連續隨機變數

- 何謂連續隨機變數？
  - 結果是長度、時間、重量、體積等
  - 實驗所定義的隨機變數的**值域**為**實數區間**
  - 除非另有說明，我們假設連續型隨機變數之值域為所有**實數**所成的集合
  - **連續隨機變數**無法用**離散型隨機變數**的機率函數分配機率

連續隨機變數 (Cont.)

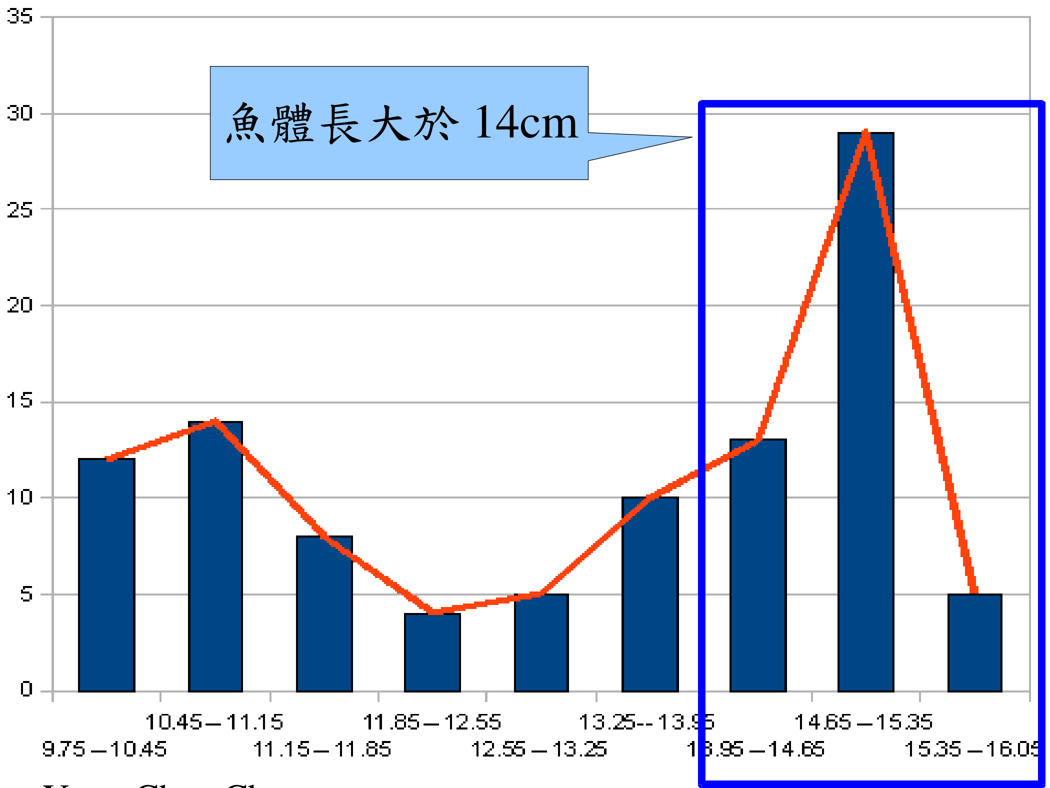
- 例：某水產養殖場水池裡養了一群虱目魚，隨機撈取一條量測體長
  - 問，當隨機撈到魚的體長大於 14cm 的**機率**為何？
  - 魚的體長 X 是連續隨機變數
  - 量測 100 條魚後得知，最短 10.1cm ，最長 15.8cm
  - 統計量測到的魚體長資料
  - 魚的體長分為 9 個區間，則 $\frac{15.8-10.1}{9}=0.63$

連續隨機變數 (Cont.)

- 假設以 0.7 為區間長度，並得到下列統計資料
- 魚體長大於 14cm 機率約為

$$\frac{13+29+5}{100}=0.47$$

- 若增加量測的魚數量則區間會增加，且直條圖會平滑成 $y=f(x)$ 曲線



連續隨機變數 (Cont.)

- 魚體長落在某一區間的機率與區間所包含的面積成正比
- $y = f(x)$ 函數稱為隨機變數 X 之**機率密度函數**

定義

若 f 為實數函數，且滿足

$(1)\ f(x) \geq 0$

$(2)\ \int_{-\infty}^{\infty} f(x)\,dx = 1$

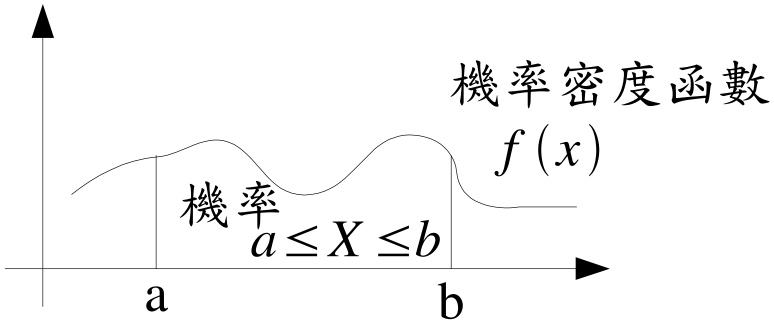
連續隨機變數 (Cont.)

- 若連續隨機變數 X 之機率密度函數為 $y = f(x)$，則對 $a \le b$

$$P(a < X < b) = \int_a^b f(x)\,dx$$

- 若於上式取 $b = a$ 則 $P(X = a) = \int_a^a f(x)\,dx = 0$
- 因此，在連續隨機變數出現任一點的機率均為 0，且 $P(a < X < a) = P(a < X \le a) = P(a \le X < a) = P(a \le X \le a)$

連續隨機變數 (Cont.)



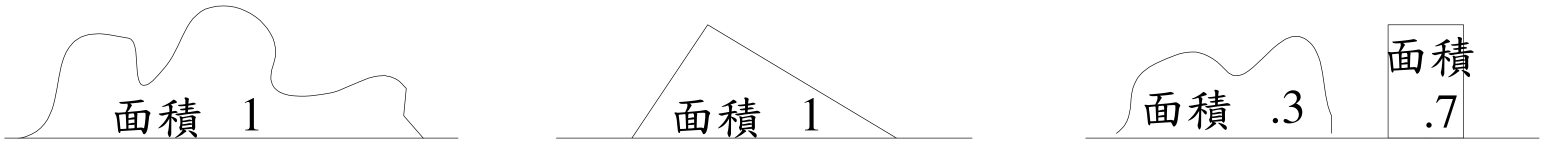
- 並非所有先前提到機率函數都可做為隨機變數 X 的密度
- 必須注意的是，機率永遠不會是**負值**，且

$$P(-\infty < X < \infty) = 1$$

- 因此，$f(x) \ge 0$ for all x

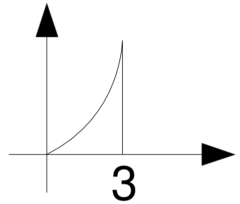
連續隨機變數 (Cont.)

