造成100人死亡的機會	造成1,000人死亡的機會
(1)人爲因素		
飛機失事	每2年一次	每2,000年一次
火災	每7年一次	每200年一次
爆炸	每16年一次	每120年一次
毒氣中毒	每100年一次	每1,000年一次
(2)自然因素		
龍捲風	每5年一次	甚小
颶風	每5年一次	每25年一次
地震	每20年一次	每50年一次
隕星衝擊	每100,000年一次	每1,000,000年一次
(3)核子事故		
100個核能電廠	每100,000年一次	每1,000,000年一次

表2. 一些災害與核能電廠安全之比較

平均值表面上的大小,有時也不一定能正確的解釋事情,見下例。

例11.曾有人抱怨美國加州大學柏克萊分校(University of California at Berkeley)的研究生院對女生較不公。經過統計在某年有8,442個男生申請,有4,321個女生申請,其中有約44%的男生被錄取,有約35%的女生被錄取。假設男生與女生的能力差不多,則上述數據顯示出女生的錄取率顯然較低。各個不同的研究所是分別辦理收學生的工作。若以各研究所的資料來看,有些研究所收的男生比率較高,有些則是女生高,甚至對女生有利的研究所還較多,究竟這是怎麼回事。

由於該校有一百多個研究所,資料太多,若選出6個較大的研究所來看,可得表3及表4(見Freedman et al. (1991))。

男女相比,研究所A對男生最不利,研究所E對女生最不利。但若6個研究所一起看,有44%的男生被錄取,有30%的女生被錄取,相差14%。此矛

盾是如何產生的？

男生

所別	申請人數	錄取率
A	825	62%
B	560	63%
C	325	37%
D	417	33%
E	191	28%
F	373	6%

表3.

女生

所別	申請人數	錄取率
A	108	82%
B	25	68%
C	593	34%
D	375	35%
E	393	24%
F	341	7%

表4.

如果仔細看表3及表4, A, B二研究所較易申請,有一半以上的男生申請;而C, D, E, F四個研究所較難申請,有超過90%的女生申請。換句話說,男生較多申請較容易進的研究所,女生較多申請較難進的研究所。因此不能光由表面上的公平與否來推測對女生是否不公。

這種例子很多。譬如說去餐廳的顧客通常會覺得餐廳生意不錯,為什麼?因為大部分的人是在餐廳生意不錯的時候去的,而生意不好時就是那個時段去的人數較少。所以長期下來,比較多人的印象是餐廳常生意不錯。例如,假設在一個月30天中,有8天各有100位顧客,有22天各只有10位顧客。則在一個月總共去的1,020位顧客中,有近八成認為餐廳人滿為患。但實際上30天裡,有22天生意是很差的。一個類似的問題是,某一系假設開20門課,如果做問卷,調查每位學生他所修的各門課的修課人數,則得到的平均值會大於實際各門課的平均修課人數。你有想通這原因嗎?

6 共變異數及相關係數再探

共變異數顧名思義是在量測二隨機變數 X 及 Y 同時變化的大小,正如變異數是量測一隨機變數變化的大小。

$\text{Cov}(X, Y)$ 可以是正的也可以是負的。若大的 X 之值便對應大的 Y 之值,且小的 X 之值便對應小的 Y 之值,則 $\text{Cov}(X, Y)$ 很可能為正。而若 X 與 Y 移動的方向相反, $\text{Cov}(X, Y)$ 便很可能為負。此因

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))),$$

若 X 與 Y 都在其期望值之同一側,則 $(X - E(X))(Y - E(Y))$ 為正,而若 X 與 Y 在其期望值之異側,則 $(X - E(X))(Y - E(Y))$ 為負。就視那一種情況的“加權平均”較大, $\text{Cov}(X, Y)$ 便為正或負。 $\text{Cov}(X, Y)$ 也會有等於0的時候。

若以 X, Y 分別表父親的身高及兒子的身高,我們會預期 $\text{Cov}(X, Y)$ 為正。若以 X, Y 分別表體重及跳高高度,我們會預期 $\text{Cov}(X, Y)$ 為負。

做為量測二隨機變數 X 及 Y 間之變化情況,共變異數有一重大的缺點,就是其值與量測 X 與 Y 之尺度有關。首先不難驗證

$$(49) \quad \text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y),$$

其中 a, b, c, d 為常數。現令 X, Y 分別表學生入學及一年後之身高。在以公分為單位時,假設得到 $\text{Cov}(X, Y) = 13.6$ 。若改以公尺為單位,而得 $X^* = X/100, Y^* = Y/100$,此時 $\text{Cov}(X^*, Y^*) = 0.00136$ 。觀測雖未改變,但因尺度的不同,共變異數之大小卻可能有很大的差異。為了消除這種對尺度的依賴,相關係數因應而生(見(16)式)。

