隨機變數

黃文璋

國立高雄大學應用數學系

1 隨機變數

假設有一機率空間$(\Omega,\mathcal{F},\mathcal{P})$。此機率空間可能是觀測某一隨機現象而得，或就是給定而不論其由來。

樣本空間中的元素，不一定是數字，可以是紅球及白球，也可能是正面及反面等。做民意調查，想了解民眾對某議題的看法，選項可能有極力贊成、贊成、不贊成、極不贊成，及沒意見等。我們一向比較喜歡處理數字的問題。對上述情況，可將紅球視為1，白球視為0；正面視為1，反面視為0；極力贊成視為5，贊成視為4，不贊成視為3，極不贊成視為2，沒意見視為1。也就是將實驗的結果數量化。

調查選民對某候選人支持與否，以1表支持，0表不支持。雖是數量化了，但若調查1,000位選民，則樣本空間中共有$2^{1,000}$個元素，每一元素為一串長達1,000個的0或1的數字，其中第i個數字為1，表第i個選民支持該候選人，0表不支持。這麼大的樣本空間不但難以掌握，也顯得累贅。因我們可能只對其中共有多少人支持該候選人感到興趣。而究竟是那幾個選民支持，此訊息並不重要。因此可令

$$X = 1{,}000\text{個數字中1的個數}。$$

則對X, 樣本空間成為$\{0,1,2,\cdots,1{,}000\}$, 減小許多。

即使原先樣本空間中的元素就已是數字, 因種種目的, 我們仍可能會考慮它的某一函數。例如, 若隨機選取一蘋果, 量測其重量, 則可取樣本空間$\Omega=(0,\infty)$。若蘋果每單位重量的價格為a元, 則將蘋果的重量ω, 對應至$X=a\omega$, 而得其價格。當然我們也可以一開始就是觀測一隨機選取的蘋果之價格, 由於$a>0$, 此時樣本空間仍可取為$(0,\infty)$。

由於以上這些原因, **隨機變數**(random variable)的概念便自然地產生了。所謂隨機變數, 就是一個定義在樣本空間Ω的實函數。若以X表一隨機變數, 則此函數的定義域為Ω, 對應域為實數集合R, 值域則為R的一個子集合。換句話說對$\forall\omega\in\Omega$, $X(\omega)\in R$。

即使不論上述所提的那些原因, 在數學裡我們本來就經常在考慮各種函數。有些應用性的問題, 常可化為解跟某一函數有關的問題。函數在數學中的角色太重要, 數學中常在討論函數的各種性質。這樣看來, 隨機變數的產生, 就更不奇怪了。在同一樣本空間, 往往可定義出無限多個隨機變數。定義隨機變數, 可能是因實際需要, 也可能只是純粹數學的目的。

其實在很多隨機試驗裡, 我們早就隱含地使用隨機變數了。例如, 投擲一骰子2次, 令X表點數和; 投擲一銅板20次, 令Y表所得正面數等。

對於隨機變數X, 可以想成原先觀測到ω, 再轉換至$X(\omega)$, 樣本空間由Ω轉為R(或R的一個子集合)。但也可想成所觀測到的結果就是X。如前述隨機選取蘋果的例子。又如, 前述民調的例子, 負責抽樣的人有原始資料, 但他整理好後, 交出來的統計資料, 是支持的人數X。對大部分的人而言, 觀察值就是X, 樣本空間為$\{0,1,\cdots,1{,}000\}$。由此可見隨機變數亦可是一觀測值, 視為一隨機試驗的結果, 即視為樣本。

隨機變數的概念既然是如此卑之無甚高論, 它不過就是一個函數(或僅是不同階段下的樣本), 但我們也經常聽到有人說他是在進了(統計)博士班之後, 才終於了解隨機變數的意義。這又是為什麼呢?

首先以往我們對函數, 常以f,g,h或大寫的F,G,H等命名, 參數(或稱變數), 則以x,y,z等表示。有時我們說有一函數$f(x)$, 即使因函數較多, 有時會取名為$a(x),b(x)$等, 或藉助希臘字母$\xi(x)$, $\eta(x)$等, 較少見到

以大寫的英文字母X, Y, Z當做函數。但對於隨機變數, 我們往往以大寫的X, Y, Z等命名。又以往的函數, 變數常是附在一起的, 如$f(x), g(y)$等。但隨機變數這種函數, 其變數並不常與函數一起出現。例如, 我們會寫$f(x) = 3$, $g(y) = 5$等。但對隨機變數, 往往寫成$X = 3$, $Y^2 = 7$等。尤有進者, 除了在一開始定義時, 説隨機變數為一實函數, 之後我們很少稱它為函數, 連名稱都含有"變數"。因此不易使人與以往所學的函數連結在一起。

再看一個原因。隨機變數定義於樣本空間, 樣本空間中的元素之出現情況是隨機的, 以一機率函數來描述。例如, 在數學裡解方程式是很平常的。我們説解$f(x) = 3$, 就是找出滿足$f(x) = 3$的實數x。但對隨機變數X, 我們較少説解$X = 3$。X如果是由某一Ω映至實數R, 是可以有一滿足$X(\omega) = 3$ 之解集合$\{\omega | X(\omega) = 3\}$(如同方程式可能不易解, 有時無法明確地表示出此集合)。但我們通常不會就此打住, 而會去求此集合之機率。甚至後來經過一番改造(下一節會説明), 我們常略去解$X = 3$的過程, 而直接給出$X = 3$的機率(見底下例2)。換句話説, 雖然本身是一函數, 但常另有跟機率有關的函數, 伴隨一隨機變數, 指揮該隨機變數的行為。那些函數的味道, 就比較接近我們以往所遇到的函數。而那些函數也採用一般我們熟悉的函數之符號, 如f, g, h及F, G, H等。這也可以解釋為什麼在隨機變數這一階段, 我們不採慣用的函數符號。

第三個原因是, 由於在樣本空間Ω上, σ-體$\mathcal{F}$選取的不同, 使得並非每一由Ω映至R的函數, 皆可當做隨機變數。一般要求(尤其在較高等的機率論中)還要滿足

$$(1) \qquad 對 \forall x \in R, \{\omega | X(\omega) \le x\} \in \mathcal{F}。$$

為什麼要滿足此條件, 也留在下一節再説明。如果$\mathcal{F}$取得小些, 則有些Ω上的實函數, 就不一定滿足條件(1), 因此就非隨機變數了。數學家一向愛好此道, 給一定義, 列出一些條件, 總會讓它有時滿足, 有時不滿足, 問題因此變化多端。雖然增加挑戰性, 對初學者而言, 迷惘也隨之而生。幸好在統計學裡, $\mathcal{F}$常取得夠大。例如, 若Ω為一可數的集合, 我們通常取$\mathcal{F}$為包

含Ω之所有子集合所形成的集合, 因此條件(1)自然成立。如此一來所有由Ω映至R的函數, 便皆為隨機變數了。

其他還有若干初學者不易了解隨機變數的概念之原因, 我們陸續會解釋。

隨機變數常以大寫的英文字母表示, 而它的觀測值則以對應的小寫字母表示。有時會說"隨機變數X取值x", 或說"求$X=x$之機率"。

定義一隨機變數後, 其值域可視為新的樣本空間, 那原來的機率函數(定義在原來樣本空間所產生之σ-體), 是否仍適用此隨機變數?

假設樣本空間$\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$, 且如一般, 取$\mathcal{F}$為包含$\Omega$之所有子集合所形成的$\sigma$-體。定義一隨機變數$X$, 其值域以$\Omega_1=\{x_1,x_2,\cdots,x_m\}$, $m\leq n$, 表之。觀測到$X=x_i$, 若且唯若原試驗的結果為$\omega\in\Omega$, 且使得$X(\omega)=x_i$。則可取Ω_1為新的樣本空間, Ω_1之所有子集合所形成的σ-體以$\mathcal{F}_1$表之, 而新的機率函數為P_X, 其中

$$P_X(X=x_i)=P(\{\omega|\omega\in\Omega\text{且}X(\omega)=x_i\}), \tag{2}$$

如此便得一新的機率空間(Ω_1,$\mathcal{F}_1$,P_X)。要注意的是, 比較好的寫法是$P_X(\{X=x_i\})$, 因$\{X=x_i\}$才是一集合, 不過為了簡便, 我們常簡單地寫成$P_X(X=x_i)$, 有時寫成$P_X\{X=x_i\}$。可以驗證P_X的確滿足機率函數的公理。雖然P_X與P為不同的函數(定義域便不同了), 但因有(2)式, 今後如果不會混淆, 我們通常只寫$P(X=x_i)$, 且將P解釋為"機率"的意思。

例1. 投擲一銅板3次, 並觀測3次所得結果。則樣本空間$\Omega=\{$正正正, 正正反, 正反正, 反正正, 正反反, 反正反, 反反正, 反反反$\}$, 其中"正正反"表頭兩次出現正面, 第三次出現反面, 餘類推。且設每一樣本之機率為1/8(因此這是一公正的銅板)。令X表共得之正面數。顯然X之值域為$\{0,1,2,3\}$。則

$$P(X=1)=P(\{\text{正反反},\text{反正反},\text{反反正}\})=\frac{3}{8}\text{。}$$

同理可得$P(X=0)=1/8$, $P(X=2)=3/8$, $P(X=3)=1/8$。

次令$Y=$正面數$-$反面數。則Y之值域為$\{-3,-1,1,3\}$, 且

$$P(Y=-3)=P(\{反反反\})=\frac{1}{8}。$$

同理可得$P(Y=-1)=P(Y=1)=3/8$, $P(Y=3)=1/8$。

例2. 投擲一公正的銅板100次,並觀測100次所得結果。則樣本空間Ω中共有2^{100}個元素。由於此銅板為公正, 所以每個樣本之機率均相同, 皆為$1/2^{100}$。令X表所得之正面數。欲求$X=10$之機率。因每一樣本皆為一串長達100的“正、反”文字, 本例不像上例可輕易列舉出對應$X=10$之原事件。但利用排列組合的技巧, 可得

$$P(X=10)\ =\ \frac{\Omega中有10個“正”之元素個數}{\Omega之元素個數}=\frac{\binom{100}{10}}{2^{100}}。$$

又對$i=0,1,\cdots,100$,

$$P(X=i)=\frac{\binom{100}{i}}{2^{100}}。$$

雖然我們並未展示出事件$\{X=i\}$究竟為何, 但其機率卻可求出。

一般而言, 令Ω_1表隨機變數X所引發出來的樣本空間, 則對$\forall A\subset\Omega_1$, 我們定義引發出來的機率函數

$$P(X\in A)=P(\{\omega|\omega\in\Omega, 且X(\omega)\in A\})。$$

例3. 如果Ω本來就是實數的一個子集合, 則$X(\omega)=\omega$為一很自然的隨機變數。例如, 一袋中有10張紙牌, 分別寫數字1至10。隨機地取一張, 觀測所得之點數。則樣本空間$\Omega=\{1,2,\cdots,10\}$。令$X(\omega)=\omega$, 便得一隨機變數。另一種講法是自袋中隨機取一張紙牌, 並令X表所得之點數。這種講法就是直接引進隨機變數。