- 且 $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- 上式之所以成立必須滿足$f(\infty)=0, f(-\infty)=0$
- 一些典型的密度



連續隨機變數 (Cont.)

- 設隨機變數 X 有密度如 $f(x)=\begin{cases}\dfrac{x^2}{9} & \text{where } 0\le x\le 3\\ 0 & \textit{otherwise}\end{cases}$



- 因 $f(x)\ge 0$，所以這是個合理的密度，且

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_0^3 \frac{1}{9}x^2\,dx=\frac{x^3}{27}\Big|_0^3=\frac{27}{27}-\frac{0}{27}=1$$

- 因此 $P(1\le X\le 2)=\int_1^2 \frac{1}{9}x^2\,dx=\frac{x^3}{27}\Big|_1^2=\frac{2^3}{27}-\frac{1^3}{27}=\frac{7}{27}$

$$P(X\le 1)=\int_{-\infty}^{1} f(x)\,dx=\int_0^1 \frac{1}{9}x^2\,dx=\frac{x^3}{27}\Big|_0^1=\frac{1^3}{27}-\frac{0^3}{27}=\frac{1}{27}$$

$$P(X\ge 3)=\int_3^{\infty} 0\,dx=0$$

連續隨機變數 (Cont.)

- 方程式 $f(x)=\begin{cases} \dfrac{x^2}{9} & \text{where } 0 \le x \le 3 \\ 0 & otherwise \end{cases}$
- 可改寫成 => $f(x)=\dfrac{x^2}{9}$ where $0 \le x \le 3$
- 但不可只寫 $f(x)=\dfrac{x^2}{9}$

不是合理的密度

- 若 X 為一連續機率變數則 $P(X=x)=0$
- 因為 $\int_x^x f(x)\,dx=0$

即使 $X=x$ 是可能的，它的機率還是 0

連續隨機變數 (Cont.)

- 導出 $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$
- 以 $f(x) = \begin{cases} \dfrac{x^2}{9} & \text{where } 0 \le x \le 3 \\ 0 & otherwise \end{cases}$ 為例
- $P(1 \le X \le 2) = P(1 \le X < 2) = P(1 < X \le 2) = P(1 < X < 2) = \dfrac{7}{27}$

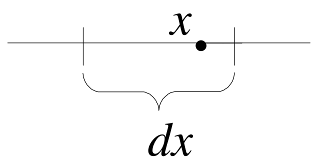
$$\int_1^2 \frac{x^3}{27}\,dx = \frac{x^3}{27}\Big|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

$$P(1 \le X < 2) = P(1 \le X \le 2) - P(X = 2)$$

$$\int_1^2 \frac{x^2}{9}\,dx - \int_2^2 \frac{x^2}{9}\,dx = \frac{x^3}{27}\Big|_1^2 - 0 = \frac{8}{27} - \frac{1}{27} - 0 = \frac{7}{27}$$

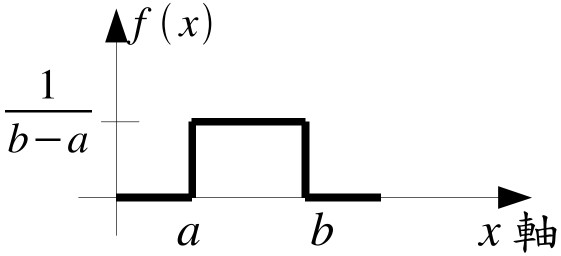
連續隨機變數 (Cont.)

- 事件 $X \approx x$
  - 設 x 的密度函數為 $f(x)$
  - 當 X 很接近 x 情況，以 $X \approx x$ 表示之
  - 利用 密度×長度 的公式可得到 $P(X \approx x) = f(x)\,dx$



- 均勻分佈隨機變數的密度
  - 設 X 均勻分佈在 $[a, b]$ 區間，即 $P(X = \text{事件}) = \dfrac{\text{符合的長度}}{\text{全部的長度}}$
  - 將一個單位的機率事件攤在 b-a 呎的長度，單位機率為 $\dfrac{1}{b-a}$，且其機率密度為 $f(x) = \dfrac{1}{b-a}$ where $a \le x \le b$

連續隨機變數 (Cont.)



- 在區間 $[a, b]$ 上，因 X 是均勻分佈，故 $f(x)$ 有固定的高度
- 再者，由於全部的面積必定是 1 ，且底是 b-a ，因此，其高度必定是 $\frac{1}{b-a}$
- 例：若 X 是均勻分佈在 [-3, 7] ，則

$$f(x)=\frac{1}{10} \text{ where} -3 \leq x \leq 7$$

分佈函數

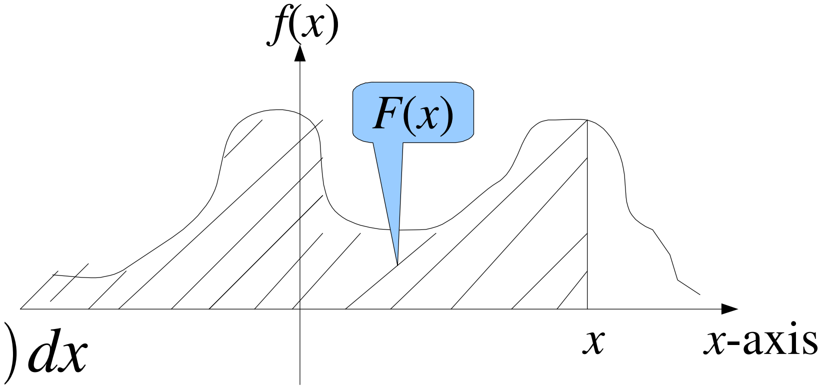
- 密度函數：用 $f(x)$ 表示
- 分佈函數：用 $F(x)$ 表示
- 何謂分佈函數？
  - 令 $f(x)$ 是隨機變數 X 的密度函數，則這些機率是由一個函數所形成，此函數稱之為分佈函數，並用 $F(X)$ 表示之。

分佈函數(Cont.)

- 隨機變數的(累積)分佈函數
  - 令 x 的分佈函數為 $F(x)=P(X\le x)$
  - 意思為，截至 x 為止的機率總和
  - 因此，常有人稱之為**累積分佈函數**，並以縮寫 cdf 表示
  - ★注意★　每一個隨機變數，不管是離散、連續或其他，均存在**分佈函數**

分佈函數(Cont.)

- 在連續隨機變數的情況下
  - 除了有 ***分佈函數***外，還有 ***密度函數*** $f(x)$
  - 以圖形說明
  - 以式子表示

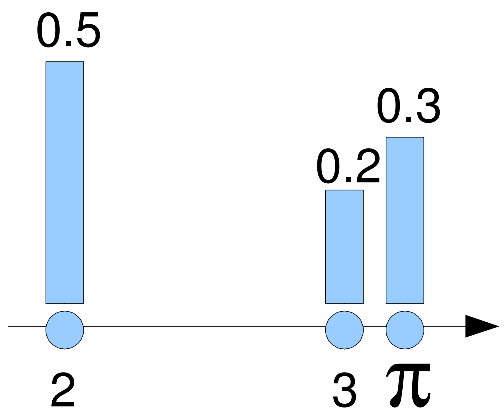


$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

  - 相對於圖中 $F(x)$ 所指的連續區域

分佈函數(Cont.)

