# 連續隨機變數

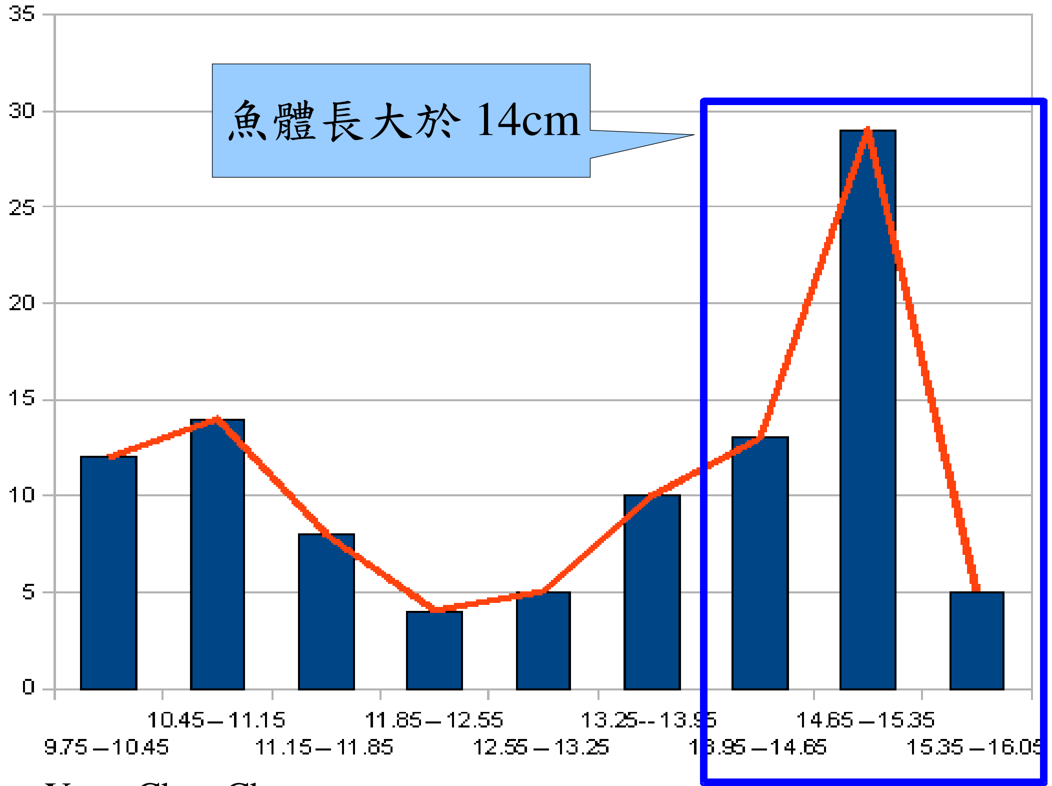
- 何謂連續隨機變數？
  - 結果是長度、時間、重量、體積等
  - 實驗所定義的隨機變數的**值域**為**實數區間**
  - 除非另有說明，我們假設連續型隨機變數之值域為所有**實數**所成的集合
  - **連續隨機變數**無法用**離散型隨機變數**的機率函數分配機率

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

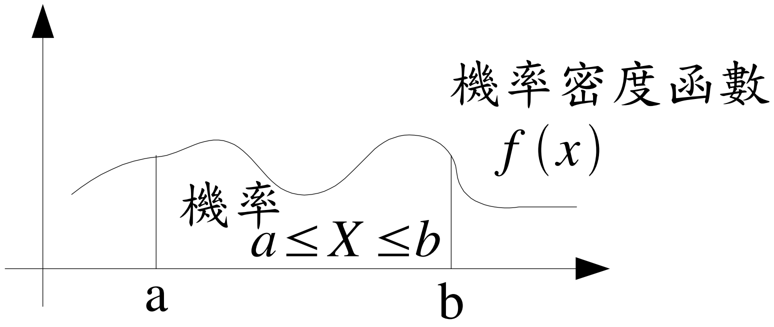
# 連續隨機變數 (Cont.)

- 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.)