其他如令Y表某袋中的紅球數, 令Z表某校戴眼鏡的學生人數, 都是直接引進隨機變數之例。紅球、戴眼鏡本來皆非數字, 但因我們將觀測的結

果數量化, 或者說考慮數值上的特性, 便產生了隨機變數。

在同一樣本空間中, 可定義無數個隨機變數。即使$\Omega = \{0, 1\}$只有兩個元素, 對任二實數a, b, 令$X(0) = a$, $X(1) = b$, 便得一隨機變數。

常數也是一隨機變數。即對一實數c, 若$X(\omega) = c$, $\forall \omega \in R$, 則X為一隨機變數。有時以$X \equiv c$表之。這種只取一個值之隨機變數, 由於並不會變, 雖然仍視為隨機變數, 但稱之為**退化的**(degenerate)隨機變數。

例4. 對例1中的Ω, 依序令Z(正正正) $= a_1$, Z(正正反) $= a_2, \cdots$, Z(反反反) $= a_8$, 其中$a_1, a_2, \cdots, a_8$皆為實數, 則Z為一隨機變數。可看出若取$a_1 = 3$, $a_2 = a_3 = a_4 = 2$, $a_5 = a_6 = a_7 = 1$, $a_8 = 0$, 即得例1中之X; 若取$a_1 = 3$, $a_2 = a_3 = a_4 = 1$, $a_5 = a_6 = a_7 = -1$, $a_8 = -3$, 即得例1中的Y。任一定義於Ω上的隨機變數, 皆可經由適當地選取$a_1, \cdots, a_8$而得。

再給一特殊但常見之隨機變數。

例5. 設A為樣本空間Ω之一子集合。令I_A表A之**指示函數**(indicator function), 即若A發生, 則$I_A = 1$, 否則$I_A = 0$。I_A便是"指示"事件A是否發生。因此

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

則I_A為一隨機變數, 它只取1與0兩個值, 且$P(I_A = 1) = P(A)$, $P(I_A = 0) = 1 - P(A)$。指示函數是除了退化的隨機變數外, 最簡單的隨機變數。利用指示函數, 可將每一事件對應一隨機變數。求事件之機率, 便可視為求隨機變數之機率。

2 分佈函數及機率密度函數

隨機變數已經是一函數, 對一隨機變數X, 我們又引出另一重要的函數, 即**累積分佈函數**(cumulative distribution function), 通常簡稱為**分佈**

函數, 又稱**分布函數**, 或**分配函數**:

(3) $$F_X(x) = P(X \le x), x \in R \text{ 。}$$

如果不會混淆, $F_X(x)$也可只寫成$F(x)$ 。對隨機變數X, 給定一$x \in R$, $F(x)$乃一機率值 。由於我們常對形式為$\{X \le x\}$的事件之機率有興趣, 所以特別以函數$F(x)$表此機率值 。一些常會用到的與X有關的機率值, 都可藉由F表示出來 。分佈函數可說是與隨機變數的關係極為密切之一函數 。

分佈函數是一定義域為實數集合, 對應域為區間$[0,1]$的函數 。由機率空間, 隨機變數, 至分佈函數, 當只給分佈函數時, 源頭之機率空間為何, 可能毫不清楚 。自完全不同的機率空間, 有可能會得到相同的分佈函數 。

由於是以(3)式來定義分佈函數, 這可解釋為何在上一節裡, 我們曾提到, 須滿足條件(1)才是隨機變數 。否則若有一$x \in R$, 使得$\{X \le x\} \notin \mathcal{F}$, 即$\{X \le x\}$非事件, 因此便不知$\{X \le x\}$之機率, 如此就得不到$F(x)$了 。

例6. 投擲一公正的銅板3次, 令X表所得之正面數 。則因$P(X=0) = 1/8$, $P(X=1) = 3/8$, $P(X=2) = 3/8$, $P(X=3) = 1/8$, 故

$$F(x) = \begin{cases} 0, & x < 0, \\ P(X=0) = \frac{1}{8}, & 0 \le x < 1, \\ P(X=0) + P(X=1) = \frac{1}{2}, & 1 \le x < 2, \\ P(X=0) + P(X=1) + P(X=2) = \frac{7}{8}, & 2 \le x < 3, \\ \sum_{i=0}^{3} P(X=i) = 1, & x \ge 3 \text{ 。} \end{cases}$$

圖1 為$F(x)$的圖形 。

我們說明一下圖1 。分佈函數F, 乃定義在整個實數上, 而非只定義在X取值的集合$\Omega_1 = \{0,1,2,3\}$ 。例如,

$$F(2.7) = P(X \le 2.7) = P(X = 0,1,2) = P(X \le 2) = \frac{7}{8} \text{ 。}$$

換句話說$F(2.7) = F(2)$ 。事實上對$2 \le x < 3$, $F(x)$之值皆為$7/8$ 。F的圖形, 在$\forall x_i \in \Omega_1$, 有一**跳升**(jump), 跳升的高度即$P(X = x_i)$ 。又$x < 0$時,

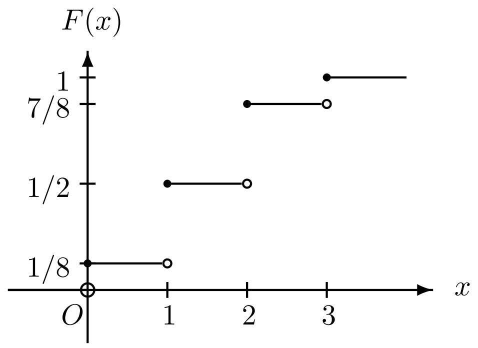


圖1　例6中$F(x)$之圖形

$F(x)=0$, 因X不會取負值。而$x>3$時, $F(x)=1$, 因X不會取大於3之值。也可看出F的圖形為一**階梯函數**(step function), 並且是**右連續**(right continuous)。即

$$\lim_{t \downarrow x} F(t)=F(x), x \in R,$$

其中$t \downarrow x$表t由x的右側漸減趨近至x。又$t \uparrow x$表t由x的左側漸增趨近至x。

退化的隨機變數, 其分佈函數亦為階梯函數。即若$X \equiv c$, 則$F(x)=0$, $x<c$, $F(x)=1$, $x \geq c$。反之若X有前述分佈函數, 則知$X \equiv c$。

我們將分佈函數具有的性質, 列成一定理。

定理2.1 設F為一**分佈函數**, 則有下述性質:

(i) $0 \leq F(x) \leq 1$, $\forall x \in R$;

(ii) F為一非漸減函數;

(iii) F為一右連續函數;

(iv) $\lim_{x \rightarrow-\infty} F(x)=0$, 且$\lim_{x \rightarrow \infty} F(x)=1$。

在上述定理中, 條件(i)其實可以不用, 因(ii)與(iv)便導致(i)成立。不過我們還是寫出來, 使分佈函數的性質更清楚。

反之, 若一函數具有定理1中之(ii)-(iv)三個性質, 必為某一隨機變數之

分佈函數, 見下定理, 證明則略去。

<u>**定理2.2**</u> 設G為一非漸減且右連續之函數, 並滿足$\lim_{x\to-\infty} G(x)=0$, $\lim_{x\to\infty} G(x)=1$。則存在一隨機變數X, 以G為其分佈函數。

由於定理2, 任一非漸減、右連續, 且滿足$\lim_{x\to-\infty} G(x)=0$, 及$\lim_{x\to\infty} G(x)=1$之函數, 皆可視為分佈函數。不過定理2是針對存在性, 事實上, 以G為分佈函數之隨機變數並不唯一。讀者在上一節已看到, 樣本空間不必須都是區間$(0,1)$。

如果我們只討論機率, 便只須知隨機變數之分佈, 機率空間並無關緊要。甚至有時隨機變數的"名字"亦不重要, 稱為X, 稱為Y皆可, 出身(定義在那一機率空間)不同也沒關係。不同的隨機變數, 可以有相同的分佈函數(見底下例9)。X,Y,Z等, 在此不過是一個"名字"而已。

設隨機變數X之分佈函數為F, 則對$\forall a<b$, 因$\{X\le a\}\subset\{X\le b\}$, 故由機率函數的性質得

$$\begin{aligned}P(a<X\le b) &= P(\{X\le b\}\setminus\{X\le a\})=P(X\le b)-P(X\le a)\\ &= F(b)-F(a)\text{。}\end{aligned}$$

我們以F表示出一些與X有關之常見事件的機率如下: 先令$F(x-)=\lim_{t\uparrow x}F(t), x\in R$, 表$F$在$x$之**左極限值**。如果$F$在$x$連續, 則$F(x-)=F(x)$。$F(x)-F(x-)$表$F$之圖形在$x$跳升之高度。對任二$a,b\in R$,

$$\begin{aligned}&P(X=a)=F(a)-F(a-),\\ &P(X>a)=1-F(a),\\ &P(X\ge a)=1-F(a-)=1-F(a)+P(X=a),\\ &P(X\le b)=F(b),\\ &P(X<b)=F(b-)=F(b)-P(X=b)\text{。}\end{aligned}$$

又對$a<b$,

$$\begin{aligned}&P(a<X\le b)=F(b)-F(a),\\ &P(a<X<b)=F(b-)-F(a)=F(b)-F(a)-P(X=b),\\ &P(a\le X\le b)=F(b)-F(a-)=F(b)-F(a)+P(X=a),\end{aligned}$$

$$P(a \le X < b) = F(b-) - F(a-) = F(b) - F(a) + P(X=a) - P(X=b) \text{ 。}$$

例7. 持續投擲一出現正面的機率為p之銅板, $0 < p < 1$, 直到出現一正面才停止。令X表總共之投擲數。則

$$P(X = x) = p(1-p)^{x-1}, x \ge 1 \text{ 。} \tag{4}$$

因此對$x = 1, 2, \cdots$,

$$\begin{aligned} F(x) = P(X \le x) &= \sum_{i=1}^{x} P(X = i) = \sum_{i=1}^{x} p(1-p)^{i-1} \\ &= 1 - (1-p)^x \text{ 。} \end{aligned}$$

F之圖形仍為階梯函數, 在兩整數之間是水平的。一隨機變數, 若具有這種分佈函數, 稱為有**幾何分佈**(geometric distribution)。

有了F, 可得X落在任一區間之機率。如

$$\begin{aligned} P(3.1 < X \le 7.2) &= P(3 < X \le 7) = F(7) - F(3) \\ &= (1 - (1-p)^7) - (1 - (1-p)^3) \\ &= (1-p)^3 - (1-p)^7 \text{ 。} \end{aligned}$$