- 在離散隨機變數的情況下
  - 設隨機變數 X 只有在 2, 3, 及π時有機率存在，即 $P(X=2)=0.5$, $P(X=3)=0.2$ 及 $P(X=\pi)=0.3$
  - 則 x 的機率函數為 $P(x)=\begin{cases} 0.5 & \text{if } x=2 \\ 0.2 & \text{if } x=3 \\ 0.3 & \text{if } x=\pi \end{cases}$

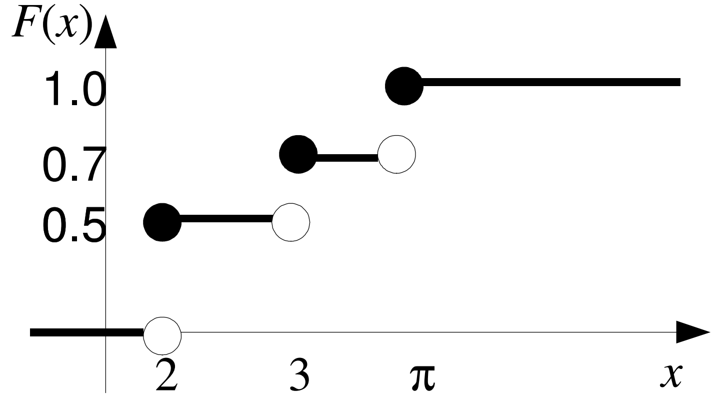


分佈函數(*Cont.*)

- 設 $F(x)$ 是 X 的分佈函數，表示為累積機率

$$F(x)=\begin{cases} 0 & if \quad x<2 \\ 0.5 & if \quad 2\le x<3 \\ 0.7 & if \quad 3\le x<\pi \\ 1 & if \quad x\ge\pi \end{cases}$$

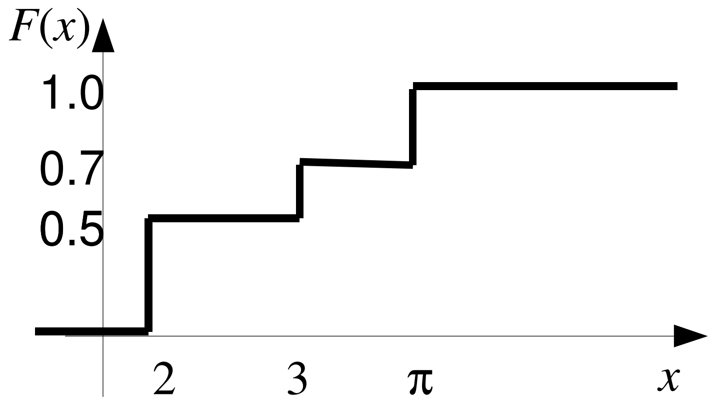
圖示



- 離散隨機變數的分佈是**步階函數**，高度從 0 到 1
- 在所有 X 的可能值都有跳躍，當 x_0 時，跳躍的大小是 $P(X=x_0)$

分佈函數(Cont.)

- 電腦畫的圖形
  - 雖不正確卻很好用
  - 一種方便的表示法

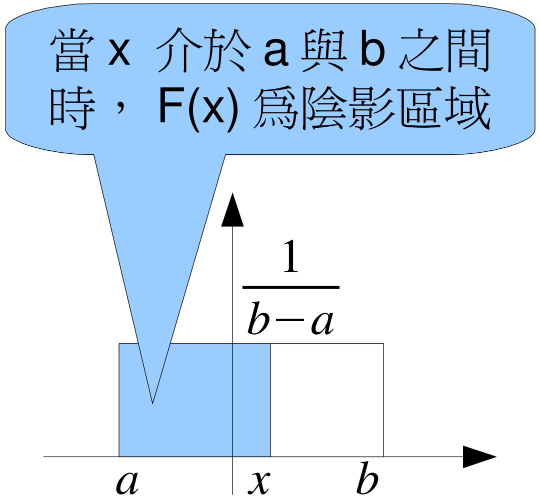


- 因應新圖表示法，分佈函數改為

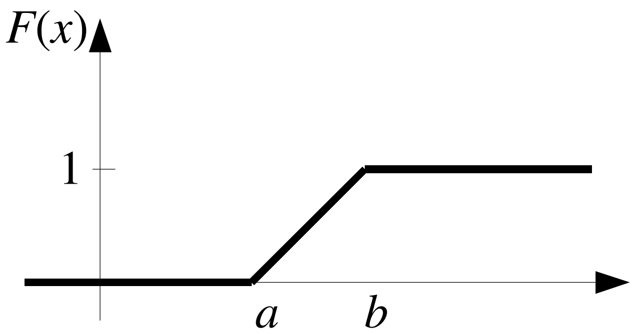
$$F(x)=\begin{cases} 0 & \text{if} \quad x \le 2 \\ 0.5 & \text{if} \quad 2 \le x \le 3 \\ 0.7 & \text{if} \quad 3 \le x \le \pi \\ 1 & \text{if} \quad x \ge \pi \end{cases}$$

分佈函數(Cont.)

- 均勻分佈隨機變數的分佈函數
  - X 均勻分佈在 $[a, b]$ 區間
  - 密度函數 $f(x)$
  - 累積面積 (機率) 是 1
  - 所以 $F(x) = 1$



$$F(x)=\begin{cases} 0 & \text{if} \quad x\le a \\ \dfrac{x-a}{b-a} & \text{if} \quad a\le x\le b \\ 1 & \text{if} \quad x\ge b \end{cases}$$

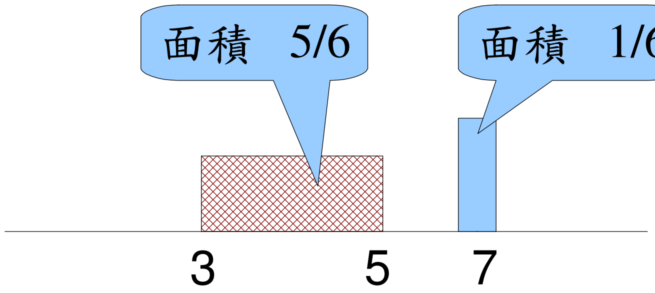


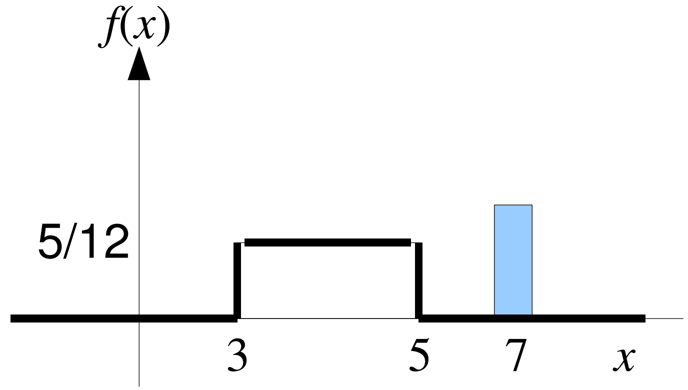
分佈函數(Cont.)