$\rho(X, Y)$ 與 $\text{Cov}(X, Y)$ 之符號相同,且 $\rho(X, Y) = 0$ 若且唯若 $\text{Cov}(X, Y) = 0$ 。 $\rho(X, Y)$ 仍保持 $\text{Cov}(X, Y)$ 中所傳送的關於 X 與 Y 之相依性,且此時(利用(49)式)

$$(50) \quad \rho(aX + b, cX + d) = \frac{ac}{|ac|}\rho(X, Y), \quad ac \neq 0。$$

所以只要尺度之改變,並未使 X, Y 之值的方向變成相反(即 $ac > 0$),則相關係數不會改變。

相關係數究竟是代表什麼指標呢?它主要是反映二隨機變數之分佈的線性關係之強度及符號(正負關係)。因此相關係數為0,僅表二變數之線

性相關性很低,而非表二變數機率上無關(probabilistically unrelated,即獨立)。見下例。

例12. 設 X, Y 為離散型之隨機變數,且

$$P(X = -1, Y = 1) = P(X = 1, Y = 1) = \frac{1}{3},$$
$$P(X = -2, Y = 4) = P(X = 2, Y = 4) = \frac{1}{6},$$

至於對其他的 (x, y) , $P(X = x, Y = y) = 0$ 。則

$$E(X) = 0, \text{Var}(X) = 2,$$
$$E(Y) = 2, \text{Var}(Y) = \frac{38}{9},$$

且

$$E(XY) = 0。$$

因此

$$\text{Cov}(X, Y) = 0,$$

且

$$\rho(X, Y) = 0。$$

另一方面, X 與 Y 間有一很強的非線性關係,因對每一 (x, y) , 只要 $P(X = x, Y = y) \neq 0$, 則 $y = x^2$ 。即

$$P(Y = X^2) = 1。$$

因此雖由 $\rho(X, Y) = 0$, 得知 X 與 Y 間缺乏(機率上的)線性關係,但並未排除 X 與 Y 間有一非線性的關係。

底下我們證明對任二隨機變數 X, Y , $|\rho(X, Y)| \leq 1$, 且 $|\rho(X, Y)| = 1$, 若且唯若 $P(X = aY + b) = 1$, 其中 a, b 為某二常數, 且 $a \neq 0$ 。由於相關係數的值介於正負1之間, 且當其值等於+1或-1時, 二變數有一線性關

係之機率為1,這便說明了如前所述,相關係數是用來量測二變數之線性相依情形。

我們先證底下著名的Schwarz不等式(Schwarz's Inequality)。Schwarz (1843-1921)為近代德國著名的數學家。

定理3. 設 X, Y 為二隨機變數,且 $E(X^2), E(Y^2) < \infty$ 。則

$$(51) \quad E^2(XY) \leq E(X^2)E(Y^2)。$$

又(51)中等式成立,若且唯若 $P(X=0)=1$,或存在一常數 a ,使得 $P(Y=aX)=1$ 。

證明. 若 $P(X=0)=1$,則 $P(XY=0)=1$, $E(XY)=0$ 且 $E(X^2)=0$,此時(51)左右皆為0,故等式成立。另外,若 $P(Y=aX)=1$,亦可證明此時(51)之左右皆等於 $(aE(X^2))^2$,故等式仍成立。底下設 $P(X=0) < 1$,因此 $E(X^2) > 0$ 。

現對 $\forall \lambda > 0$,

$$(52) \quad E((Y - \lambda X)^2) = \lambda^2 E(X^2) - 2\lambda E(XY) + E(Y^2)$$

為一恆非負之 λ 的二次式。又因 λ^2 之係數 $E(X^2) > 0$,故此二次式在 $\lambda = \lambda_0 = E(XY)/E(X^2)$ 有極小,且

$$0 \leq E((Y - \lambda_0 X)^2) = E(Y^2) - E^2(XY)/E(X^2),$$

故(51)式成立。若(51)中等式成立,則 $E((Y - \lambda_0 X)^2) = 0$,此時 $P(Y = \lambda_0 X) = 1$ 。證畢。

將Schwarz不等式應用至隨機變數 $X - E(X)$ 及 $Y - E(Y)$,得

$$E^2((X - E(X))(Y - E(Y))) \leq E((X - E(X))^2)E((Y - E(Y))^2),$$

即

$$(53) \quad (\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y)。$$

亦即證出對任二隨機變數 X, Y ,

$$(54) \quad |\rho(X, Y)| \leq 1。$$

我們又有下述所要的推論。

系理1. 設 X, Y 為二隨機變數, 且 $E(X^2), E(Y^2) < \infty$ 。則 $|\rho(X, Y)| = 1$, 若且唯若存在二常數 $a, b, a \neq 0$, 使得 $P(Y = aX + b) = 1$ 。