本例亦顯示, 對一分佈函數F, 不一定存在一$x \in R$, 使得$F(x) = 1$。當然也不一定存在一$x \in R$, 使得$F(x) = 0$。

底下為分佈函數連續的例子。

例8. 設

$$F_1(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \le x < 1, \\ 1, & x \ge 1 \text{ 。} \end{cases} \quad F_2(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \ge 0 \text{ 。} \end{cases}$$

圖2分別給出F_1及F_2之圖形。F_1及F_2為連續函數, 滿足定理1中之4個條件, 故皆為分佈函數。分佈函數如F_1, 便稱有在區間$(0,1)$之**均勻分**

佈(uniform distribution), 以$U(0,1)$表之; 分佈函數如F_2, 便稱有參數λ之**指數分佈**(exponential distribution), 以$\mathcal{E}(\lambda)$表之, 其中$\lambda > 0$ 。又我們也可定義在區間(a,b)之均勻分佈, 其中$a < b$, 並以$U(a,b)$表之, 即分佈函數$F(x) = (x-a)/(b-a), x \in [a,b)$, 且$F(x) = 0, x < a, F(x) = 1, x \geq b$ 。另外, 也可有$U[a,b)$, $U(a,b]$或$U[a,b]$分佈 。

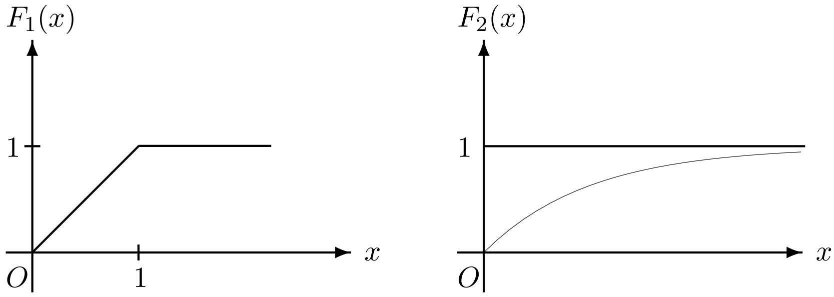


圖2 例8中$F_1(x)$及$F_2(x)$之圖形

隨機變數X之分佈函數F若為連續函數, 便稱X為**連續型**; 若F為階梯函數, 便稱X為**離散型** 。也可稱分佈函數為連續型或離散型 。不過也有既非連續型亦非離散型的隨機變數 。例如, 令

$$F(x) = \begin{cases} \frac{1}{2}\frac{1}{1+e^{-x}}, & x < 0, \\ \frac{1}{2} + \frac{1}{2}\frac{1}{1+e^{-x}} & ,x \geq 0 \text{ 。} \end{cases}$$

F滿足定理1中之4個條件, 故為分佈函數 。但F之圖形並非階梯函數, 且除了在$x = 0$有一高度為$1/2$之跳升(所以$P(X = 0) = 1/2$), 其餘地方為連續 。故F既非連續型, 亦非離散型 。

離散型隨機變數, 其分佈函數的圖形, 有跳升之處的集合, 必為可數的, 可能是一有限集合, 或可數的無限集合 。

有時我們給一分佈函數“$F(x), x > a$”, 其中a為一常數, 就表$F(x) = 0, \forall x \leq a$ 。又若說有一分佈函數“$F(x)$, $a < x < b$”, 就表$F(x) = 0$, $x \leq a$, 且$F(x) = 1$, $x \geq b$ 。其他形式(如$F(x)$, $x < b$)也有類似的說明 。

要注意的是, 不但在同一樣本空間, 可以定義出不同的隨機變數, 而分佈函數卻相同, 在不同的樣本空間, 亦可分別定義隨機變數, 且有相同的分佈函數 。

我們再給二隨機變數**分佈相同**(identically distributed)之定義。

定義1. 二隨機變數X, Y, 若其分佈函數相同, 即滿足

$$P(X \le x) = P(Y \le x), x \in R, \tag{5}$$

則稱**分佈相同**, 並以$X \stackrel{d}{=} Y$表之。

設以F, G分別表X及Y之分佈函數。X與Y分佈相同, 表$F(x) = G(x)$, $\forall x \in R$。也可說X與Y有**共同的分佈**。但分佈相同與隨機變數相同是不一樣的, 甚至二差異很大的隨機變數, 都可有相同的分佈函數。我們以下二例來說明。

例9. 投擲一公正的銅板3次, 令X表所得之正面數, Y表所得之反面數。則當然$X \ne Y$。因若$X = 0$則$Y = 3$。事實上, 可看出X與Y二"函數"不會有相等的時候, 即$X(\omega) \ne Y(\omega)$, $\forall \omega \in \Omega$, 其中$\Omega$見例1(如當$\omega =$正正正, 則$X(\omega) = 3 \ne 0 = Y(\omega)$)。但因

$$P(X = i) = P(Y = i), i = 0, 1, 2, 3,$$

故X與Y有相同的分佈函數, 即$X \stackrel{d}{=} Y$。

例10. 投擲一公正的銅板3次, 令Z表所得之正面數。投擲一公正的骰子3次, 令W表所得點數為偶數的次數。Z與W為兩個不相干的隨機變數, 二者當然不等。

隨機變數乃函數, 而二函數f與g相等的定義是要滿足

(i) f與g有相同的定義域,

(ii) 對每一定義域中的x, $f(x) = g(x)$。

現因Z的定義域為例1中之Ω, 共有8個元素, 而W之定義域為

$$\{111, 112, \cdots, 666\},$$

共有216個元素, 其中ijk表投擲骰子第1次得i, 第2次得j, 第3次得k, $1 \le$

$i, j, k \leq 6$, W與Z之定義域不同, 故$Z \neq W$。但明顯可看出

$$P(Z = i) = P(W = i), i = 0, 1, 2, 3,$$

故$Z \stackrel{d}{=} W$。

在例9中, X與Y雖為不同的隨機變數(只是分佈相同), 但二者的關係卻知道, 即$X + Y = 3$, 或者說$Y = 3 - X$。我們也可求諸如$X > Y$, $X = Y + 1$等事件之機率。但在例10中, $Z = 3$時, W是多少, 是完全不知道的。而事件$Z > W$, 或$Z = W + 1$的機率, 也無法求得。這是屬於兩個隨機變數的問題, 要知道Z與W之聯合分佈(見第5節)才行。光知道Z的分佈, 及W的分佈, 並無法求同時與Z, W有關的機率。

雖$X \stackrel{d}{=} Y$時, 不一定導致$X = Y$, 但$X = Y$, 卻必導致$X \stackrel{d}{=} Y$。此因$X = Y$(二函數相等), 表$X(\omega) = Y(\omega)$, $\forall \omega \in \Omega$, 故對$\forall x \in R$,

$$P(X \leq x) = P(\{\omega | X(\omega) \leq x\}) = P(\{\omega | Y(\omega) \leq x\}) = P(Y \leq x)\text{。}$$

二隨機變數X, Y, 亦有可能滿足諸如$X < Y$, $X < Y + 1$, $X > 3Y$, 或$X^2 = Y$等。例如, 若令$Y = X + 1$, 則有$X < Y$。記住隨機變數即函數, 因此函數會有的性質或關係, 隨機變數自然也可以有。

對一隨機變數X, 除了伴隨分佈函數F, 亦有一與其關係密切的函數, 即**機率密度函數**(probability density function), 或是**機率質量函數**(probability mass function), 前者是對連續型, 後者是對離散型。不過為了簡便, 通常我們不分連續或離散, 皆說是機率密度函數。底下我們給其定義。

設X為一隨機變數, 以F為分佈函數。若X為離散型, 則下述函數f(若為區隔可寫f_X), 便稱為X之機率密度函數:

$$f(x) = P(X = x), x \in R\text{。} \tag{6}$$

由於離散型隨機變數, 其分佈函數的圖形, 有跳升之處的集合為可數的, 而此集合就是使$P(X = x)$為正之所有x的集合。設以C表此集

合, 則對$\forall x \notin C$, $f(x) = 0$。因此離散型的機率密度函數, 有時表示成$f(x), x \in C$, 其中C為某一可數的集合, 可能是$\{1, 2, \cdots, 10\}$, 也可能是非負整數的集合$\{0, 1, 2, \cdots\}$等。至於$x \notin C$時的$f(x)$往往不提, 因其值皆為0。所以即是將此機率密度函數的定義域取成C。可看出對一實數上的集合A,

$$P(A) = P(X \in A) = \sum_{x \in A \cap C} f(x),$$

且

$$F(x) = \sum_{t \le x, t \in C} f(t), x \in R \text{ 。}$$

例11. 設隨機變數X之機率密度函數為

$$f(x) = \begin{cases} (8/15)(1/2)^x, & x = 0, 1, 2, 3, \\ 0, & \text{其他 。} \end{cases}$$

則可求出其分佈函數為

$$F(x) = \begin{cases} 0, & x < 0, \\ 8/15, & 0 \le x < 1, \\ 12/15, & 1 \le x < 2, \\ 14/15, & 2 \le x < 3, \\ 1, & x \ge 3 \text{ 。} \end{cases}$$

為一階梯函數。

又在求$F(2.5)$時, 可依定義得

$$\begin{aligned} F(2.5) &= \sum_{x \le 2.5} f(x) = f(0) + f(1) + f(2) \\ &= \frac{8}{15} + \frac{4}{15} + \frac{2}{15} = \frac{14}{15}, \end{aligned}$$

或利用已求出之$F(x)$, 因$2.5 \in [2, 3)$, 故$F(2.5) = 14/15$。

圖3給出$f(x)$及$F(x)$之圖形。

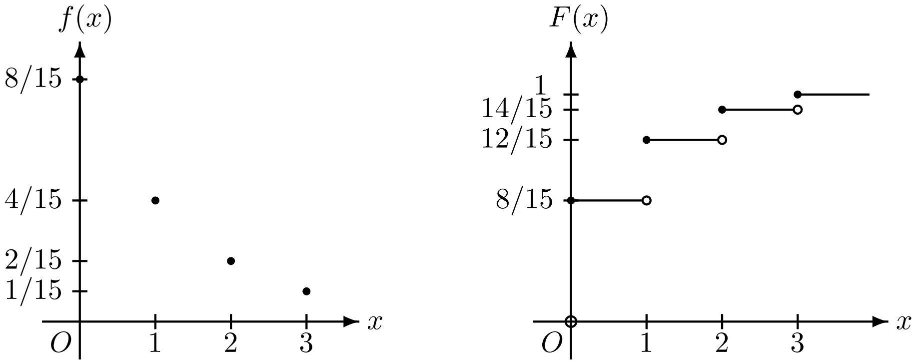


圖3 例11中$f(x)$及$F(x)$之圖形

離散型機率密度函數之圖形, 有時以線段表之。而對只取整數值之隨機變數, 其機率密度函數的圖形, 有時以一些小長方形表示, 每一小長方形之底部中點為取值處, 高度為該點之機率, 因此小長方形面積和為1。這樣表示機率密度函數的圖形, 稱為**直方圖**(histogram)。例11中之$f(x)$亦可繪成如圖4中之二圖, 右圖即直方圖。