- 混合隨機變數實例—丟骰子
  - 2 點→玩家得 7 元
  - 非 2 點→玩家得 [3, 5] 間隨機選取的金額
  - 設 X 是玩家贏的金額
  - 以連續隨機變數而言 $P(X = x)$ 應為 0
  - 本例， $P(X = 7) = 1/6$ ，所以 X 非連續 (X 沒有密度函數)

分佈函數 (Cont.)

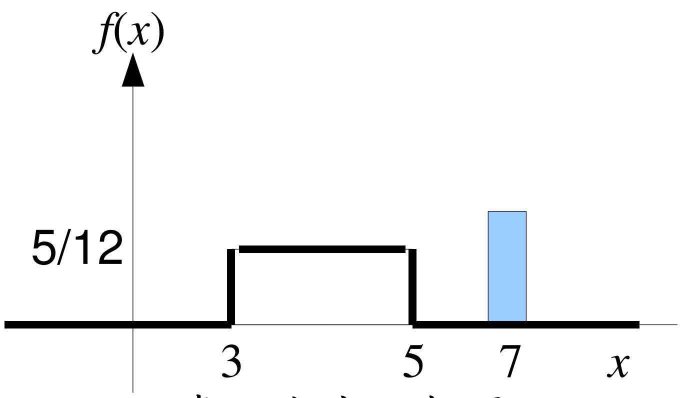
- X 亦不是離散的，因 X 可在 [3, 5] 區間中挑任一個數
- 因此，X 沒有機率函數，但可以題意虛擬一個密度
- 先以圖形表示



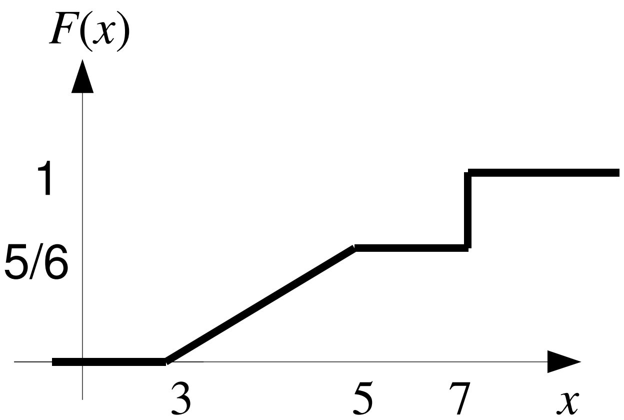


$$g(x)+\frac{1}{6}\delta(x-7)\ \text{where}\ g(x)=\frac{5}{12}\ \text{for}\ 3\le x\le 5$$

分佈函數 (Cont.)



虛擬密度函數圖示



分佈函數圖示

- 分佈函數性質
  - F 為非遞減 (F 能增或停留在同一高度，但不會下降)
  - $F(-\infty)=0, F(\infty)=1$

分佈函數(Cont.)

- 典型的分佈函數圖示

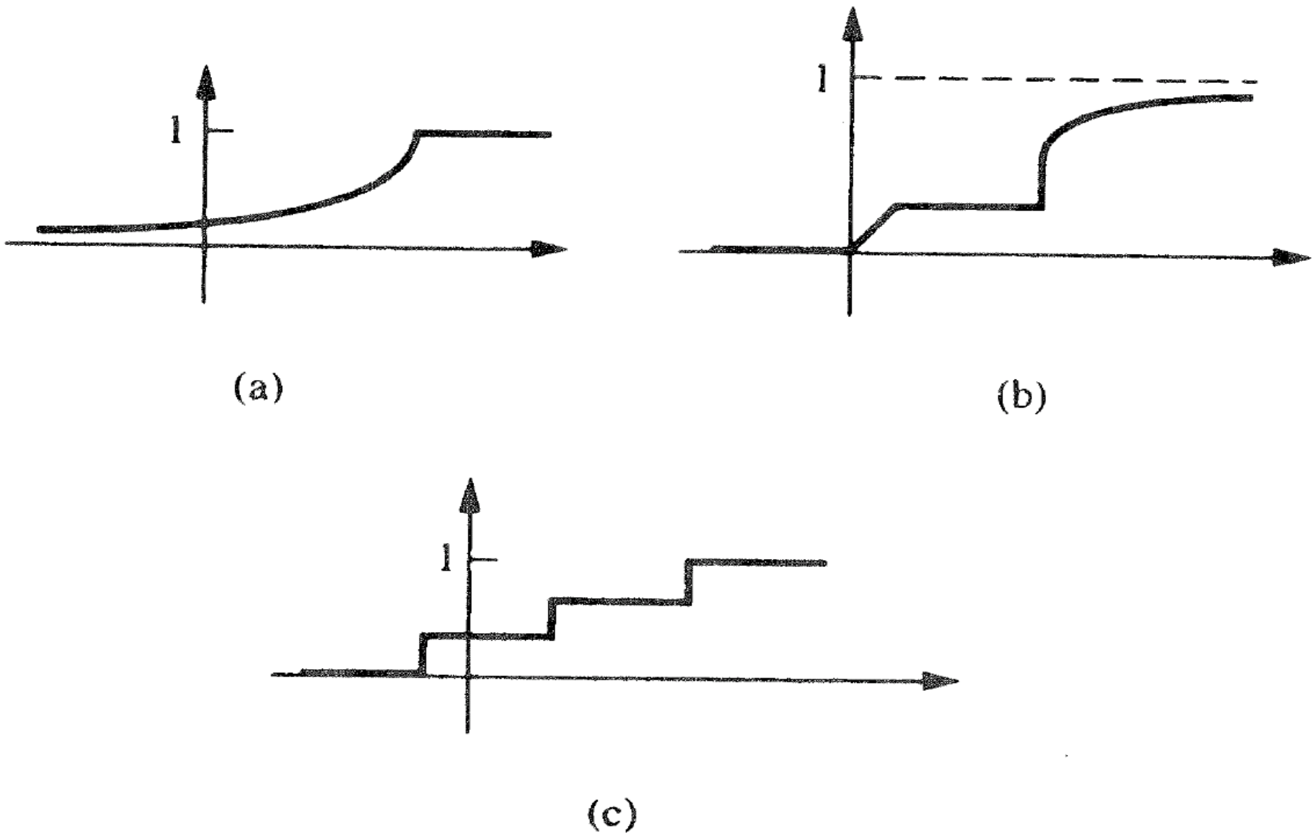


圖 9　典型的分佈函數

分佈函數 (*Cont.*)

