6. 常用離散機率分佈

Common Discrete Probability Distributions

- 二項式分佈與多項式分佈。
- 幾何分佈。
- 負二項式分佈。
- 超幾何分佈。
- 卜瓦松分佈。

6.1 二項式分佈與多項式分佈

- 白努利程序 (Bernoulli Process)
  - 實驗由重複的試驗(trial)組成。
  - 每次試驗的結果可以被歸類為「成功」或「失敗」兩類。
  - 試驗的結果是成功的機率是不變的，不會因為做了多少次試驗而改變。
  - 每次試驗都是獨立的，與其他試驗的結果無關。

二項式分佈 (Binomial Distribution)

假設一白努利試驗其成功的機率為 p 而失敗的機率為 q = 1- p。令隨機變數 X 為在 n 次白努利試驗中成功的次數，則該離散二項式隨機變數 X 的機率分佈為

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \cdots, n$$

範例6.1

- The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

範例6.2

- A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%. (a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20? (b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

二項式分佈與多項式分佈(contd.)

定理 6.1

二項式分佈 $b(x;n,p)$ 的平均值 $\mu = np$，變異數 $\sigma^2 = npq$。

How to prove it?

範例6.3

- It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community. In order to gain some insight into the true extent of the problem, it is determined that some testing is necessary. It is too expensive to test all of the wells in the area, so 10 are randomly selected for testing. (a) Using the binomial distribution, what is the probability that exactly 3 wells have the impurity, assuming that the conjecture is correct? (b) What is the probability that more than 3 wells are impure?.

範例6.4

- Find the mean and variance of the binomial random variable of Example 6.1, and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

範例6.5

- Consider the situation of Example 6.3. The notion that 30% of the wells are impure is merely a conjecture put forth by the area water board. Suppose 10 wells are randomly selected and 6 are found to contain the impurity. What does this imply about the conjecture? Use a probability statement.

二項式分佈與多項式分佈(contd.)

多項式分佈 (Multinomial Distribution)

如果一個試驗有 k 種結果 $E_1, E_2, \cdots, E_k$，且這 k 種不同結果的發生機率依次為 $p_1, p_2, \cdots, p_k$，其中 $p_1 + p_2 + \cdots + p_k = 1$。在 n 次獨立試驗中，令隨機變數 $X_1, X_2, \cdots, X_k$ 分別代表結果為 $E_1, E_2, \cdots, E_k$ 的次數，則這些隨機變數的機率分佈為

$$f(x_1, x_2, \cdots, x_k; p_1, p_2, \cdots, p_k, n) = \binom{n}{x_1, x_2, \cdots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \quad \text{其中} \sum_{i=1}^{k} x_i = n \quad \text{及} \quad \sum_{i=1}^{k} p_i = 1$$

範例6.6

- For a certain airport with three runways, in the ideal setting of a computer simulation, the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet: Runway 1: p1 = 2/9, Runway 2: p2 = 1/6, Runway 3: p3 = 11/18. What is the probability that 6 randomly arriving airplanes are distributed in the following fashion? Runway 1: 2 airplanes, Runway 2: 1 airplane, Runway 3: 3 airplanes.

6.2 超幾何分佈

- 超幾何實驗 (Hypergeometric Experiment)
  - 從 N 個項目中隨機選取 n 個樣本，且其樣本都不置回。
  - 在 N 個樣本空間中，有 k 個項目被歸類為成功，而有(N – k)個項目被歸類為失敗。

超幾何分佈 (Hypergeometric Distribution)

超幾何實驗中，成功的數目是一個隨機變數 X，其機率分佈稱為超幾何分佈：$h(x;N,n,k)=\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$

Why x has this range?

其中 $\max\{0, n-(N-k)\}\le x\le \min\{n,k\}$。上式代表：在 N 個項目中隨機挑選 n 個樣本，其中 x 個項目是從 k 個項目中挑選，另外 (n – x) 個項目是從 (N – k) 個項目中挑選。

範例6.7

- A particular part that is used as an injection device is sold in lots of 10. The producer deems a lot acceptable if no more than one defective is in the lot. A sampling plan involves random sampling and testing 3 of the parts out of 10. If none of the 3 is defective, the lot is accepted. Comment on the utility of this plan.