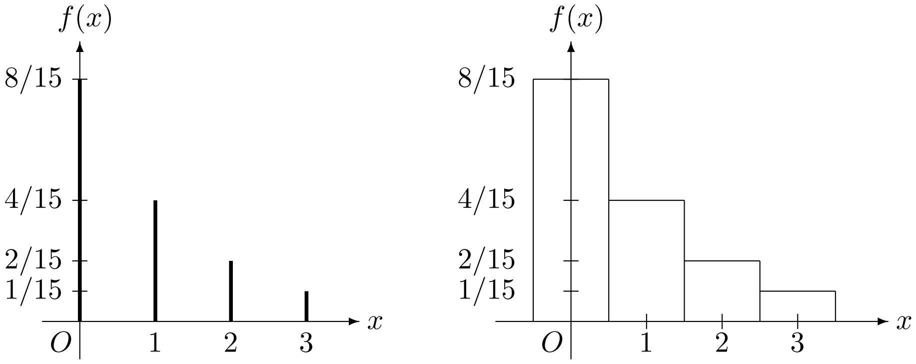


圖4 例11中$f(x)$之圖形

其次若X為連續型, 則其機率密度函數, 乃定義為滿足下式之非負函數f:

(7)
$$F(x)=\int_{-\infty}^{x} f(t)dt, x\in R。$$

由定理1, 因$\lim\limits_{x\to\infty} F(x)=1$, 故知$f$除了非負之外, 尚須滿足

$$\int_{-\infty}^{\infty} f(x)dx = 1 \text{。} \tag{8}$$

即$f(x)$之圖形與x軸間之面積為1。

對連續型的隨機變數X, 由於$P(X=x)=0$, $\forall x\in R$, 即X取任一定值的機率為0。故在求X落在某一區間的機率時, 該區間是否包含端點是無關緊要的。即對$a<b$,

$$\begin{aligned} P(a<X<b) &= P(a\le X<b)=P(a\le X\le b) \\ &= P(a<X\le b)=F(b)-F(a) \text{。} \end{aligned}$$

若以機率密度函數表示, 上述任一機率皆等於

$$P(a<X<b) = \int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_{a}^{b} f(t)dt \text{。}$$

連續型隨機變數, 其機率密度函數, 也可有寫成$f(x)$, $c\le x\le d$, 的型式, 即表$f(x)=0$, $\forall x>d$或$x<c$, 即所對應的隨機變數, 取值不會大於d也不會小於c。

又由**微積分基本定理**(fundamental theorem of calculus), 若f為連續函數, 則由(7)式,

$$F^{'}(x)=f(x), x\in R \text{。} \tag{9}$$

可藉由上式得到此時之機率密度函數。又二隨機變數, 若有相同的機率密度函數, 則其分佈相同。

<u>例12</u>. 在例8中, 對應F_1及F_2之機率密度函數分別為

$$f_1(x)=\begin{cases} 0, & x\le 0, \\ 1, & 0<x<1, \\ 0, & x\ge 1; \end{cases} \quad f_2(x)=\begin{cases} 0, & x<0, \\ \lambda e^{-\lambda x}, & x\ge 0 \text{。} \end{cases}$$

或簡單地寫成$f_1(x)=1,\ 0<x<1,\ f_2(x)=\lambda e^{-\lambda x},\ x\geq 0$。對連續型的隨機變數, 機率密度函數在有限個點的值改變, 是不用在意的。因此前述f_1寫成

$f_1(x)=1, 0\leq x<1,\quad f_1(x)=1, 0<x\leq 1,\quad$ 或$f_1(x)=1, 0\leq x\leq 1,$

等皆可。這3個f_1亦皆滿足$F_1(x)=\int_{-\infty}^{x}f_1(t)dt,\ x\in R$, 因此皆可當做對應$F_1$之機率密度函數。

必須一提的是, 連續型的分佈不一定有機率密度函數, 此因一連續函數不一定可微。也就是有可能分佈函數F為連續, 但不存在函數f, 使得(7)式成立。

由分佈函數可得到機率密度函數, 反之亦有下述定理。

定理2.3 一函數f若滿足下述二條件, 必為某一隨機變數之機率密度函數:

(i) $f(x)\geq 0, \forall x\in R,$

(ii) $\sum_x f(x)=1$(離散型), 或$\int_{-\infty}^{\infty}f(x)dx=1$(連續型)。

由上述定理知, 機率密度函數是可以很容易得到的, 對一非負的函數g, 且$\sum_x g(x)$(離散型), 或$\int_{-\infty}^{\infty}g(x)dx$為正且有限, 則經除以該和或積分的值, 便立即可得一機率密度函數, 見下例。

例13. 設$g(x)=3^{-x}, x=1,2,3,\cdots$。因$\sum_{x=1}^{\infty}g(x)=1/2$, 故$f(x)=2g(x)=2\cdot 3^{-x}, x=1,2,3,\cdots$, 即成為一機率密度函數。

另外, 設$g_1(x)=xe^{-3x}, x>0$。因$\int_0^{\infty}xe^{-3x}dx=1/9$, 故$f_1(x)=9g_1(x)=9xe^{-3x}, x>0$, 亦成為一機率密度函數。假設隨機變數$X$以$f_1$為機率密度函數, 則分佈函數$F(x)=P(X\leq x)=\int_0^x f_1(u)du=1-e^{-3x}-3xe^{-3x},\ x\geq 0$。

對連續型的隨機變數, 其機率密度函數$f(x)$與x軸間之面積等於1; 對離散型的隨機變數, 其機率密度函數之高度和等於1。通常我們較少去繪隨機變數的圖形(雖然此為一函數), 但分佈函數及機率密度函數的圖形倒

是常繪的。

對一隨機變數X, 以$F(x)$為分佈函數, $f(x)$為機率密度函數, 有時可以$X \sim F(x)$或$X \sim f(x)$表之。至於若$X \overset{d}{=} Y$, 也可以$X \sim Y$表之。對一隨機變數X, 我們常會說, X遵循某一分佈, 或說X之機率法則為$\cdots$。分佈就如同隨機變數, 機率上的**法則**(law), 是隨機變數該遵循的。而給定機率密度函數, 或分佈函數, 或任何其他可以唯一決定分佈函數之**轉換**(見第4節)等, 皆稱為**給定分佈**, 或說**決定分佈**。

3 變數代換

在很多試驗裡, 對一隨機變數X, 有時我們可能會對X之某一函數有興趣。譬如說量測得到圓的半徑X, 但我們想知道圓的面積$Y = \pi X^2$。即使不論實務上的需要, 在第1節我們已指出, 在數學裡我們常會討論與一變數有關的函數之性質或行為。對一隨機變數X, 有時我們會想知道它的一個函數$Y = g(X)$的行為。Y便稱為X之**變數代換**(change of variable)。而由X至$Y = g(X)$, 也可視為一種隨機變數之**轉換**(transformation of a random variable)。Y仍為一隨機變數。若X的分佈知道, 則因Y的值是由X的值所決定, 故Y的分佈, 可由所給之X的分佈來決定。

對每一實數的子集合A,

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A))\text{。} \tag{10}$$

在此

$$g^{-1}(A) = \{x | x \in R, g(x) \in A\}\text{。}$$

對於新的隨機變數Y, 由(10)式, 欲求Y落在某一集合A之機率, 仍要回到關於X的機率。但若我們求出Y的分佈函數(或機率密度函數), 當然就可以不用理會X了。

若X為離散型的隨機變數, 則$Y = g(X)$的機率密度函數, 只要留意由X至Y的變換是否為$1-1$(等價於g是否為嚴格單調函數), 則藉由(10)式,

通常可很快地得到。底下為一典型的例子。

例14. 設X為一離散型的隨機變數, 且機率密度函數為

$$f_X(x)=\begin{cases}0.15, & x=0,3,\\ 0.20, & x=1,2,\\ 0.30, & x=4,\\ 0, & \text{其他}。\end{cases}$$

令$Y=X^2$, $Z=(X-2)^2$。我們分別來決定Y及Z之分佈。

首先X之可能的值為$0,1,2,3,4$, Y之可能的值則為$0,1,4,9,16$。因X只取非負的值, 故由X至Y的變換為$1-1$(雖$g(x)=x^2$, $x\in R$, 並非嚴格單調函數, 但在$x\geq 0$處, g為嚴格單調)。又

$$P(Y=y)=P(X^2=y)=P(X=\sqrt{y}), y=0,1,4,9,16\text{。}$$

故得Y之機率密度函數為

$$f_Y(y)=P(Y=y)=\begin{cases}0.15, & y=0,\\ 0.20, & y=1,\\ 0.20, & y=4,\\ 0.15, & y=9,\\ 0.30, & y=16,\\ 0, & \text{其他}。\end{cases}$$

可驗證$\sum_y f_Y(y)$仍等於1。

其次Z之可能的值為$0,1,4$。$X=0$及4, 皆對應$Z=4$; $X=1$及3, 皆對應$Z=1$; $X=2$, 則對應$Z=0$。故由X至Z, 並非$1-1$的變換。Z之機率密度函數為

$$f_Z(z)=P(Z=z)=\begin{cases}f_X(2)=0.20, & z=0,\\ f_X(1)+f_X(3)=0.35, & z=1,\\ f_X(0)+f_X(4)=0.45, & z=4\text{。}\end{cases}$$

可驗證$\sum_z f_Z(z)$仍等於1。

底下再給一例。

例15. 設隨機變數X之機率密度函數為

$$f(x)=\frac{1}{31}2^{4-x}, x=0,1,2,3,4\text{。}$$

令$Y=X^2+1$。此為由X至Y之$1-1$變換, $X=\sqrt{Y-1}$。Y取值在$1,2,5,10,17$。故Y之機率密度函數為

$$f_Y(y)=\frac{1}{31}2^{4-\sqrt{y-1}}, y=1,2,5,10,17\text{。}$$

其次令$Z=g(X)=X^3-3X^2+X+2$。則因

$$g(0)=2,\quad g(1)=1,\quad g(2)=0,\quad g(3)=5,\quad g(4)=22,$$

故此仍為由X至Z之$1-1$變換。雖不易給出$X=g^{-1}(Z)$之g^{-1}的型式, 但仍可寫出Z之機率密度函數如下:

$$\begin{aligned}
f_Z(0)&=P(X=2)=\frac{4}{31},\\
f_Z(1)&=P(X=1)=\frac{8}{31},\\
f_Z(2)&=P(X=0)=\frac{16}{31},\\
f_Z(5)&=P(X=3)=\frac{2}{31},\\
f_Z(22)&=P(X=4)=\frac{1}{31}\text{。}
\end{aligned}$$