- 由分佈函數計算機率
  - 分佈函數 $F(x)$ 若在 $x = x_0$ 處有**跳躍** (jump) ，那一定是**向上跳躍** (jump up)
  - 而在 $x = x_0$ 這一點會向上跳躍顯然是在 x_0 的情況大量發生進而使機率提高

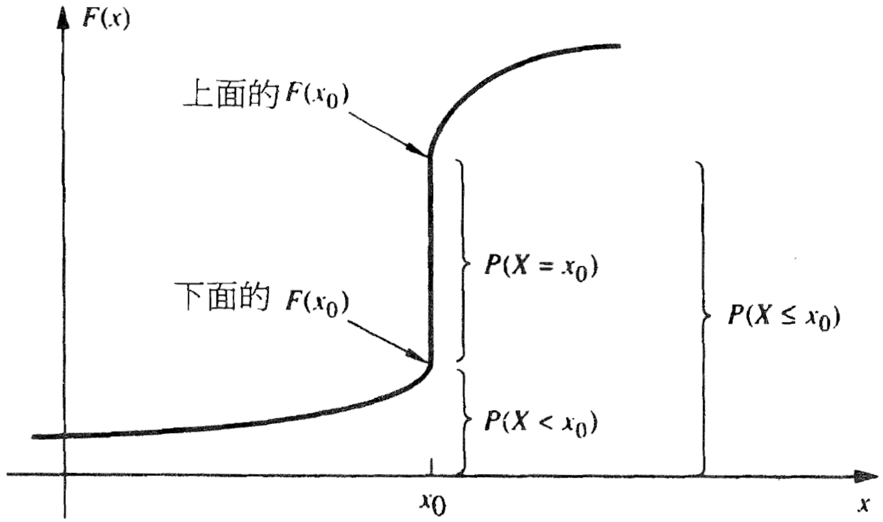


圖 10 分佈函數的跳躍

分佈函數 (Cont.)

- 因此可整理如下：
  - $P(X = x_0) =$ 跳躍的大小
  - $P(X \leqq x_0) =$ 當 x_0 狀況發生後機率會變成跳躍的高點 $=$ 上面的 $F(x_0)$
  - $\mathrm{P}(X < x_0) =$ 當 x_0 狀況發生前機率是跳躍的低點 $=$ 下方的 $F(x_0)$

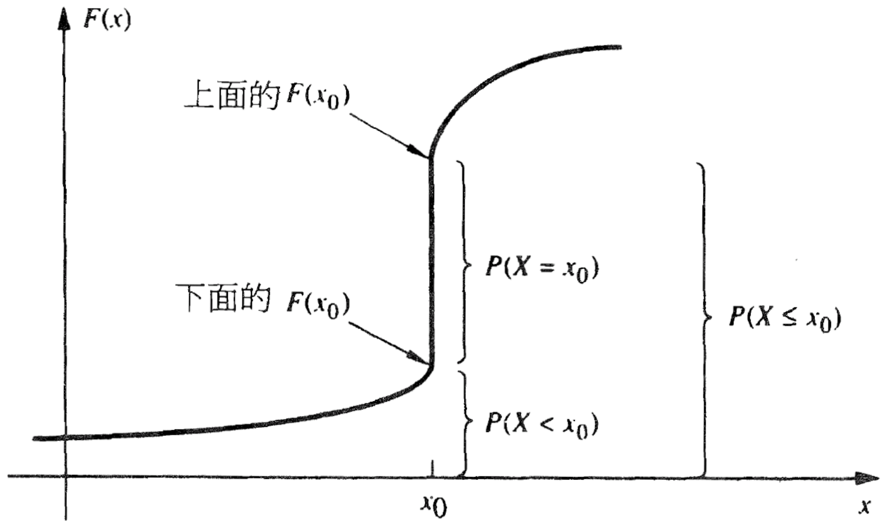


圖 10 分佈函數的跳躍

分佈函數(Cont.)

- 此外
  - $P(X > x_0) = 1 - P(X \leqq x_0)$
    $= 1 - 上面的\ F(x_0)$
  - $P(X \geqq x_0) = 1 - P(X < x_0)$
    $= 1 - 下面的\ F(x_0)$

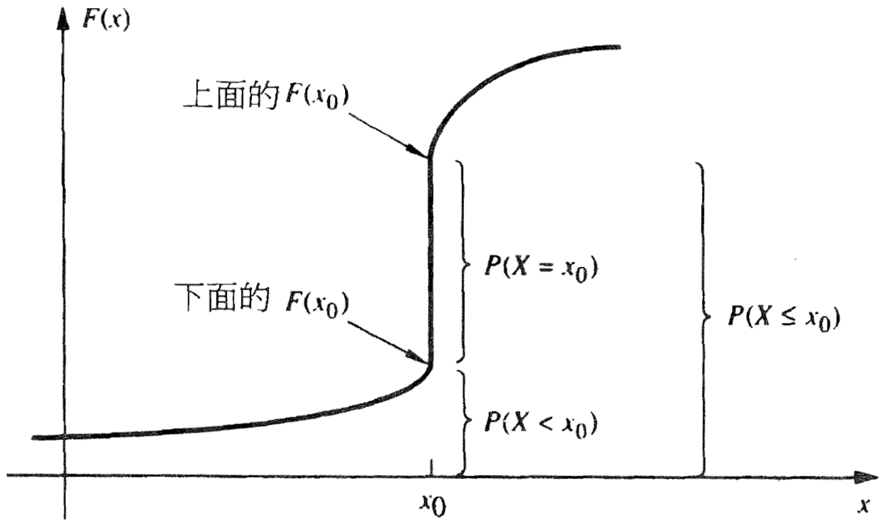


圖 10 分佈函數的跳躍

分佈函數(Cont.)

- 若在 a 點及 b 點有跳躍的情況
  - $P(a \leqq X \leqq b) = P(X \leqq b) - P(X < a)$

    $=$ 上面的 $F(b)$ – 下面的 $F(a)$

    (包括在 b 點及 a 點的機率)
  - $P(a < X \leqq b) = P(X \leqq b) - P(X \leqq a)$

    $=$ 上面的 $F(b)$ – 上面的 $F(a)$

    (包括在 b 點的機率)

分佈函數(*Cont.*)

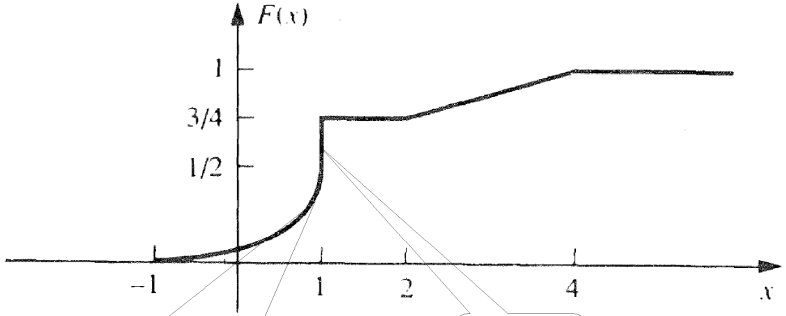
- $P(a \leqq X < b) =$ 下面的 $F(b) -$ 下面的 $F(a)$
- $P(a < X < b) =$ 下面的 $F(b) -$ 上面的 $F(a)$

- 若在 X 的分佈函數中沒有跳躍，則
  - $P(X=x_0) = 0$
  - $P(X<x_0) = P(X\leqq x_0) = F(x_0)$
  - $P(X>x_0) = P(X\geqq x_0) = 1 - F(x_0)$
  - $P(a \leqq X\leqq b) = P(a < X\leqq b) = P(a \leqq X<b) = P(a < X<b) = F(b) - F(a)$

分佈函數(Cont.)