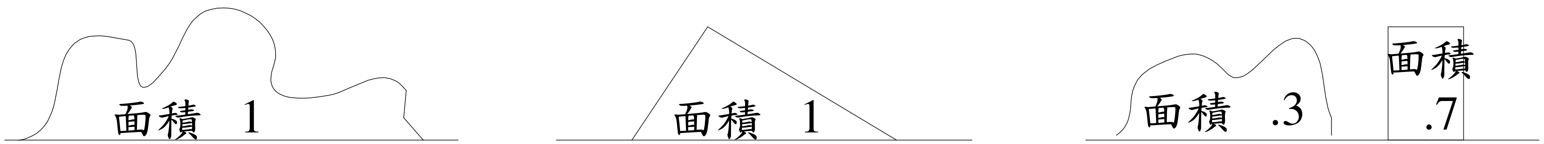

- 並非所有先前提到機率函數都可做為隨機變數 $X$ 的密度
- 必須注意的是，機率永遠不會是**負值**，且

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

- 因此， $f(x) \ge 0$ for all $x$

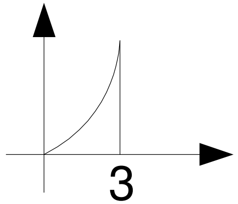
# 連續隨機變數 (Cont.)

- 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 $f(x)=\begin{cases} \dfrac{x^2}{9} & \text{where } 0\le x\le 3 \\ 0 & otherwise \end{cases}$




- 因 $f(x) \geq 0$ ，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 $f(x) = \begin{cases} \frac{x^2}{9} & \text{where } 0 \le x \le 3 \\ 0 & \textit{otherwise} \end{cases}$

- 可改寫成 => $f(x) = \frac{x^2}{9} \text{ where } 0 \le x \le 3$
- 但不可只寫 $f(x) = \frac{x^2}{9}$

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

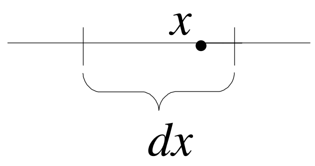
$$\int_1^2 \frac{x^3}{27}\,dx = \frac{x^3}{27}\Big|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

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

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

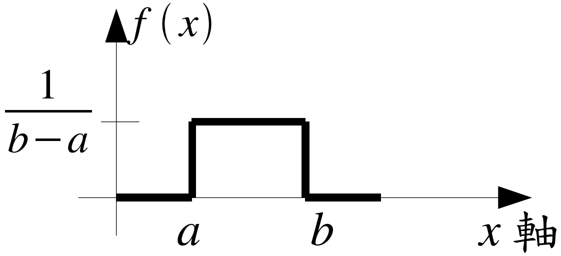
# 連續隨機變數 (Cont.)

- 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

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

# 分佈函數

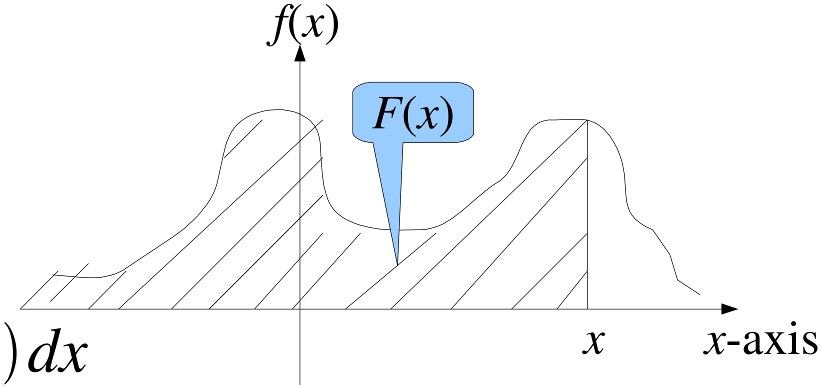
- 密度函數：用 $f(x)$ 表示
- 分佈函數：用 $F(x)$ 表示
- 何謂分佈函數？
  - 令 $f(x)$ 是隨機變數 $X$ 的密度函數，則這些機率是由一個函數所形成，此函數稱之為分佈函數，並用 $F(X)$ 表示之。

# 分佈函數 (*Cont.*)

- 隨機變數的 ( 累積 ) 分佈函數
  - 令 $x$ 的分佈函數為 $F(x)=P(X\leq x)$
  - 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

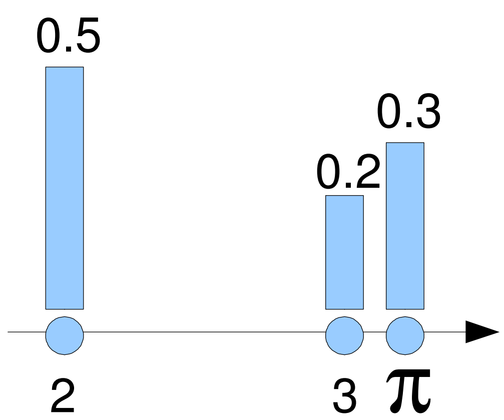



$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

  - 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 $\pi$ 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 $P(x) = \begin{cases} 0.5 & \text{if } x = 2 \\ 0.2 & \text{if } x = 3 \\ 0.3 & \text{if } x = \pi \end{cases}$