要知所謂函數, 只要每一定義域中的元素, 在對應域中恰有一元素與其對應即可, 不一定要能給出一個適用定義域中每一元素的型式。

其次我們給二連續型隨機變數的例子。

例16. 設隨機變數X有$\mathcal{U}(0,1)$分佈。即其機率密度函數為$f(x)=1,\ 0<x<1$。令$Y=-\lambda^{-1}\log(1-X)$, 其中$\lambda>0$為一常數。機率與統計裡出現

的對數符號log, 通常為自然對數, 即以e為底。此為一由X至Y之$1-1$變換。由於X介於$0,1$之間, 故Y取值在$(0,\infty)$。則對$\forall y>0$,

$$\begin{aligned} F_Y(y)=P(Y\le y) &= P(-\lambda^{-1}\log(1-X)\le y)=P(\log(1-X)\ge -\lambda y) \\ &= P(X\le 1-e^{-\lambda y})=1-e^{-\lambda y}, y>0 \text{。} \end{aligned}$$

因此Y之機率密度函數為

$$f_Y(y)=\frac{dF_Y(y)}{dy}=\begin{cases} \lambda e^{-\lambda y}, & y>0, \\ 0, & y\le 0 \text{。} \end{cases}$$

即Y有$\mathcal{E}(\lambda)$分佈。

例17. 設隨機變數X之機率密度函數為

$$f_X(x)=\frac{2}{\pi(1+x^2)}, x>0 \text{。}$$

令$Y=1/X$。則此為一由X至Y之$1-1$變換, 且Y取值在$(0,\infty)$。我們想決定Y之機率密度函數。

對$\forall y>0$,

$$\begin{aligned} F_Y(y) &= P(Y\le y)=P(1/X\le y)=P(X\ge 1/y)=P(X>1/y) \\ &= 1-P(X\le 1/y)=1-F_X(1/y) \text{。} \end{aligned}$$

因此

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy}=f_X(\frac{1}{y})\cdot\frac{1}{y^2}=\frac{2}{\pi(1+1/y^2)}\cdot\frac{1}{y^2} \\ &= \frac{2}{\pi(1+y^2)}, y>0 \text{。} \end{aligned}$$

由於X與Y之機率密度函數相同, 故X與Y有相同的分佈, 即$X\stackrel{d}{=}Y$。對此特別的分佈, 雖經過一倒數的變換, 所得之新隨機變數, 其分佈卻未改變。

在上述推導中, 由於只是要求Y之機率密度函數, 所以我們並未將$F_Y(y)$明確地寫出。事實上

$$\begin{aligned} F_Y(y) &= 1 - F_X(1/y) = 1 - \int_0^{1/y} f_X(x)dx \\ &= 1 - \int_0^{1/y} \frac{2}{\pi(1+x^2)}dx = 1 - \frac{2}{\pi}\arctan(\frac{1}{y}), y > 0 \text{。} \end{aligned}$$

將此$F_Y(y)$對y微分, 便得到前述$f_Y(y)$, $y > 0$。

在例17的推導過程中, 有

$$f_Y(y) = f_X(\frac{1}{y}) \cdot \frac{1}{y^2} \text{。}$$

即Y之機率密度函數, 可將X之機率密度函數中的參數x, 以$1/y$取代, 再乘上$1/y$對y微分的絕對值。而由$Y = g(X) = 1/X$, 解出

$$X = g^{-1}(Y) = \frac{1}{Y} \text{。}$$

故$f_Y(y)$可寫成

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \text{。}$$

上述這些討論與觀察, 便引出底下兩個一般的定理。

定理3.4 設隨機變數X之分佈函數為F_X, 機率密度函數為f_X。令$Y = g(X)$, Y之分佈函數以F_Y表之。又令

(11) $$\Omega_1 = \{x | f_X(x) > 0\},$$

(12) $$\Omega_2 = \{y | \text{存在} x \in \Omega_1, \text{使得} y = g(x)\} \text{。}$$

(i) 若g在Ω_1為嚴格漸增函數, 則$F_Y(y) = F_X(g^{-1}(y))$, $\forall y \in \Omega_2$;

(ii) 若g在Ω_1為嚴格漸減函數, 則$F_Y(y) = 1 - F_X(g^{-1}(y)-)$, $\forall y \in \Omega_2$。

定理4對連續及離離散型的隨機變數皆適用。當Y之機率密度函數為

連續, 則利用定理4, 可由微分分佈函數而得。我們列成一定理。

定理3.5 設隨機變數X之機率密度函數為f_X。令$Y = g(X)$, 其中g為一嚴格單調函數, Ω_1及Ω_2分別定義在(11)及(12)式中。設f_X在Ω_1連續, 且g^{-1}在Ω_2有一連續的導數, 則Y之機率密度函數為

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|, y \in \Omega_2, \tag{13}$$

且$f_Y(y) = 0$, $y \notin \Omega_2$。

例18. 設X有$\mathcal{E}(\lambda)$分佈。令$Y = \sqrt{X}$。則$\Omega_1 = \Omega_2 = (0, \infty)$, 且$X$之機率密度函數$f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, 在Ω_1中連續; $g^{-1}(y) = y^2$在Ω_2之導數$2y$為連續。故Y之機率密度函數為$f_Y(y) = f(y^2)\cdot 2y = 2\lambda y e^{-\lambda y^2}, y > 0$。$Y$稱為具有參數$\lambda$之**瑞萊分佈**(Rayleigh distribution)。

在做變數代換時, 函數g有時不為嚴格單調, 因此定理5就不適用。但往往會有g在一個個的集合中, 分別為嚴格單調(想想諸如多項函數、正弦及餘弦函數的圖形), 可參考一般初等的機率或統計的書。因此在每一個集合中, 皆可解出$X = g^{-1}(Y)$, 則仍有辦法求出Y之分佈。

4 期望值

對一隨機變數, 我們常想粗略地知道其值究竟多大, **期望值**(expectation, 或稱expected value, mean), 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。

設隨機變數X之機率密度函數為f, 則其期望值以$E(X)$表示, 定義為

$$E(X) = \begin{cases} \int_{-\infty}^{\infty} x f(x)dx, & \text{連續型}, \\ \sum_x x f(x), & \text{離散型}, \end{cases} \tag{14}$$

只要上述積分或和存在。此處$\sum_x x f(x)$表對所有可能的x(即$f(x) \neq 0$)相加。通常以μ_X(或只以μ)表$E(X)$。

假設有10個數: $1,1,2,2,2,3,3,3,4,4$, 則其平均為

$$\frac{1}{10}(1+1+2+2+2+3+3+3+4+4)=2.5\text{ 。}$$

由於數字中有重複: 有2個1, 3個2, 3個3, 2個4, 所以平均亦可表為

$$\frac{1}{10}(1\cdot 2+2\cdot 3+3\cdot 3+4\cdot 2)=2.5,$$

所得相同。第二種作法是我們常採用的, 特別是在數字較多或較大時。又上式左側可改寫為

$$1\cdot\frac{2}{10}+2\cdot\frac{3}{10}+3\cdot\frac{3}{10}+4\cdot\frac{2}{10},$$

其中2/10, 3/10, 3/10, 2/10, 可分別視為10個數字中, $1,2,3,4$所佔的份量。所以第二種作法乃是一種加權平均。隨機變數期望值的定義, 便是基於加權平均的想法。這點由離散型隨機變數期望值的定義, 特別可以看出來。

至於X之一函數$g(X)$的期望值, 則定義為

$$E(g(X))=\begin{cases}\int_{-\infty}^{\infty}g(x)f(x)dx, & \text{連續型},\\ \sum_x g(x)f(x), & \text{離散型 。}\end{cases} \tag{15}$$

亦可經由變數代換, 先求出$Y=g(X)$之機率密度函數, 再利用(14)式而求得$E(Y)$。

底下給幾個例子。

例19. 設離散型隨機變數X之機率密度函數為

$$f(0)=\frac{1}{16},\quad f(1)=\frac{2}{16},\quad f(2)=\frac{3}{16},\quad f(3)=\frac{4}{16},$$
$$f(4)=\frac{4}{16},\quad f(5)=\frac{1}{16},\quad f(6)=\frac{1}{16}\text{ 。}$$

則

$$E(X)=\sum_{x=0}^{6}xf(x)=\frac{47}{16}\text{ 。}$$

又

$$E((X-3)^2)=\sum_{x=0}^{6}(x-3)^2 f(x)=\frac{37}{16}\ 。$$

在求$E((X-3)^2)$時, 也可利用變數代換。令$Y=(X-3)^2$。則Y取值0, 1, 4, 9, 且

$$\begin{aligned}
f_Y(0)&=P(X=3)=\frac{4}{16},\\
f_Y(1)&=P(X=2)+P(X=4)=\frac{7}{16},\\
f_Y(4)&=P(X=1)+P(X=5)=\frac{3}{16},\\
f_Y(9)&=P(X=0)+P(X=6)=\frac{2}{16}\ 。
\end{aligned}$$

故

$$E((X-3)^2)=E(Y)=\sum_y y f_Y(y)=\frac{37}{16}\ 。$$

其中$\sum$是對所有的y相加, 即$0,1,4,9$。

例20. 設隨機變數X之機率密度函數為$f(x)=2x, x\in[0,1]$。則

$$\begin{aligned}
E(X)&=\int_0^1 x\cdot 2x dx=\int_0^1 2x^2 dx=\frac{2}{3},\\
E(X^2)&=\int_0^1 x^2\cdot 2x dx=\int_0^1 2x^3 dx=\frac{1}{2}\ 。
\end{aligned}$$

另外, 亦可利用變數代換。令$Y=X^2$, 先求出Y之機率密度函數為

$$g(y)=2\sqrt{y}\cdot\frac{1}{2\sqrt{y}}=1, y\in[0,1]\ 。$$

如此

$$E(X^2)=E(Y)=\int_0^1 y\cdot 1dy=\frac{1}{2},$$

仍得到相同的結果。

例21. 在微積分裡有$\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$。故

$$f(i) = \frac{6}{\pi^2}\frac{1}{i^2}, i = 1, 2, \cdots, \tag{16}$$

為一機率密度函數。但因$\sum_{i=1}^{\infty} 1/i = \infty$, 故若隨機變數$X$之機率密度函數如(16)式, 則$X$之期望值不存在。

另外, 因$\sum_{i=1}^{\infty} 1/i^4 = \pi^4/90$, 故

$$g(i) = \frac{90}{\pi^4}\frac{1}{i^4}, i = 1, 2, \cdots, \tag{17}$$

為一機率密度函數。又設隨機變數Y之機率密度函數如(17)式所給。則

$$E(Y) = \frac{90}{\pi^4}\sum_{i=1}^{\infty}\frac{1}{i^3} < \infty,$$
$$E(Y^2) = \frac{90}{\pi^4}\sum_{i=1}^{\infty}\frac{1}{i^2} = \frac{90}{\pi^4}\frac{\pi^2}{6} = \frac{15}{\pi^2} < \infty。$$