- 設 x 有分佈函數如

$$F(x)=\begin{cases} 0 & \text{if} \quad x\le -1 \\ \frac{1}{8}(x+1)^2 & \text{if} \quad -1\le x\le 1 \\ \frac{3}{4} & \text{if} \quad 1\le x\le 2 \\ \frac{1}{8}x+\frac{1}{2} & \text{if} \quad 2\le x\le 4 \\ 1 & \text{if} \quad x\ge 4 \end{cases}$$



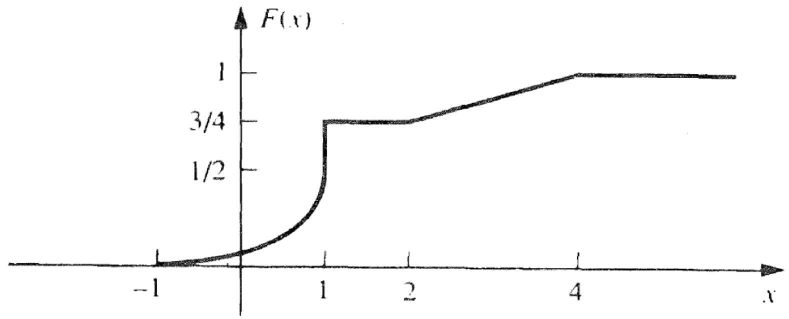
$$\frac{1}{8}(x+1)^2\Big|_{x=1}=\frac{1}{2}$$

¼
跳躍

- $P(X=1)$ = 跳躍大小 = ¼
- $P(X\leqq 1)$ = 上面的 $F(1)$= ¾
- $P(X<1)$ = 下面的 $F(1)$= ½
- $P(X>1)$ = 1- 上面的 $F(1)$= 1 - ¾ = ¼

分佈函數(*Cont.*)

- $P(X \geqq 1)$ = 1- 下面的 $F(1)$= 1 - ½ = 1/



- $P(X=0) = 0$
- $P(X \leqq 0) = P(X<0) = F(0) = 1/8$
- $P(X \geqq 0) = P(X>0) = 1 - F(0) = 7/8$
- $P(2 \leqq X \leqq 3) = P(2 < X < 3) = F(3) - F(2) = 7/8 - 3/4 = 1/8$
- $P(1 \leqq X \leqq 3)$ = $F(3)$ - 下面的 $F(1)$ = 7/8 - ½ = 3/8
- $P(1 < X \leqq 3)$ = $F(3)$ - 上面的 $F(1)$ = 7/8 - ¾ = 1/8
- $P(1 < X \leqq 2) = 0$, $P(X>4) = 0$

連續隨機變數的分佈函數

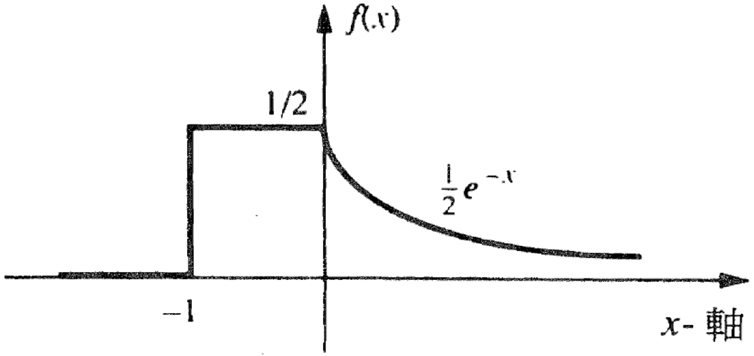
- 隨機變數都有分佈函數
- 不是每個隨機變數都有**密度函數**
- 有密度函數的隨機變數稱之為**連續**隨機變數
- 連續隨機變數的特性
  - $P(X=x)=0$
  - 連續隨機變數的分佈函數 $F(x)$ 不會跳躍

從密度函數到分佈函數

- 從密度函數求得分佈函數
  - $F(x) =$ 在 $f(x)$ 圖形下方所累積的面積
  - $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$
- 例：x 的密度

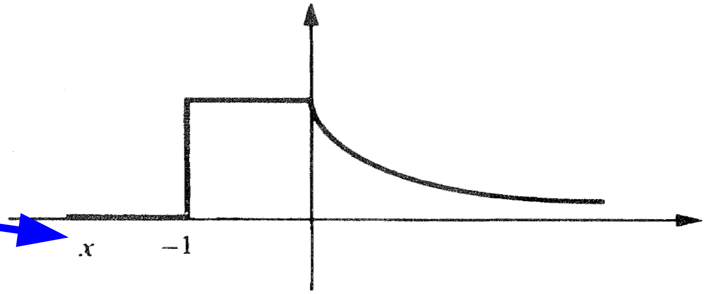
$$f(x) = \begin{cases} \dfrac{1}{2} & where \quad -1 \le x \le 0 \\ \dfrac{1}{2}e^{-x} & where \quad x \ge 0 \end{cases}$$

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e \approx 2.7182818\ldots$$



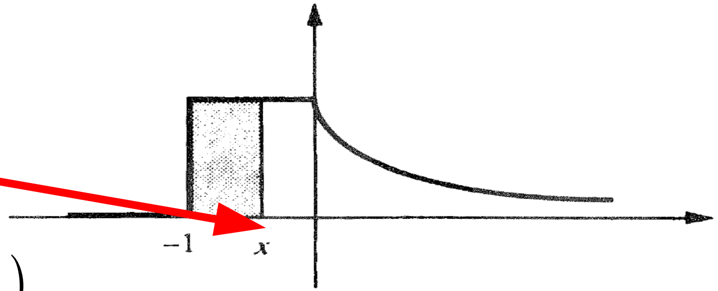
從密度函數到分佈函數 (Cont.)

