4. 隨機變數與機率函數

Random Variables and Probability Functions

- 學習隨機變數的概念與定義。
- 熟悉各種機率函數(PMF、PDF、CDF)。

4.1 隨機變數的觀念

- 統計實驗(statistical experiment)是用來描述一個會產生數個偶然性結果的實驗程序。
- 測試數個電子元件就是一個典型的統計實驗。

定義 4.1
隨機變數(random variable；r.v.) 是一個實值函數(real-valued function)，其定義域為樣本空間而其值域為實數。每個樣本空間的樣本點依照隨機變數都有一個定義的實數。

範例4.1

- A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

範例4.2

- Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. How to define a random variable X to describe the status of a component?

範例4.3

- Suppose a sampling plan involves sampling items from a process until a defective is observed. Establish the sample space, and find the value x of the random variable X that represents the number of items that are observed.

範例4.4

- Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. What is the range of x?

範例4.5

- Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. What is the range of x?

定義 4.2
如果一個樣本空間中包含有限多個樣本點，或者為可數之無窮多個樣本點，則我們稱其為離散的樣本空間 (discrete sample space)。

定義 4.3
如果樣本空間包含不可數之無窮多個的樣本點，其數目等於一實數線上所有的可能點，則我們稱其為連續的樣本空間 (continuous sample space)。

4.2 離散機率分佈(配)

離散型機率密度函數

- 數學函數 f(x) 稱為基於一離散隨機變數 X 的機率函數(probability function)、機率質量函數(probability mass function；PMF)、機率分佈(配)函數(probability distribution function)、機率分佈(配)(probability distribution)。
- What exactly is f(x)?

定義 4.4
f(x)為離散隨機變數 X 的機率函數，則對每個可能的結果 x，滿足

1. $f(x) \geq 0$　　2. $\sum_{x} f(x) = 1$　　3. $P(X = x) = f(x)$

範例4.6

- A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

範例4.7

- If 50% of a foreign brand of cars sold by an agency are equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

定義 4.5
若函數f(x)為一離散的隨機變數X的機率分佈，則該隨機變數的累積分佈(配)函數(cumulative distribution function；CDF) 定義為

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ , } \forall -\infty < x < \infty$$

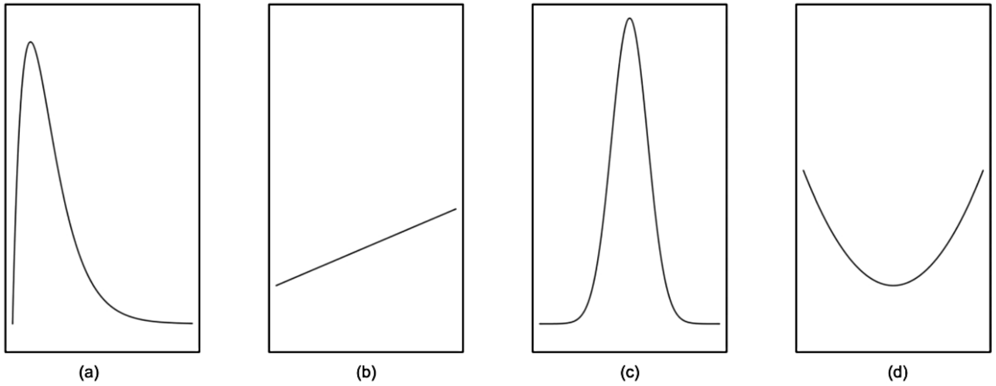
- What exactly is F(x)?

範例4.8

- Find the cumulative distribution function of the random variable X in Example 4.7. Using F(x), verify that f(2) = 3/8. Find F(0.7) and F(1.6). For the random variable X, make the following graphs: probability mass function plot, probability histogram, and discrete cumulative distribution function plot.

4.3 連續機率分佈(配)

- 若X為連續隨機變數，且用函數f(x)來表示其機率分佈，則f(x)稱為連續隨機變數X的機率密度函數(**probability density function；PDF)**，或簡稱密度函數(**density function)**。
- 雖然函數f(x)的定義域是在一連續隨機變數上，但在其值域的表現，f(x)可能會包含有限個不連續點。以下是幾個典型的密度函數的圖形，橫軸是x，縱軸是f(x)。
- f(x)被稱為機率密度函數，係因：
- For a continuous random variable, why do we consider the probability of an interval rather than a point value? Can you explain it using the following figures?