即$E(Y)$及$E(Y^2)$皆存在。當然$E(Y^3)$便不存在了。

例22. 設隨機變數X有**標準的柯西分佈**(Cauchy distribution,, Augustin-Louis Cauchy, 1789-1857, 為法國著名數學家), 以$\mathcal{C}(0,1)$表之。即其機率密度函數為

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

則X之期望值不存在。此因

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\pi(1+x^2)}dx,$$

為一**瑕積分**(improper integral)。而依定義, 此瑕積分存在, 若且唯若

$$\int_0^{\infty}\frac{x}{\pi(1+x^2)}dx, \quad \int_{-\infty}^{0}\frac{x}{\pi(1+x^2)}dx$$

皆要存在。但此二積分皆不存在, 分別為∞及$-\infty$。故$E(X)$不存在。

由上述瑕積分存在的定義, 可看出$E(X)$存在, 若且唯若$E(|X|)$存在。期望值有不少性質。例如, **一有界的**(bounded)隨機變數, 期望值必存在。在此對一隨機變數X, 若存在一常數$M > 0$, 使得$P(|X| \le M) = 1$, 便稱該隨機變數為**有界的**。若X有$\mathcal{U}(0,1)$分佈, 便為有界; 有$\mathcal{C}(0,1)$分佈, 則非有界。我們再列出一些性質如下。

<u>**定理4.6**</u> 設X為一隨機變數, a, b, c為常數, g_1, g_2為二函數, 使得$E(g_1(X))$, $E(g_2(X))$皆存在。則

(i) $E(ag_1(X) + bg_2(X) + c) = aE(g_1(X)) + bE(g_2(X)) + c$;

(ii) 若$g_1(x) \ge 0, \forall x \in R$, 則$E(g_1(X)) \ge 0$;

(iii) 若$g_1(x) \ge g_2(x), \forall x \in R$, 則$E(g_1(X)) \ge E(g_2(X))$;

(iv) 若$a \le g_1(x) \le b, \forall x \in R$, 則$a \le E(g_1(X)) \le b$。

由上定理之(i), 特別地, 只要$E(X)$存在, 則

$$E(aX + b) = aE(X) + b,\ E(b) = b \text{。} \tag{18}$$

即一常數之期望值等於該常數。由於$E(X)$為一常數(已非隨機了), 故其期望值仍為本身。即$E(E(X)) = E(X)$。

期望值這個名詞也許會令有些人感到困擾。投擲一公正的骰子一次, 令X表所得之點數。則X之機率密度函數為$f(x) = 1/6, x = 1, 2, \cdots, 6$。因此

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{21}{6} = 3.5 \text{。}$$

投擲一骰子, 點數之期望值為3.5。點數為整數, 無論怎麼投擲都得不到3.5。那怎會期望得到3.5呢?

就暫且不要去想期望值字面上的含義。不過期望值是一隨機變數之“最佳”常數的代表值。所謂最佳, 指的是誤差平方的期望值最小。見下

定理。

<u>定理4.7</u> 設隨機變數X滿足$E(X^2) < \infty$。則對所有$a \in R$, $E((X-a)^2)$之極小值發生在$a = E(X)$處。

果園裡生產的蘋果, 大小頗有差異。有人問你, 蘋果每個重量大約是多少, 如何回答呢? 人們常想以一常數a來代表一隨機變數X。很難說那一個a最好, 誤差$X - a$仍是隨機的。但在誤差平方$(X-a)^2$之期望值最小下, 上定理指出, 期望值$E(X)$便是最好的選擇。

其次我們定義**動差**(moment)。設n為一整數, 則若$E(X^n)$存在, 便稱此為隨機變數X之**n次動差**。而$E((X-\mu)^n)$稱為X的**n次中心動差**(nth central moment), 其中$\mu = E(X)$。特別地, 二次中心動差稱為X的**變異數**(variance), 以σ_X^2, $\mathrm{Var}(X)$或簡單地以σ^2表之, 即

$$\mathrm{Var}(X) = E((X - E(X))^2)\text{。} \tag{19}$$

$\mathrm{Var}(X)$之正的平方根, 以σ_X或σ表之, 稱為X之**標準離差**(standard deviation), 或**標準差**、**均方差**。另外, 如同期望值, 對於n次動差, 仍有$E(X^n)$存在, 若且唯若$E(|X|^n)$存在。

在機率論裡我們引進了隨機變數的概念。例如, 學生不再如小學的練習裡, 每位體重40公斤, 求25位學生的總體重, 而是讓每位學生的體重有一機率分佈。雖然這樣的模式較合理, 但因無法掌握隨機的量之大小, 我們才又想要有一代表值。由於具有很多好的性質, 期望值遂常被拿來當做隨機變數之代表值。期望值像是隨機變數分佈之一個核心, 隨機變數可能取的值, 散佈在期望值的左右。其他亦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)等。眾數早期稱為**密集數**。在此一實數m若滿足

$$P(X \le m) \ge \frac{1}{2}, \quad \text{且} P(X \ge m) \ge \frac{1}{2}, \tag{20}$$

便稱為隨機變數X(或其分佈)之一**中位數**。依此定義中位數不一定唯一。當然如果X之分佈函數為連續且嚴格漸增, 則中位數便唯一。此時

中位數為唯一滿足$F(x)=0.5$之x。若X之分佈對稱於m, 則X之中位數為m。另外, 若$f(a)$為X之機率密度函數的一絕對極大值, 便稱a為X(或其分佈)之一**眾數**。對X為離散型的隨機變數, 其眾數表X最可能取的值。眾數不一定存在也不一定唯一。而且中位數及眾數有時雖存在但卻不易求出。

例23. 設X為一離散型的隨機變數, 且$P(X=0)=1/4$, $P(X=1)=1/2$, $P(X=2)=1/4$。則X有唯一的中位數1:

$$P(X\leq 1)=P(X=0)+P(X=1)=\frac{3}{4}\geq\frac{1}{2},$$
$$P(X\geq 1)=P(X=1)+P(X=2)=\frac{3}{4}\geq\frac{1}{2}。$$

故1為一中位數。又任一$a<1$,

$$P(X\leq a)\leq P(X=0)=\frac{1}{4}<\frac{1}{2},$$

任一$a>1$,

$$P(X\geq a)\leq P(X=2)=\frac{1}{4}<\frac{1}{2},$$

皆非中位數。

另外, 可看出X之眾數亦為1。

例24. 設X為一離散型的隨機變數, 且$P(X=0)=0.1$, $P(X=1)=0.4$, $P(X=2)=0.4$, $P(X=3)=0.1$。則任一$a\in[1,2]$皆為中位數, 而$1,2$皆為眾數。

例25. 設隨機變數X之分佈函數為$F(x)=x^2, 0\leq x<1$。令$x^2=1/2$, 解出$x=\sqrt{2}/2\doteq 0.707$。即X之中位數為$\sqrt{2}/2$。也可以說F之中位數為$\sqrt{2}/2$。若取X之機率密度函數為$f(x)=2x$, $0\leq x\leq 1$, 則X之眾數為1。若取X之機率密度函數為$f(x)=2x$, $0\leq x<1$, 則眾數不存在。

我們常要藉助統計做各種決策。投資某項事業, 可能成功而獲利若干, 可能失敗而損失若干。那到底要不要投資呢? 不能只看到成功機率很

大(說不定獲利很小), 也不能只看到獲利很大(說不定成功機率很低), 算一算淨所得之期望值, 是一簡單的決策依據。我們平常在判斷出門要不要帶雨傘, 也是先盤算一下期望損失(先估計下雨的機率, 及下雨不帶傘的損失等)。期望值在做決策時, 常扮演重要的角色。

<u>**例26.**</u> 某賭局的賭法如下: 投擲一公正的銅板, 若第n次投擲才首度出現正面, 則賭徒可獲得a^n元, 其中$0 < a \leq 2$, 為一定值。試依期望所得, 決定賭徒每次賭之前, 該付多少元, 此賭局才是公正。在此公正賭局, 乃照一般的認知, 指所付的錢等於期望所得, 即讓期望淨所得為0。

解.不難看出第n次投擲才首度出現正面之機率為2^{-n}, $n \geq 1$。令X表賭徒賭1次之所得。先設$0 < a < 2$, 則

$$E(X) = \sum_{n=1}^{\infty} a^n \cdot 2^{-n} = \sum_{n=1}^{\infty} (\frac{a}{2})^n = \frac{a/2}{1-a/2} = \frac{a}{2-a}\text{。}$$

至於若$a = 2$, 則

$$E(X) = \sum_{n=1}^{\infty} 2^n \cdot 2^{-n} = \sum_{n=1}^{\infty} 1 = \infty\text{。}$$

例如, 當$a = 1.999$時, $E(X) = 1{,}999$, 即每次賭之前要付1,999元。一般可能會認為每次賭能得到1.999^n元, 與得到2^n元, 無太大差異。的確, 例如, $n = 10$時, $1.999^n \doteq 1{,}018.89$, 而$2^{10} = 1{,}024$, 差異不大。但對前者, 每次賭只要付1,999元, 對後者, 則要付∞的錢。但有誰願意付(而且也沒有)∞的錢去玩一賭局呢? 再仔細觀察。對$a = 2$, 易知

$$P(\text{所得} \geq 2^n) = 2^{-n+1}, \text{ 且} P(\text{所得} \leq 2^{n-1}) = 1 - 2^{-n+1}, \ n \geq 1\text{。}$$

例如, 當$n = 11$, 有機率$1 - 2^{-10} \doteq 0.999$, 所得不超過1,024元。又當$n = 21$, 所得不超過$2^{20} = 1{,}048{,}576$(約百萬元)之機率, 約為0.999999, 已夠接近1了。要得到百萬元以上, 可說極不容易。但卻要先付∞的錢, 才能玩一次。但這可稱公正賭局呢! 這就是著名的**彼得堡詭論**(Petersburg paradox, Daniel Bernoulli, 1700-1782, 所提出)。當$a = 2$時, 如何設計出

一較合理的付費方式, 來參與此賭局, 可參考黃文璋(2010)pp.295-297。

不論期望值是多好的一個隨機變數之代表值, 難免會偏離該隨機變數, 如何量測偏差究竟有多大呢? 這就是變異數與標準差的功能。

設有一隨機變數X, 且設期望值$E(X)$存在。則X與$E(X)$之**離差**為$X-E(X)$。此值有正有負, 但期望值$E(X-E(X))=E(X)-E(X)=0$。離差之期望值為0, 此亦為期望值之一性質。但我們往往是對離差之絕對值較感興趣。此正如射飛鏢時, 有時上偏有時下偏, 總不能得意地說平均命中紅心。