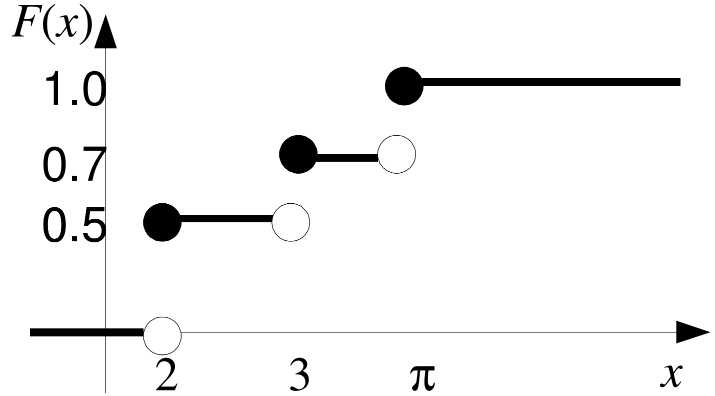


# 分佈函數(Cont.)

- 設 $F(x)$ 是 $X$ 的分佈函數，表示為累積機率

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

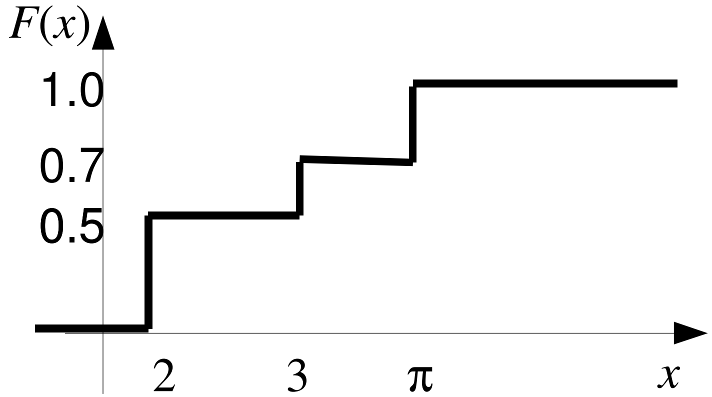
圖示




- 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

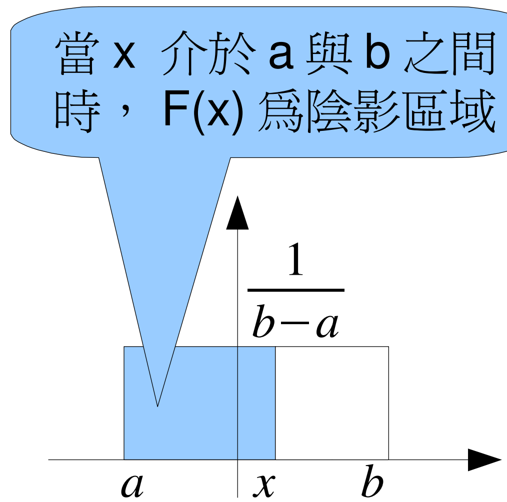



- 因應新圖表示法，分佈函數改為

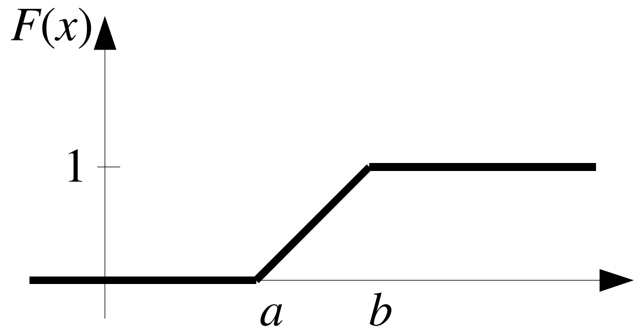
$$F(x)=\begin{cases} 0 & if \quad x\le 2 \\ 0.5 & if \quad 2\le \pi \le 3 \\ 0.7 & if \quad 3\le x\le \pi \\ 1 & if \quad x\ge \pi \end{cases}$$