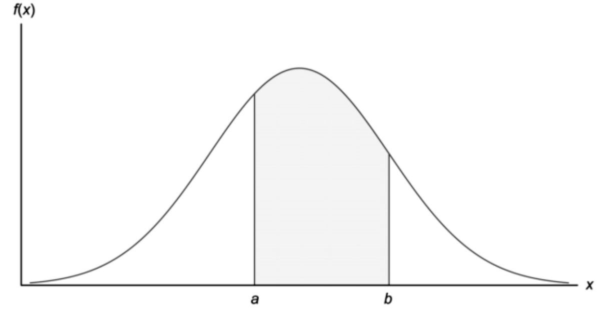
4.3 連續機率分佈(contd.)

- 若我們關心 $(a < X < b)$ 之事件，則該事件的機率值等於機率密度函數 f(x) 在座標 x = a 和 x = b 之間的陰影面積。由微積分的積分定義可得

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

$$P(a < X \le b) = ?$$

$$P(a \le X \le b) = ?$$



定義 4.6
函數f(x)是一個連續隨機變數之機率密度函數(probability density function)，則其定義域是整個實數，且其值滿足

1. $f(x) \ge 0 \text{ , } \forall x \in \Re$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$

範例4.9

- Suppose that the error in the reaction temperature, in ◦C, for a controlled laboratory experiment is a continuous random variable X having the probability density function defined as :

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Verify that f(x) is a density function.

(b) Find P(0 < X ≤ 1).

定義 4.7

若 f(x) 是一連續隨機變數 X 的密度函數，則 F(x) 其累積分佈函數 (cumulative distribution function)定義為 $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt, \quad -\infty < x < \infty$

- What about P(X < x) and P(a < X < b)?
- How to obtain f(x) from F(x)?

範例4.10

- For the density function of Example 4.9, find F(x), and use it to evaluate $P(0 < X \leq 1)$.

範例4.11

- The Department of Energy (DOE) puts projects out on bid and estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b \\ 0, & \text{elsewhere} \end{cases}$$

Find F(y) and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

4.4 聯合機率分佈

- 我們可能會需要數個隨機變數來記錄同時發生的實驗結果。比方，在測量受控制的化學藥品實驗中，化學反應所產生的化學沈澱物的質量P以及釋放出的氣體流量V，其變數組 (p, v) 組成了一個2維(two-dimensional)的樣本空間。
- 假設 X 和 Y 為 2 個離散隨機變數，對於 X 和 Y 它們同時發生的可能性的機率分佈，可以由一個聯合機率分佈函數(joint probability distribution) $f(x, y)$ 來表示，亦即在隨機變數 X 和 Y 的可能值域範圍內，對任一樣本點 (x, y) 都有一個對應的函數值 $f(x, y)$，使得 $f(x, y) = P(X = x, Y = y)$。So, what exactly is f(x, y)?

定義4.8
函數 $f(x, y)$ 是離散隨機變數 X 和 Y 的聯合機率分佈函數(joint probability distribution) 或是聯合機率質量函數 (joint probability mass function)，它滿足以下的條件：

1. $f(x, y) \geq 0, \forall (x, y)$
2. $\sum_x \sum_y f(x, y) = 1$
3. $P(X = x, Y = y) = f(x, y)$

For any region A is the xy plane, $P[(X, Y) \in A] = \sum\sum_{(x, y) \in A} f(x, y)$

範例4.12

- Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) the joint probability function $f(x, y)$

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) \mid x + y \le 1\}$

4.4 聯合機率分佈(contd.)

定義 4.9
函數 $f(x,y)$ 是 X 和 Y 連續隨機變數的聯合機率密度函數 (joint probability density function)，它滿足以下的條件：

1. $f(x,y) \geq 0,\ \forall (x,y)$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$
3. $P[(X,Y) \in A] = \iint_A f(x,y)\,dxdy$, for any region A in the xy plane.