以$|X-E(X)|$表離差之絕對值, 此量仍為一隨機變數, 如何衡量其大小呢? 取期望值為一辦法。也就是求量測離差之絕對值的"平均"。但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, 往往有些麻煩(想想求$\int_{-3}^{3}|\sin x-0.3x|dx$的例子)。對於$E(|X-E(X)|)$, 要討論何時$X\geq E(X)$, 何時$X<E(X)$, 這些不等式不見得好解。因此通常的作法, 是考慮離差平方之期望值, 而這就是**變異數**。由於是平方的期望值, 像是距離的平方。因此若將變異數開平方, 就像使其回到距離, **標準差**便產生了。

<u>**例27.**</u> 假設有A,B二項投資。投資A於一年後, 有1/2的機率賺一倍的錢, 有1/4的機率賠一半的錢, 有1/4的機率不賺不賠。投資B一年後, 有3/4的機率賺一半的錢, 有1/4的機率不賺不賠。若放10,000元於投資A, 則一年後之獲利X, 為一離散型的隨機變數, 機率密度函數為(略去"元", 底下同)

$$f_X(x)=\begin{cases}1/2, & x=10{,}000,\\ 1/4, & x=-5{,}000,\\ 1/4, & x=0。\end{cases}$$

若放10,000於投資B, 則一年後之獲利Y, 亦為一離散型的隨機變數, 機率密度函數為

$$f_Y(y)=\begin{cases}3/4, & y=5{,}000,\\ 1/4, & y=0。\end{cases}$$

依上述條件, 可得

$$E(X) = \sum_x x f_X(x) = \frac{1}{2} \cdot 10{,}000 + \frac{1}{4} \cdot (-5{,}000) + \frac{1}{4} \cdot 0 = 3{,}750 \text{ 。}$$

$$E(Y) = \sum_y y f_Y(y) = \frac{3}{4} \cdot 5{,}0000 + \frac{1}{4} \cdot 0 = 3{,}750 \text{ 。}$$

即投資A與投資B獲利之期望值相同 。雖然有相同的期望值, 但X之分佈較Y之分佈, 散佈的較開, 所以我們猜測X之變異數較大 。底下來驗證此猜測 。

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_x (x - 3{,}750)^2 f_X(x) \\
&= \frac{1}{2}(10{,}000 - 3{,}750)^2 + \frac{1}{4} \cdot (-5{,}000 - 3{,}750)^2 + \frac{1}{4}(0 - 3{,}750)^2 \\
&= 42{,}187{,}500 \text{ 。} \\
\mathrm{Var}(Y) &= \sum_y (y - 3{,}750)^2 f_Y(y) \\
&= \frac{3}{4}(5{,}000 - 3{,}750)^2 + \frac{1}{4}(0 - 3{,}750)^2 \\
&= 4{,}687{,}500 \text{ 。}
\end{aligned}$$

由於$\mathrm{Var}(X) > \mathrm{Var}(Y)$, 表示投資$A$之變異較大 。若喜愛冒險者, 說不定會採投資$A$, 有時會有較高獲利 。個性保守者, 可能會採投資B, 穩紮穩打 。

例28. 某項食品是以機械填裝, 每包標示16公斤 。若容量不足, 顧客可能會抱怨, 若容量過多, 公司也有損失 。令X表每包之容量 。假設X有$U[15.8, 16.2]$分佈, 即機率密度函數為$f(x) = 2.5,\ 15.8 \le x \le 16.2$ 。則

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{15.8}^{16.2} 2.5 x dx = 16.0 \text{ 。}$$

所以平均而言, 機器填裝的容量是正確的 。但

$$P(X < 16) = P(X > 16) = 0.5,$$

即各有0.5的機率, 會容量不足及容量超過。我們來求其變異數。

$$\mathrm{Var}(X)=\int_{-\infty}^{\infty}(x-16)^2f(x)dx=\int_{15.8}^{16.2}2.5(x-16)^2dx=0.0133\text{。}$$

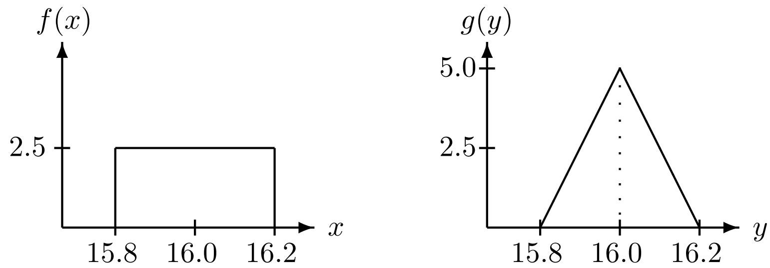
標準差為$\sqrt{\mathrm{Var}(X)}\doteq 0.115$。

為了降低變異, 公司引進新機器, 新的容量Y之機率密度函數為

$$g(y)=\begin{cases}25(y-16)+5,\ 15.8\le y<16,\\ 25(16-y)+5,\ 16\le y<16.2\text{。}\end{cases}$$

則仍有$E(Y)=16$, 但$\mathrm{Var}(Y)=1/150\doteq 0.00\overline{6}$, 標準差$\doteq 0.0816$。變異數下降一半以上, 標準差則下降約29%。

圖5給出X及Y機率密度函數之圖形。由圖形亦可看出Y之變異數較X之變異數小: 較多的機率配置在Y之期望值的近傍。



(a) X之機率密度函數　(b) Y之機率密度函數

圖5 例28中X及Y機率密度函數之圖形

我們先給變異數之一些性質。

<u>**定理4.8**</u> 設X為一隨機變數, 則下述二式成立, 其中a,b為常數。

$$\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X), \tag{21}$$

$$\mathrm{Var}(X)=E(X^2)-(E(X))^2\text{。} \tag{22}$$

由於變異數一定不是負的(非負隨機變數$(X-E(X))^2$之期望值必為非負, 見定理6之(ii)), 故由(22)式得

$$E(X^2)\ge(E(X))^2\text{。} \tag{23}$$

又由(21)式可看出(取$a = 0$)常數的變異數為0。這點由(22)式亦可看出。即若存在一常數b, 使得$P(X = b) = 1$, 則因$E(X^2) = b^2$, $E(X) = b$, 故$\text{Var}(X) = 0$。這當然是合理的。既然是常數(可視為一退化的隨機變數), 就沒有變異, 每次觀測皆得到該常數。反之, 若$\text{Var}(X) = 0$, 則存在常數b, 使得$P(X = b) = 1$。

<u>例29</u>**.** 投擲一公正的骰子, 令X表所得之點數。試求X之變異數及標準差。

解.前面已求出$E(X) = 3.5 = 7/2$。又$E(X^2) = \sum_{i=1}^{6} i^2/6 = 91/6$。故

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12},$$

且X之標準差為$\sqrt{35/12} \doteq 1.7078$。

<u>例30</u>**.** **常態分佈**(normal distribution)為一極重要的分佈。設X有參數μ及σ^2之常態分佈, 以$\mathcal{N}(\mu, \sigma^2)$表之,其中$\mu \in R$, $\sigma^2 > 0$。則X之機率密度函數為

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

當$\mu = 0, \sigma = 1$, 即得**標準常態分佈**(standard normal distribution), 以$\mathcal{N}(0, 1)$表之, 其機率密度函數為

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

常態分佈有很多好的性質, 其中之一是當X有$\mathcal{N}(\mu, \sigma^2)$分佈, 則$(X - \mu)/\sigma$有$\mathcal{N}(0, 1)$分佈。因此

$$P(X \le x) = P(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}) = P(Z \le \frac{x - \mu}{\sigma}),$$

其中, Z有$\mathcal{N}(0, 1)$分佈。即對任一常態分佈, 其機率值皆可經由$\mathcal{N}(0, 1)$分佈之機率值得到。這是其他分佈少有的性質。而由X至$(X - \mu)/\sigma$之變換, 便稱為**標準化**(standardized)。對於$\mathcal{N}(\mu, \sigma^2)$分佈, 可求出

$$E(X) = \mu, \text{Var}(X) = \sigma^2,$$

即二參數分別為其期望值及標準差。由標準常態分佈的機率值表, 可得當X有$\mathcal{N}(\mu,\sigma^2)$分佈,

$$\begin{aligned}
P(|X-\mu| \leq \ \sigma) &= P(|Z| \leq 1) = 0.6826,\\
P(|X-\mu| \leq 2\sigma) &= P(|Z| \leq 2) = 0.9545,\\
P(|X-\mu| \leq 3\sigma) &= P(|Z| \leq 3) = 0.9974 \text{。}
\end{aligned}$$

即X與期望值差異不超過1標準差, 2標準差, 3標準差之機率, 分別約為0.6826, 0.9545及0.9974(其他分佈當然不會也是這3個值)。對一常態分佈, 觀測值會偏離期望值3個標準差之機率, 僅約為0.003。

對一隨機變數, 於給出一次及二次動差(變異數是基於二次動差)後, 對其分佈便有了基本的了解。若知道更多次的動差, 對其分佈就愈了解。所謂分佈, 顧名思義, 就是一隨機變數取值之散佈情況。由於散佈著, 我們才又想以一些基本的量來描述散佈的情況。這就是動差的功能。

大家看, 我們由常數推廣至隨機變數, 由於隨機的難以掌握, 又想以一些簡單的量來描述此隨機的量。數學上就是如此常在由簡入繁, 又由繁入簡。

另外, 亦可以一些**轉換**, 來描述隨機變數的分佈。如**母函數**(generating function, 或稱**生成函數**), **拉普拉斯轉換**(Laplace transform), **特徵函數**(characteristic function)及**動差母函數**(moment generating function)等, 乃幾個常見可以唯一決定隨機變數之分佈的轉換。細節在此不再討論。

5 多維隨機變數

如同函數可以有多個變數, 我們也可以有**多維**(或說**多變數**,**多變量**)**隨機變數**。**聯合機率密度函數**, **聯合分佈函數**也就對應產生。例如, 二隨機變數X, Y之聯合分佈函數$F(x,y) = P(X \leq x, Y \leq y)$。統計分析要先收集資料, 而很多資料有不只一個變數, 多變數的討論, 為統計學中重要的

題材。本文只是對隨機變數給一初步的介紹, 因此不多著墨之。我們只給幾個概念。

定義2. 設二隨機變數X, Y, 以$f(x,y)$為聯合機率密度函數, X, Y之機率密度函數分別以$f_X(x)$及$f_Y(y)$表之。若

$$f(x,y) = f_X(x)f_Y(y),\ x,y \in R, \tag{24}$$

則稱X與Y**相互獨立**(mutually independent), 或**獨立**。

定理5.9 設二隨機變數X, Y, 以$F(x,y)$為聯合分佈函數, X, Y之分佈函數分別為$F_X(x)$及$F_Y(y)$。則X與Y獨立, 若且唯若

$$F(x,y) = F_X(x)F_Y(y),\ x,y \in R; \tag{25}$$

若且唯若存在函數$G(x)$及$H(y)$, 使得對$\forall x,\ y \in R$,

$$F(x,y) = G(x)H(y)。 \tag{26}$$

二隨機變數獨立, 表知道其中之一的值, 對另一變數的分佈, 沒有影響。在某些特殊的情況, 有可能一隨機變數為另一隨機變數之函數, 兩者可說關係極度密切, 但卻相互獨立。