證明. 首先, 由相關係數之定義知 $\text{Var}(X)$ 及 $\text{Var}(Y)$ 皆不能為0, 因此 $P(X = 0) = 1$ 不能成立。利用定理3於(53)式, 即得 $|\rho(X, Y)| = 1$, 若且唯若存在一常數 a , 且 $a \neq 0$ (否則 $\text{Var}(Y) = 0$), 使得 $P(Y - E(Y) = a(X - E(X))) = 1$ 。不難看出此即我們要證的結果。

在系理1中, 若 $a > 0$, 則 $\rho(X, Y) = 1$, 若 $a < 0$, 則 $\rho(X, Y) = -1$ 。二隨機變數即使相關係數很高, 只是表其分佈上所顯現的現象, 並不一定導致二隨機現象間一定有密切的關係。例如, 令 X 表某地可樂之月銷售量, Y 表該地每月至醫院腸胃科就診的人數。則 X 與 Y 之相關係數可能很高。但這並不是說喝可樂易拉肚子, 而很可能是夏天天氣炎熱, 可樂需求量因而增大, 而天氣炎熱時病蟲又易滋長, 因此腸胃有問題的人也變多。如此導致 X 與 Y 之相關係數很高。至於喝可樂與拉肚子間倒不見得有因果關係。

例13. 設 X, Y 為在某礦物中, 兩種金屬分別所佔之比率。又設 X, Y 之聯合p.d.f. 為

$$f(x, y) = \begin{cases} 360xy^2(1-x-y), & 0 \leq x+y \leq 1, x, y \geq 0, \\ 0 & \text{其他。} \end{cases}$$

則可驗證 X, Y 分別有 $Be(2, 5)$ 及 $Be(3, 4)$ 分佈。因此

$$E(X) = \frac{2}{2+5} = \frac{2}{7}, \quad E(Y) = \frac{3}{3+4} = \frac{3}{7}。$$

又

$$(55) \quad E(XY) = \int_0^1 \int_0^{1-x} xyf(x, y)dydx$$

$$\begin{aligned}
&= 18 \int_0^1 x^2(1-x)^5 dx \\
&= 18 \cdot \frac{\Gamma(3)\Gamma(6)}{\Gamma(9)} \\
&= 18 \cdot \frac{2!5!}{8!} = \frac{18}{168} = \frac{3}{28},
\end{aligned}$$

故得

$$\text{Cov}(X, Y) = \frac{3}{28} - \frac{2}{7} \cdot \frac{3}{7} = -\frac{3}{196}。$$

此處 X, Y 之共變異數為負值並非不合理。因 $X + Y$ 不能大於 1, 故若其一較大, 會使另一較小。因此 X 與 Y 之變化為朝相反的方向, 從而有負的共變異數並不足為奇。

註 3. 若隨機變數 X 之 p.d.f. 為

$$(56) \quad f(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}, \quad 0 < x < 1,$$

其中 $r, s > 0$, 則稱 X 有 $\text{Be}(r, s)$ 分佈。由此得對 $\forall r, s > 0$,

$$(57) \quad \int_0^1 x^{r-1}(1-x)^{s-1} dx = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}。$$

習 題

1. 試證(15)式及(17)式成立。
2. 對一隨機變數 X , 試證 $E(X)$ 存在, 若且唯若 $E(|X|)$ 存在。
3. 對一隨機變數 X , 試證若 $E(X^m)$ 存在, $m \geq 1$, 則 $E(X^k)$ 存在, $\forall 0 \leq k \leq m-1$ 。
4. 試證 $\text{Cov}(\sum_{i=1}^m \alpha_i X_i, \sum_{j=1}^n \beta_j Y_j) = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \text{Cov}(X_i, Y_j)$ 。
5. 試求 $\mathcal{C}(\theta, a)$ 分佈之中間值及眾數。