# 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 $F(x) = 1$




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

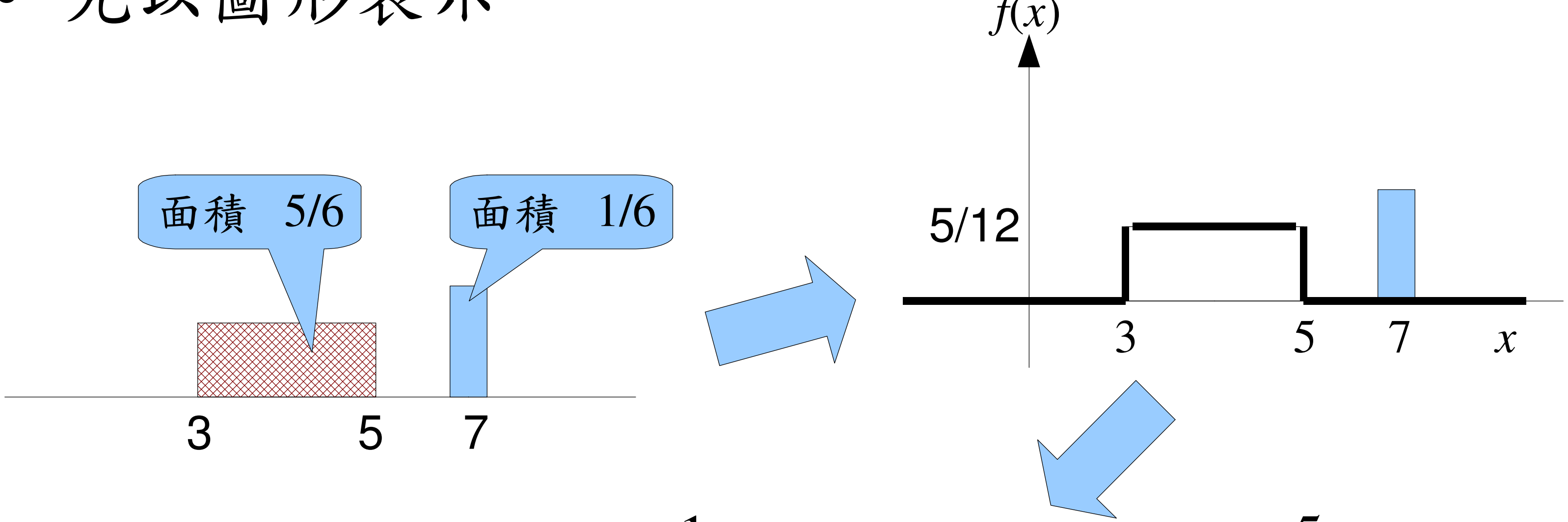


# 分佈函數 *(Cont.)*

- 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 )

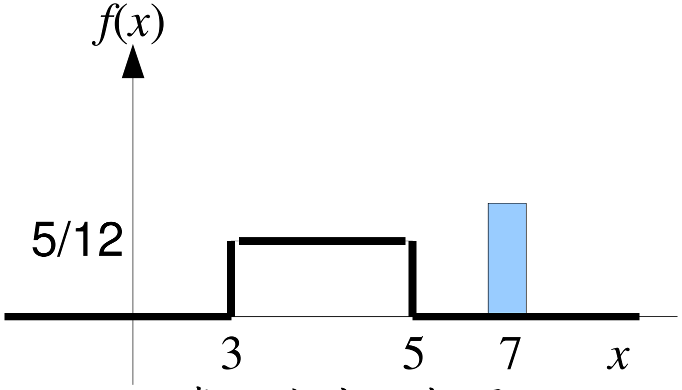
# 分佈函數(Cont.)

- $X$ 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

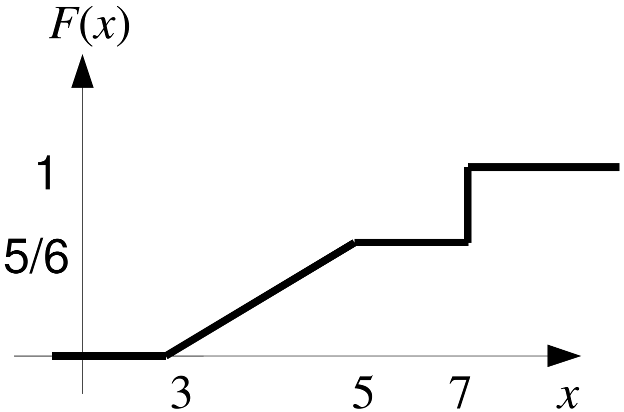



$$g(x)+\frac{1}{6}\delta(x-7)\ \text{where}\ g(x)=\frac{5}{12}\ \text{for}\ 3\le x\le 5$$