- 情況一　$x < -1$



$$F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_{-\infty}^{x} 0\,dx=0$$

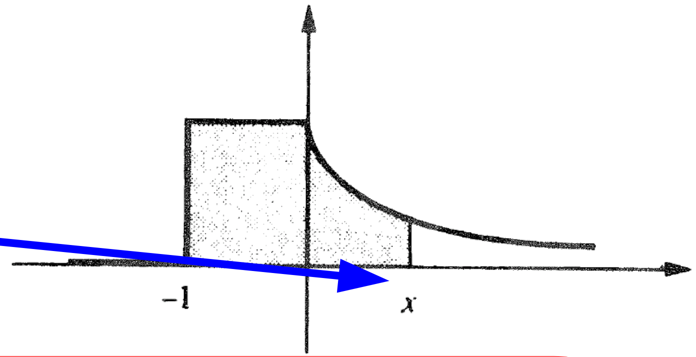
- 情況二　$-1 < x < 0$



$$F(x)=\int_{-1}^{x} f(x)\,dx=\frac{1}{2}x\Big|_{-1}^{x}=\frac{1}{2}x-(-\frac{1}{2})=\frac{1}{2}(x+1)$$

從密度函數到分佈函數 (Cont.)

- 情況三 $x>0$



$$F(x)=\boxed{\int_{-1}^{0}\frac{1}{2}dx}+\boxed{\int_{0}^{x}\frac{1}{2}e^{-x}dx}=\frac{1}{2}+\frac{1}{2}(1-e^{-x})$$

$$\int_{-1}^{0}\frac{1}{2}dx=\frac{1}{2}x\Big|_{-1}^{0}$$
$$=0-(\frac{1}{2}\cdot-1)$$
$$=0-(-\frac{1}{2})$$
$$=\frac{1}{2}$$

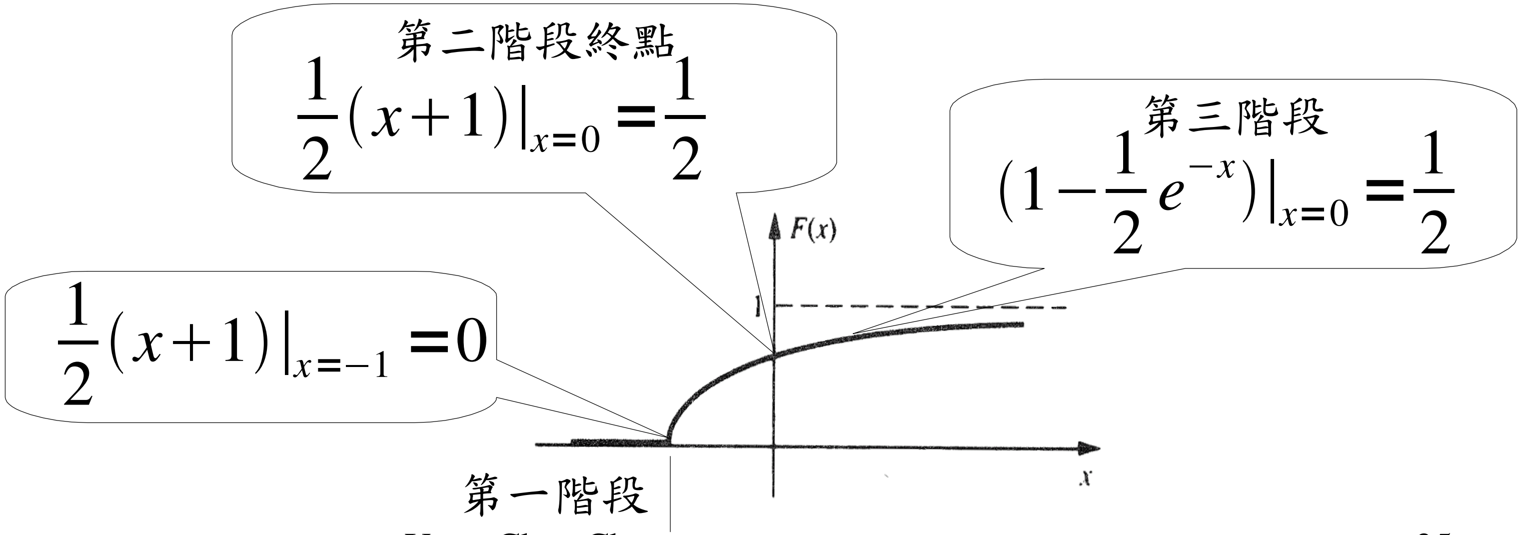
$$\int_{0}^{x}\frac{1}{2}e^{-x}dx=-\frac{1}{2}e^{-x}\Big|_{0}^{x}$$
$$=-\frac{1}{2}e^{-x}-(-\frac{1}{2}e^{-0})$$
$$=-\frac{1}{2}e^{-x}-(-\frac{1}{2})$$
$$=\frac{1}{2}-\frac{1}{2}e^{-x}$$
$$=\frac{1}{2}(1-e^{-x})$$

從密度函數到分佈函數 (Cont.)

- 整理

$$F(x)=\begin{cases} 0 & \text{if} \quad x\le -1 \\ \frac{1}{2}(x+1) & \text{if} \quad -1\le x\le 0 \\ 1-\frac{1}{2}e^{-x} & \text{if} \quad x\ge 0 \end{cases}$$

- 測試 $F(-\infty)=0$ 及 $F(\infty)=1-\frac{1}{2}e^{-\infty}=1-0=1$



從密度函數到分佈函數 (Cont.)

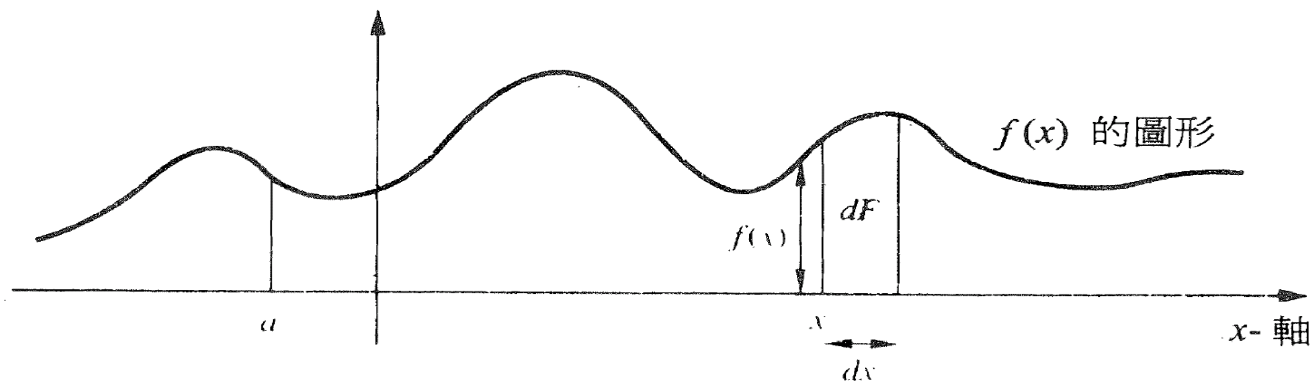
- ★注意★
  - $F(x)$ 不是 $\int f(x)\,dx$
  - $F(x)$ 不是 $\int_5^7 f(x)\,dx$
  - $F(x)$ 是累積分佈函數，因此 $F(\mathrm{i})$ 總是 $\int_{-\infty}^{x} f(x)\,dx$

從沒有跳躍的分佈函數到密度

- 微積分第一基本定理
  - If $F(x)=\int_a^x f(x)dx$ 則 $F'(x)=f(x)$
  - 即只要對分佈函數 $F(x)$ 進行微分可得連續機率密度
- 第一基本定理形成原因
  - 令 $F(x)=\int_a^x f(x)dx$
  - 則 $F(x)$ 是 $f(x)$ 圖形下方所累積的面積 ($x=a$ 開始)

從沒有跳躍的分佈函數到密度 (Cont.)