同理我們也可以定義n個隨機變數之獨立。另外, 二隨機變數若不獨立, 如何來表示兩者之關係, 或說互相變化之情況, 也是統計裡常討論的題材。

n個隨機變數$X_1, \cdots, X_n$,若相互獨立, 且有共同的分佈(independent and identically distributed, 簡稱iid), 便稱此為一組**隨機樣本**(random sample)。在統計分析裡常會接觸隨機樣本, 很多統計的理論都基於隨機樣本。

定理2指出, 對一滿足分佈函數那些性質的函數, 能給一機率空間, 並在其上定義一隨機變數, 以該函數為其分佈函數。現若有二分佈函數F_1及F_2, 則可否給出一機率空間$(\Omega, \mathcal{F}, \mathcal{P})$, 且在$\Omega$上定義二隨機變

數X與Y, 使得X與Y獨立, 且各以F_1及F_2為分佈函數? 答案是肯定的, 不過其證明在此略去。

6 再論隨機性

本章介紹隨機變數, 不只在機率與統計裡, 在日常生活中, 人們也常提到"隨機"一詞, 其涵義究竟為何。

如果是說隨機變數, 則就是指非常數, 遵循某一分佈。另外, 可能會有下述幾個常見約定俗成的定義。

第一種是指**隨機樣本**。譬如說有一隨機變數X, 重複地觀測n次, 得到$X_1, \cdots, X_n$。則$X_1, \cdots, X_n$為iid, 為一組隨機樣本。這種例子很多。欲估計一銅板出現正面的機率p, 投擲n次, 假設各次間相互獨立, n次投擲得到$X_1, \cdots, X_n$, 其中$X_i = 1$, 表示第i次投擲出現正面, $X_i = 0$, 則表示出現反面, 且$P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$, 則$X_1, \cdots, X_n$便為一組隨機樣本。又如欲估計某廠牌燈泡壽命, 假設燈泡品質皆相同, 取n個測試, 分別得到使用時間$X_1, \cdots, X_n$, 則可合理地假設$X_1, \cdots, X_n$為iid, 因此為一組隨機樣本。

第二種情況是指**隨機取樣**, 則便有均勻分佈的意思。我們分連續型及離散型來說明。譬如說自區間$[a, b]$隨機地取一點, $a < b$。若取中的點以X表之, 則乃表X在$[a, b]$有均勻分佈。因此X落於$[a, b]$之任一子區間的機率, 等於此子區間之長度除以區間長度$b - a$。如果是自區間$[a, b]$隨機地取n個點, 取中的點以$X_1, \cdots, X_n$表之, 則便表$X_1, \cdots, X_n$為iid, 皆在$[a, b]$均勻分佈。其次看離散的情況。所謂自1至10, 隨機地取1數, 乃指1至10, 每一數被取中的機率皆相等, 都是1/10。又如樂透彩開獎, 想檢驗號碼是否隨機產生。假設是42取6的樂透彩, 共有$\binom{42}{6} = 5{,}245{,}786$種號碼的組合。就是要檢驗是否每一組合出現之機率皆為1/5,245,786(而非只是檢驗1至42的號碼, 是否出現的機率皆為1/42)。

不過有些情況下, 隨機取樣, 其中的"隨機"一詞意義可能並不很明確。見下例。

例31. 假設平面上有一圓, 半徑為r。在圓上**隨機**畫一條弦, 我們想求此弦長度大於圓之一內接等邊三角形邊長a的機率。令A表一隨機的弦其長度會大於a的事件。而一隨機的弦其長度可由下述三者之一決定: 此弦與圓心之距離D; 此弦的中點位置M; 此弦與圓心所張開的角θ。底下分別假設D, M, θ在其取值範圍內均勻分佈, 而求出A之機率。

(1)設D在$[0,r]$均勻分佈。此假設是基於若有一隻尺, 自與圓相切處以等速平行地向圓心移動, 隨時停止便與圓交出一條弦。由幾何性質知, 事件A發生, 若且唯若$D < r/2$, 故$P(A) = 1/2$。

(2)設M在整個圓內均勻分佈。此假設是基於由圓內任意取一點M, 當做弦的中點, 自M做垂直此點與圓心的連線便得到一弦。易見此弦長度是否大於a, 就看M是否落在半徑為$r/2$之同心圓內。故$P(A) = P(M$落在小圓內$)=$ 小圓面積/大圓面積$= 1/4$。

(3)設θ在$[0, 2\pi]$均勻分佈。此假設是基於自圓周上任取一點當做弦的一個端點, 然後自此點沿著圓周順時針等速移動, 隨時停止便得到弦的另一端點。易見事件A發生, 若且唯若此弦二端點與圓心所張開之角θ, 介於$2\pi/3$與$4\pi/3$之間, 故$P(A) = (4\pi/3 - 2\pi/3)/2\pi = 1/3$。

故在不同的假設下, 本問題的答案可以是1/2, 1/4, 或1/3。甚至在另外的假設下, 有可能得到其他答案。此問題即為**伯特朗詭論**(Bertrand paradox)。顯見在本例中, 所謂"隨機"畫一條弦, 其中隨機的意義, 顯然不若之前自區間$[a, b]$隨機取點那麼明確。若對隨機的解釋不同, 便可能得到不同的答案。

再看一有趣的例子。

例32. 假設有一無限大的袋子, 另有無限多個的球, 編號依序為1, 2, 3, $\cdots$。現進行下述實驗: 自某時刻起(時間0), 將1號至10號球放進袋中, 然後取出10號球。假設放球及取球所花時間都可忽略。1/2分鐘後, 將11號至20號球放進袋中, 然後取出20號球。又1/4分鐘後, 將21號至30號球放進袋中, 然後取出30號球。再來是又1/8分鐘後, $\cdots$, 餘此類推。我們有興趣

的問題是, 1分鐘後, 袋中有多少球。

此問題的答案很明顯, 就是袋中有無限多個球。因凡編號不是$10n$的球, $n \geq 1$, 皆在袋中。

現將情況改變。自時間0開始, 將1號至10號球放進袋中, 然後取出1號球。1/2分鐘後, 將11號至20號球放進袋中, 然後取出2號球。再1/4分鐘後, 將21號至30號球放進袋中, 然後取出3號球。再來是又1/8分鐘後, $\cdots$, 餘此類推。則1分鐘後, 袋中有多少球。

令人驚訝的是, 對此情況, 答案是袋中空無一球。怎麼可能? 每次不仍是放10球取1球, 何以第一情況1分鐘後, 袋中有無限多個球, 而在第二情況, 1分鐘後袋中卻一球皆無? 你實在無法相信。說明如下。在第二情況, 1號球是一開始(時間0)便被取走。至於n號球, $n \geq 2$, 在$(1/2)^1 + (1/2)^2 + \cdots + (1/2)^{n-1} (= 1 - (1/2)^{n-1})$分鐘後被取走。故對$\forall n \geq 1$, n號球在1分鐘後, 皆不在袋中。即得證1分鐘後, 袋中空無一球。

因此, 雖然同樣每次放10球, 取1球, 但那1個球如何取, 造成的結果差異很大。在第一情況, 只有編號為$10n$的球才被取出, $n \geq 1$。但在第二情況, 每一號球, 最終都被取出。

我們再考慮第三情況。放球的方式不變, 但取球的方式改為每次自袋裡已有的球中, "隨機"取1球。例如, 在時間0, 放進1號至10號球, 然後自此10個球中, 隨機的取1球(即1至10號球, 每號球被取中之機率皆為1/10)。對此新情況, 我們仍問, 1分鐘後, 袋中有多少個球。

解.我們將證明1分鐘後袋中空無一球的機率為1。先考慮1號球。令E_n表首n次取球後, 1號球仍在袋中的事件。則

$$P(E_n) = \frac{9}{10} \cdot \frac{18}{19} \cdot \frac{27}{28} \cdots \frac{9n}{9n+1}\text{。}$$

上式如何得到? n次取球後, 1號球仍要在袋中, 表第1次只能從10個球中, 1號之外, 其餘的9個球取; 第2次只能19個球中, 1號之外, 其餘的18個球取, 餘類推。例如, 最後一次只能自$9n+1 (= 10n - (n-1))$個球中, 1號之外的$9n$個球取。

1分鐘後, 1號球仍在袋中的事件, 即為$\bigcap_{n=1}^{\infty} E_n$。由於$E_1 \supset E_2 \supset \cdots$

$\supset E_n \supset E_{n+1} \supset \cdots$, 故利用機率函數的單調性質(見底下註1),

$$P(1分鐘後, 1號球不在袋中) = P(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\downarrow\infty} P(E_n) = \prod_{n=1}^{\infty} \frac{9n}{9n+1}\text{ 。}$$

我們如下證明$\prod_{n=1}^{\infty}(9n/(9n+1)) = 0$。首先因

$$\prod_{n=1}^{\infty} \frac{9n}{9n+1} = (\prod_{n=1}^{\infty}(1+\frac{1}{9n}))^{-1},$$

故若能證出$\prod_{n=1}^{\infty}(1+1/(9n)) = \infty$即得證。現對$\forall n \geq 1$,

$$\begin{aligned}\prod_{n=1}^{\infty}(1+\frac{1}{9n}) &\geq \prod_{n=1}^{m}(1+\frac{1}{9n}) = (1+\frac{1}{9})(1+\frac{1}{18})\cdots(1+\frac{1}{9m}) \\ &> \frac{1}{9}+\frac{1}{18}+\cdots\frac{1}{9m} = \frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}\text{ 。}\end{aligned}$$

因當$m \to \infty$時, $\sum_{i=1}^{m} 1/i \to \infty$, 故得$\prod_{n=1}^{\infty}(1+1/(9n)) = \infty$。故若令$F_i$表$i$號球1分鐘後在袋中的事件, 則$F_1 = \bigcap_{n=1}^{\infty} E_n$, 且$P(F_1) = 0$。同理可證$P(F_i) = 0$, $\forall i \geq 1$。而1分鐘後, 袋中不是空的之事件乃$\bigcup_{i=1}^{\infty} F_i$。利用**波爾不等式**(Boole's inequality), 得

$$P(\bigcup_{i=1}^{\infty} F_i) \leq \sum_{i=1}^{\infty} P(F_i) = 0\text{ 。}$$

故1分鐘後, 袋中為空之機率的確是1。

<u>註1</u>. 事件$\{E_n, n \geq 1\}$若滿足

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots,$$

便稱**漸減數列**, 若滿足

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots,$$

便稱**漸增數列**。對漸減或漸增數列的事件$\{E_n, n \geq 1\}$, 皆有

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)\text{ 。}$$

此即機率函數之**單調性質** 。

參考文獻

1. 黄文璋(2010). 機率論, 第二版 。華泰文化事業股份有限公司, 台北市 。