範例6.8

- Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample, if there are 3 defectives in the entire lot? Comment on the sampling plan.

定理 6.2
超幾何分佈其平均值 $\mu = \frac{nk}{N}$ ，其變異數 $\sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1-\frac{k}{N}\right)$

範例6.9

- Find the mean and variance of the random variable of Example 6.8 and then use Chebyshev's theorem to interpret the interval $\mu \pm 2\sigma$.

範例6.10

- A manufacturer of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly blemished. If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished?

範例6.11

- A group of 10 individuals is used for a biological case study. The group contains 3 people with blood type O, 4 with blood type A, and 3 with blood type B. What is the probability that a random sample of 5 will contain 1 person with blood type O, 2 people with blood type A, and 2 people with blood type B?

6.3 負二項式分佈以及幾何分佈

負二項式分佈 (Negative Binomial Distribution)

考慮一白努利試驗，其試驗是成功的機率為 p、是失敗的機率為 q，p+q=1。考慮一重複進行之白努利試驗，假設隨機變數 X 代表獲得第 k 次成功試驗的實驗次數，則其機率分佈為

$$b^*(x;k,p)=\binom{x-1}{k-1}p^k q^{x-k}$$

$x=k, k+1, k+2, \cdots$ Why? What does it mean?

範例6.12

- In an NBA championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B. (a) What is the probability that team A will win the series in 6 games? (b) What is the probability that team A will win the series? (c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

6.3 負二項式分佈以及幾何分佈(contd.)

幾何分佈 (Geometric Distribution)

考慮一白努利試驗，其試驗是成功的機率為 p、是失敗的機率為 q，p+q=1。考慮一重複進行之白努利試驗，假設隨機變數 X 代表獲得第 1 次成功試驗的實驗次數，則其機率分佈為

$$g(x\,;\,p) = pq^{x-1} \quad x = 1, 2, 3, \cdots$$

範例6.13

- It may be of interest to know the number of attempts necessary in order to make a connection during a "busy time". Suppose that we let p = 0.05 be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

定理 6.3
幾何分佈其平均值 $\mu = \frac{1}{p}$ ，其變異數 $\sigma^2 = \frac{1-p}{p^2}$

- In Example 6.13, during a busy time, how many attempts are expected to make a connection?

6.4 卜瓦松分佈

- 假設隨機變數代表在給定一段時間或一個特定的區域內，其總共發生之事件總數，我們稱這類的實驗為卜瓦松實驗(Poisson experiment)。在卜瓦松實驗中，等待時間可以是任何的單位，例如一分鐘、一天、一星期、一個月、一年等。
- 卜瓦松程序(Poisson process)的特性
  - 卜瓦松程序是沒有記憶性的。
  - 在一個非常短的時間間隔或很小的區域中發生1次事件的機率，與該時間間隔的長度或該區域的大小成正比，而且該機率與在此時間間隔或此區域之外已經發生了多少次事件無關。
  - 在一個非常短的時間間隔或很小的區域中會發生2次或2次事件以上的機率很小，而且是可以忽略的。

卜瓦松分佈 (Poisson Distribution)

卜瓦松隨機變數 X 的機率分佈為 $p(x;\lambda t)=\dfrac{e^{-\lambda t}(\lambda t)^{x}}{x!},\quad x=0,1,2,\cdots$

其中 λ 為發生率，即每單位時間、距離、面積或體積的事件平均發生次數，而 $e=2.71828\cdots$。

範例6.14

- Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

定理 6.4
卜瓦松分佈的平均值與變異數皆為λt。

定理 6.5
令 X 為二項式隨機變數，其機率分佈為$b(x\,;n\,,\,p)$。
當 $n \to \infty\,,\,p \to 0$，且 $np \xrightarrow{n\to\infty} \mu$ 為常數時，則 $b(x\,;n\,,\,p) \xrightarrow{n\to\infty} p(x\,;\,\mu)$

範例6.15

- In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other. (a) What is the probability that in any given period of 400 days there will be an accident on one day? (b) What is the probability that there are at most three days with an accident?