6. 試證(25)式成立。
7. 試證(27)式成立。
8. 試證(29)中二式成立。
9. 設 X, Y, Z, U , 分別有 $\mathcal{N}(\mu, \sigma^2), \mathcal{B}e(r, s), \mathcal{G}e(\theta)$, 參數 θ 自1開始之幾何分佈。試證
- (i) $E(X) = \mu, \text{Var}(X) = \sigma^2$;
 - (ii) $E(Y) = r/(r + s), \text{Var}(Y) = rs/((r + s)^2(r + s + 1))$;
 - (iii) $E(Z) = (1 - \theta)/\theta, \text{Var}(Z) = (1 - \theta)/\theta^2$;
 - (iv) $E(U) = 1/\theta, \text{Var}(U) = (1 - \theta)/\theta^2$ 。
10. 設 X 有 $\mathcal{C}(0, 1)$ 分佈, 試證 $E(e^{tX})$ 只有在 $t = 0$ 才存在。
11. 設 X, Y 為二獨立的隨機變數, 且 $E(X^4) = 2, E(Y^2) = 1, E(X^2) = 1, E(Y) = 0$ 。求 $\text{Var}(X^2Y)$ 。
12. 設 X_1, X_2, X_3 為三獨立的隨機變數, 且令 $\text{Var}(X_i) = \sigma_i^2, i = 1, 2, 3$ 。求 $\rho(X_1 - X_2, X_2 + X_3)$ 。
13. 設二隨機變數 X, Y 滿足 $\text{Var}(X) = 1, \text{Var}(Y) = 2$, 且 $\rho(X, Y) = 1/2$ 。求 $\text{Var}(X - 2Y)$ 。
14. 一盒中有3紅球及2黑球, 自此盒中隨機地取出2球(每次取出後不放回), 令 U, V 分別表其中之紅球及黑球數。求 $\rho(U, V)$ 。
15. 一盒中有編號分別為1, 2, 3之3個球, 隨機地取出2球(每次取出後不放回), 令 X, Y 分別表兩次之號碼。求 $\rho(X, Y)$ 。
16. 設一袋中有編號1至 N 之 N 個球, 依序取出 n 個, $n \leq N$, 每次取出後不放回。令 S_n 表所取出球之點數和, 求 $E(S_n)$ 及 $\text{Var}(S_n)$ 。
17. 設 X 之 p.d.f. 為一偶函數, 且 $E(X^4) < \infty$ 。試證 X 與 X^2 無相關。

18. 試證(47)式成立,並用此證明(46)式成立。
19. 將 n 個球隨機地放進 r 個盒子中,令 $X_i = 1$ 表第 i 個盒子是空的,否則令 $X_i = 0, i = 1, 2, \dots, r$ 。又令 S_r 表空盒子之個數。求(i) $E(X_i)$, (ii) $E(X_i X_j)$, (iii) $\text{Cov}(X_i, X_j)$, (iv) $E(S_r)$, (v) $\text{Var}(S_r)$ 。
20. 試證(49)式成立。
21. 試驗證例12中之 $E(X), \text{Var}(X), E(Y), \text{Var}(Y)$ 皆如所給。
22. 試驗證例13中, X, Y 分別有 $Be(2, 5)$ 及 $Be(3, 4)$ 分佈。
23. 試驗證(55)中第二等號成立。
24. 某地攤有一遊戲,玩一次要付10元。攤主放8個白棋子及8個黑棋子在一袋中。玩者自袋中摸出五個棋子。若拿到5個白的可得200元,拿到4個白的可得20元,拿到3個白的可得5元。問你是否願意玩此遊戲?
25. 甲乙兩袋中皆有紅白二種籤,每支籤可能中獎也可能不中獎,兩袋中皆是紅籤中獎率較高。今將兩袋的籤混為一袋,則新袋中紅籤中獎率是否仍較高?是否可能新袋中紅籤中獎率不到白籤的一半?
26. 在賽馬等競技中,所謂賠率是十賠一,乃指所賭的馬贏的機率為輸的機率的10倍。
(i)試導出在此情況下,該馬被認為贏的機率為何?
(ii)試問十賠一的馬贏的機率是否為五賠一的馬贏的機率之兩倍?
27. 曾有人宣稱“根據指紋專家的研究,六百四十億個指紋中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”。試依其說法及數據,給出正確的結論。
28. 設生男生女的機會皆為 $1/2$ 。某國由於民情的關係,每一家庭皆希望能有男孩,但政府為抑制人口的成長,規定每一家庭只能有一男孩,若

前幾胎皆為女孩,則可繼續生,直至生出一男孩,便須停止。問這種政策執行的結果,是否會造成社會上女多於男?並給出理由。

29. 在上題中,若將每一家庭生孩子的限制,改為下述三情況之一:
- (i) 最多只能生 m 個小孩,
 - (ii) 生出 k 個男孩便須停止,
 - (iii) 生出 k 個男孩,或孩子數達到 m 個,便須停止,
- 則結果各會有何不同?

參考文獻

1. 王天戈(1985).核能安不安全?科學月刊,第16卷第9期,653-660。
2. 黃文璋(1994).機率論講義。全華科技圖書股份有限公司,台北。
3. 黃文璋(1999).數學欣賞。華泰文化事業股份有限公司,台北。
4. Olkin, J., Gleser, L. J. and Derman, C. (1994). *Probability Models and Applications*, 2nd ed. Macmillan College Publishing Company, New York.