# 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

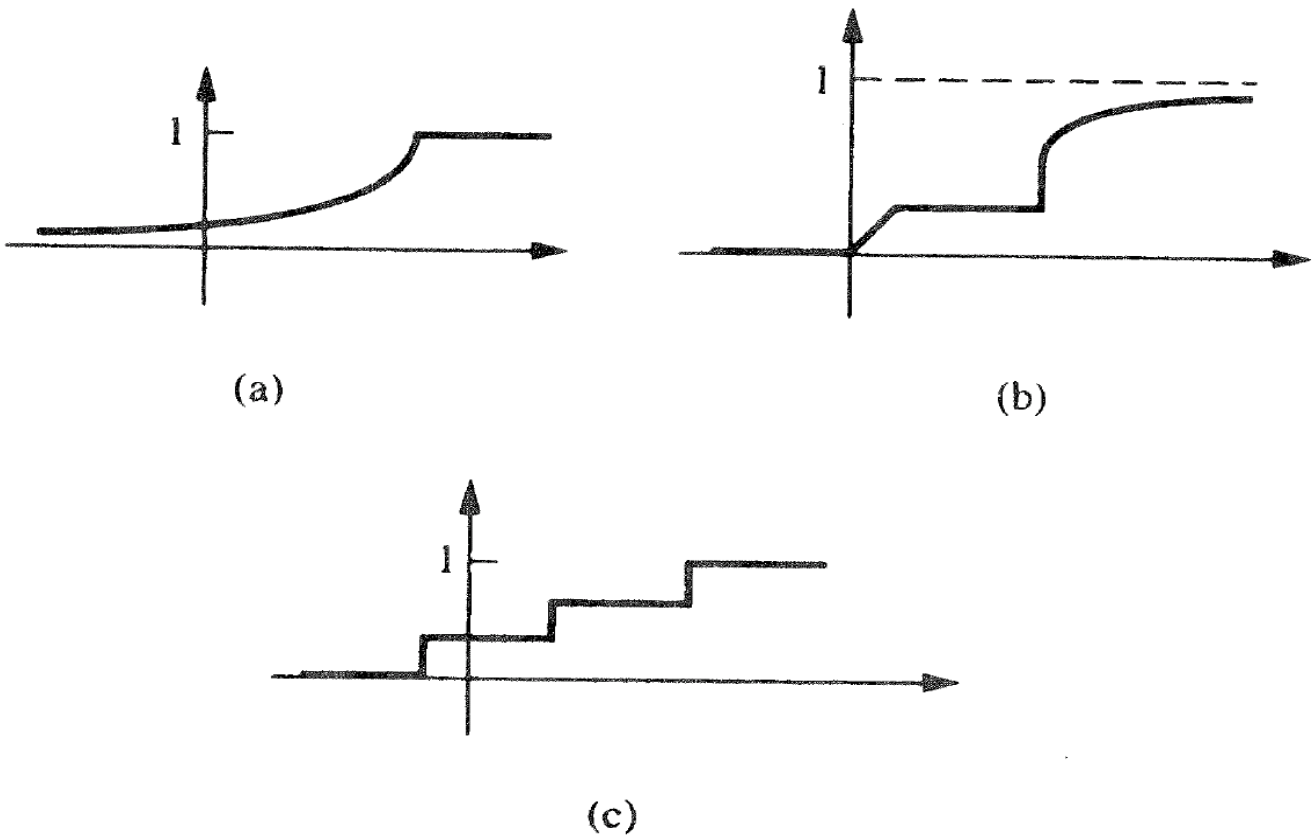



圖 9　典型的分佈函數

# 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

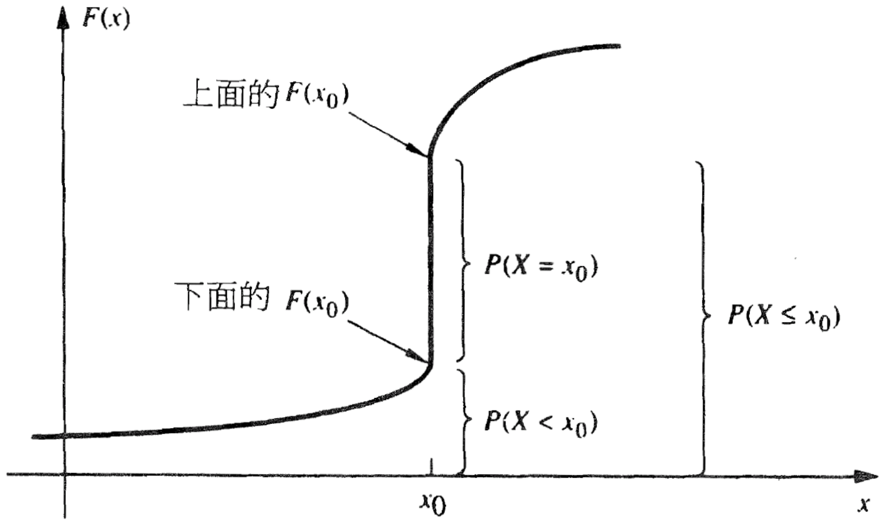



圖 10　分佈函數的跳躍

# 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 $F(x_0)$

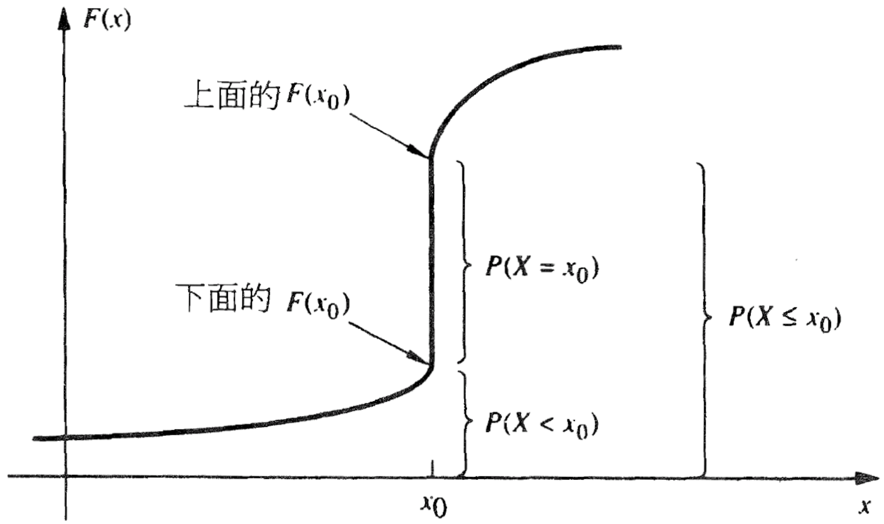