範例4.13

- A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y , respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Verify condition 2 of Definition 4.9. (b) Find $P[(X,Y) \in A]$, where $A = \left\{(x,y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\right\}$

4.4 聯合機率分佈(contd.)

定義 4.10

已知 $f(x, y)$ 為 2 個離散隨機變數 X 和 Y 的聯合機率分佈，則對於隨機變數 X 和 Y 的邊際分佈函數(**marginal distribution function**)或邊際分佈分別為

$$g(x) = \sum_{y} f(x, y) \qquad h(y) = \sum_{x} f(x, y)$$

若是連續隨機變數， $g(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad h(y) = \int_{-\infty}^{\infty} f(x, y)dx$

範例4.14

- Show that the column and row totals of the joint probability distribution table, established in Example 4.12, give the marginal distribution of X alone and of Y alone.

範例4.15

- Find g(x) and h(y) for the joint probability density function of Example 4.13.

範例4.16

- The marginal distributions g(x) and h(y) are indeed the probability distributions of the individual variables X and Y, respectively. How to verify this fact? How to calculate $P(a < X < b)$ by the marginal distribution?

4.4 聯合機率分佈(contd.)

定義 4.11
令 X 和 Y 是兩個不論是離散的或是連續的隨機變數，則已知 X = x 時，隨機變數 Y 的條件機率分佈 **(conditional probability distribution)** 定義為

$$f(y \mid x) = \frac{f(x, y)}{g(x)}, \quad \text{假設 } g(x) \neq 0 \text{，即 } g(x) > 0$$

同理，已知 Y = y，隨機變數 X 的條件機率分佈為

$$f(x \mid y) = \frac{f(x, y)}{h(y)}, \quad \text{假設 } h(y) \neq 0 \text{，即 } h(y) > 0$$

How to derive the above equations in the discrete case?

4.4 聯合機率分佈(contd.)

- How to find the probability that the discrete random variable X falls between a and b when it is known that the discrete variable Y = y?
- How to find the above probability when X and Y are continuous random variables?

範例4.17

- Referring to Example 4.12, find the conditional distribution of X, given that Y = 1, and use it to determine $P(X = 0 \mid Y = 1)$, $P(X = 1 \mid Y = 1)$, and $P(X = 2 \mid Y = 1)$. How about $P(Y=1 \mid X = 0)$?

範例4.18

- The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Find the marginal densities g(x), h(y), and the conditional density f(y | x).

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

範例4.19

- Given the joint density function $\Longrightarrow$ $f(x,y)=\begin{cases}\dfrac{x(1+3y^2)}{4}, & 0<x<2,\ 0<y<1\\ 0, & \text{elsewhere}\end{cases}$

  find g(x), h(y), f(x | y), and evaluate $P(\frac{1}{4}<X<\frac{1}{2}\,|\,Y=\frac{1}{3})$.

定義 4.12
令 X 和 Y 是隨機變數(不管是離散或連續)，給定其聯合機率分佈 $f(x,y)$ 以及 X 和 Y 個別的邊際分佈g(x)以及h(y)，則此隨機變數 X 和 Y 稱為統計獨立**(statistical independence)**，若且唯若，對樣本空間範圍內的 (x,y) 而言， $f(x,y)=g(x)h(y)$

How to derive the above equation?

範例4.20

- Show that the random variables of Example 4.12 are not statistically independent.

定義 4.13
令 $X_1, X_2, \cdots, X_n$ 為一組不管是離散的或是連續的隨機變數，已知聯合密度分佈 $f(x_1, x_2, \ldots, x_n)$以及隨機變數之個別邊際分佈 $f_1(x_1), f_2(x_2), \cdots, f_n(x_n)$，則 $X_1, X_2, \cdots, X_n$ 稱為彼此統計獨立 **(mutually statistically independent)**，若且唯若，對樣本空間範圍內的任一 $(x_1, x_2, \cdots, x_n)$ 而言，滿足方程式 $f(x_1, x_2, \cdots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$

範例4.21

- Suppose that the shelf life, in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

Let X_1, X_2, X_3 represent the shelf lives for three of these containers selected independently. Find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.