- 若 x 每次的改變量是 dx，且每當 x 改變時面積的改變量為 dF， $dF = f(x)\,dx$
- 因此 $f(x) = \dfrac{dF\text{的改變}}{dx\text{的改變}}$
- $dx \to 0$ 的極限可得 $f(x) = F'(x)$



微分

從沒有跳躍的分佈函數到密度 (Cont.)

- 例：設 X 的分佈函數是

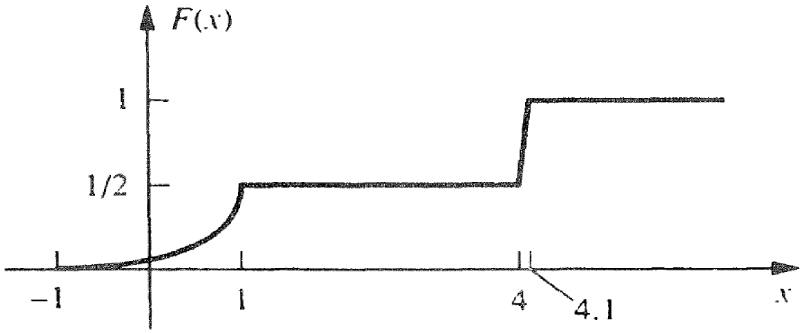
分佈函數

$$F(x)=\begin{cases} 0 & \text{if} \quad x\le 1 \\ \frac{1}{8}(x+1)^2 & \text{if} \quad -1\le x\le 1 \\ \frac{1}{2} & \text{if} \quad 1\le x\le 4 \\ 5\mathrm{x}-\frac{39}{2} & \text{if} \quad 4\le x\le 4.1 \\ 1 & \text{if} \quad x\ge 4.1 \end{cases}$$

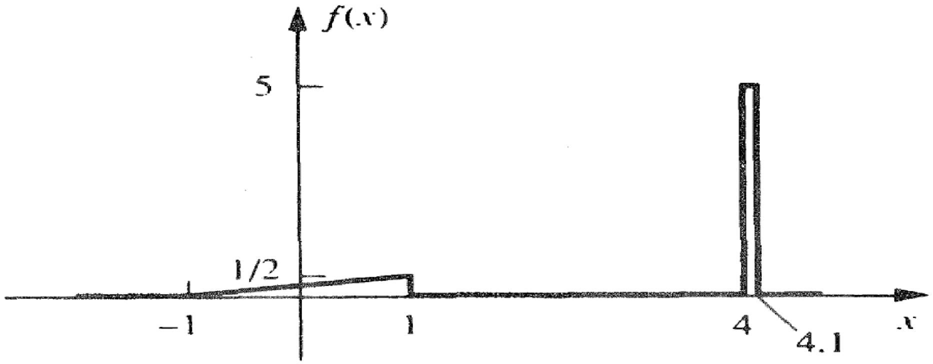
$\xrightarrow{\text{微分}}$

密度函數

$$f(x)=F'(x)=\begin{cases} 0 & \text{if} \quad x\le 1 \\ \frac{1}{4}(x+1) & \text{if} \quad -1\le x\le 1 \\ 0 & \text{if} \quad 1\le x\le 4 \\ 5 & \text{if} \quad 4\le x\le 4.1 \\ 0 & \text{if} \quad x\ge 4.1 \end{cases}$$



Yung-Chen



離散、連續與混合隨機變數之比較整理

- 離散
  - 離散隨機變數 X 可想像成是有限的數或是無限中可被計算的部份
  - 離散隨機變數的分佈函數是**步階函數**
  - 設 X 的機率函數 $p(x)$ 是由 $p(x) = P(X=x)$ 求得
  - X 不存在密度

離散、連續與混合隨機變數之比較整理 (Cont.)

- 連續
  - 連續隨機變數 X 是從無限個可能值中取出
  - 連續隨機變數的機率是從整理 (積分運算) 其密度函數 $f(x)$ 中求得
  - 即使發生 x 的情況是可能的，但其機率還是為 $P(X=x) = 0$
  - 分佈函數 $F(x)$ 不存在跳躍情況

離散、連續與混合隨機變數之比較整理 (Cont.)

- 混合 (i.e. 含有離散及連續)
  - 混合的隨機變數 X，整體而言其不被歸入離散隨機變數，也不被歸入連續隨機變數
  - X 是從無限個無法計數中取得
  - 在混合情況下，$P(X=x)\neq 0$ 這點與純連續型不同
  - x 點的分佈函數圖形上是有跳躍的且其跳躍幅度為 $P(X=x)$，此點與純連續隨機變數不同

離散、連續與混合隨機變數之比較整理 (Cont.)

- 混合的隨機變數具有一虛擬的密度
  - 由密度及機率函數組成
  - 有些點機率會突然大幅增加，其餘的部份則是由其他的機率組成

密度與分佈函數比較整理

- 每一個隨機變數 X 都存在一個分佈函數 $F(x)$
- $F(x)=P(X\leq x)$
- $F(-\infty)=0, F(\infty)=1$
- F 始終是遞增，沒有遞減的情況
- 只有在 $P(X=x_0)\neq 0$ 時 f 會在 x_0 這一點跳躍，且其跳躍幅度為 $P(X=x)$
- $P(a\leq X\leq b)=$上面的$F(b)-$下面的$F(a)$
  - 若在 a, b 沒有跳躍，則可用 $F(b)-F(a)$ 計算

密度與分佈函數比較整理 (Cont.)

- 並不是所有的隨機變數都有(合理的)密度
- 若 X 除了分佈函數 $F(x)$ 之外也存在密度 $f(x)$，則存在下列式子
- $P(a \le X \le b) = \int_a^b f(x)\,dx$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $f(x) \ge 0$
- $p(X \approx x) = f(x)\,dx$
- $F(x) = \int_{-\infty}^{x} f(x)\,dx = f(x)$下方的累積面積
- $f(x) = F'(x)$

隨機變數的分佈

- 在我們的認知裡，任一事件屬於 X，該事件的機率可在 $F(x)$ 中找到，此時我們會認為『針對隨機變數 X 是由其分佈函數 $F(x)$ 所決定』
- 想一下：
  - 若 X 是離散的：X 是由其機率函數決定
  - 若 X 是連續的：X 是由其密度函數決定
- 因此，當我們涉及『分佈』這個詞時，指的是我們有管道可求得某一事件的機率(當該事件屬於 X 時)
- 這裡所指的『管道』有可能是

  1. 分佈函數

  2. 機率函數(離散)或密度函數(連續)