圖 10　分佈函數的跳躍

# 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 $F(x_0)$

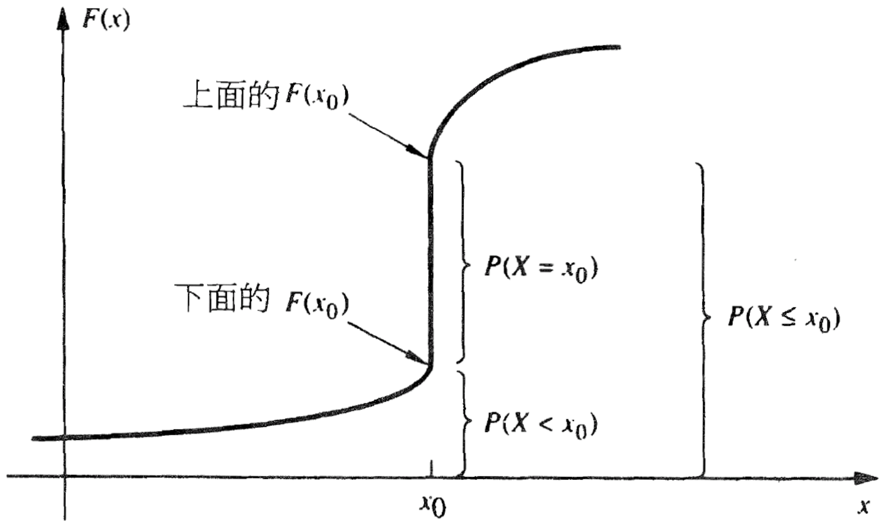



圖 10 分佈函數的跳躍

# 分佈函數(Cont.)

- 若跳躍的情況不是在一點 (e.g. $x_0$) 而是一個區間 (e.g. $[a, b]$)
  - $P(a \leqq X \leqq b) = P(X \leqq b) - P(X < a)$

    = 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 )