Chapter 6

Some Important Discrete Distributions

6.1 BERNOULLI TRIALS

A sequence of trials is performed so that

- (a) for each trial, there are only two possible outcomes, say, success and failure;
- (b) the probabilities of the occurrence of these outcomes remain the same throughout the trials;
- (c) the trials are carried out independently.

S: success, $P(S) = p$, F: failure, $P(F) = q$, $p + q = 1$.

$$\begin{aligned} &SSFFSFSSS\cdots FF \\ &FSFSSFFFS\cdots SF \\ &\vdots \end{aligned}$$

$$\begin{aligned} P(SSFFSF\cdots FF) &= P(S)P(S)P(F)P(F)P(S)P(F)\cdots P(F)P(F) \\ &= ppqqpq\cdots qq. \end{aligned}$$

6.1.1 BINOMIAL DISTRIBUTION

- The probability distribution of a random variable X representing the number of successes in a sequence of n Bernoulli trials

$$\boxed{p_X(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n,}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

BINOMIAL DISTRIBUTION

Characteristics function $\phi_X(t) = (pe^{jt} + q)^n,$

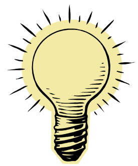
Mean $m_X = np,$

Variance $\sigma_X^2 = npq.$

燈泡壽命問題

Example 6.1. Problem: a homeowner has just installed 20 light bulbs in a new home. Suppose that each has a probability 0.2 of functioning more than three months.

(1)What is the probability that at least five of these function more than three months?



$$\sum_{k=5}^{20} p_X(k) = 1 - \sum_{k=0}^{4} p_X(k)$$
$$= 1 - \sum_{k=0}^{4} \binom{20}{k} (0.2)^k (0.8)^{20-k}$$
$$= 1 - (0.012 + 0.058 + 0.137 + 0.205 + 0.218) = 0.37.$$

(2) What is the average number of bulbs the homeowner has to replace in three months?

$$20 - E\{X\} = 20 - np = 20 - 20(0.2) = 16.$$

電話共用占線問題

Example 6.2. Suppose that three telephone users use the same number and that we are interested in estimating the probability that more than one will use it at the same time. If independence of telephone habit is assumed, the probability of exactly k persons requiring use of the telephone at the same time is given by the mass function $p_x(k)$ associated with the binomial distribution. Let it be given that, on average, a telephone user is on the phone 5 minutes per hour; an estimate of p is

每個使用者平均每小時使用5分鐘 $p = \frac{5}{60} = \frac{1}{12}.$

同時超過一個人以上使用電話機率

$$p_X(2) + p_X(3) = \binom{3}{2}\left(\frac{1}{12}\right)^2\left(\frac{11}{12}\right) + \binom{3}{3}\left(\frac{1}{12}\right)^3$$
$$= \frac{11}{864} = 0.0197.$$

Example 6.3. Problem: let X_1 and X_2 be two independent random variables, both having binomial distributions with parameters (n_1, p) and (n_2, p) respectively, and let $Y = X_1 + X_2$. Determine the distribution of random variable Y.

Answer: the characteristic functions of X_1 and X_2 are,

$$\phi_{X_1}(t) = (pe^{jt} + q)^{n_1}, \phi_{X_2}(t) = (pe^{jt} + q)^{n_2}.$$

$$\begin{aligned}\phi_Y(t) &= \phi_{X_1}(t)\phi_{X_2}(t) \\ &= (pe^{jt} + q)^{n_1+n_2}.\end{aligned}$$

By inspection, it is the characteristic function corresponding to a binomial distribution with parameters $(n_1 + n_2, p)$. Hence, we have

$$p_Y(k) = \binom{n_1+n_2}{k} p^k q^{n_1+n_2-k}, \quad k = 0, 1, \ldots, n_1 + n_2.$$

Theorem 6.1: The binomial distribution generates itself under addition of independent random variables with the same p.

$$X_1 \sim B(n_1, p)$$
$$X_2 \sim B(n_2, p)$$
$$X_1 + X_2 \sim B(n_1 + n_2, p)$$

Example 6.4. Problem: if random variables X and Y are independent binomial distributed random variables with parameters (n_1, p) and (n_2, p), determine the conditional probability mass function of X given that

$$X + Y = m, \qquad 0 \le m \le n_1 + n_2.$$

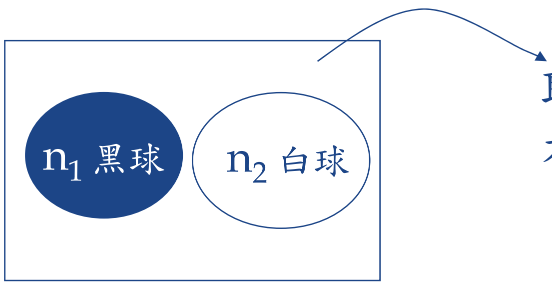
Answer: for $k \le \min(n_1, m)$, we have

$$\begin{aligned} P(X = k | X + Y = m) &= \frac{P(X = k \cap X + Y = m)}{P(X + Y = m)} \\ &= \frac{P(X = k \cap Y = m - k)}{P(X + Y = m)} = \frac{P(X = k)P(Y = m - k)}{P(X + Y = m)} \\ &= \frac{\binom{n_1}{k} p^k (1-p)^{n_1 - k} \binom{n_2}{m-k} p^{m-k} (1-p)^{n_2 - m + k}}{\binom{n_1 + n_2}{m} p^m (1-p)^{n_1 + n_2 - m}} \\ &= \boxed{\binom{n_1}{k}\binom{n_2}{m-k}\Big/\binom{n_1 + n_2}{m}}, \quad k = 0, 1, \ldots, \min(n_1, m), \end{aligned}$$

Hypergeometric Distribution 超幾何分佈

將上式 $n_2 \to n - n_1$

$$\boxed{p_Z(k) = \binom{n_1}{k}\binom{n - n_1}{m - k}\Big/\binom{n}{m}, \quad k = 0, 1, \ldots, \min(n_1, m).}$$



n 個球

取出 m 個球，
有 k 個黑球的機率

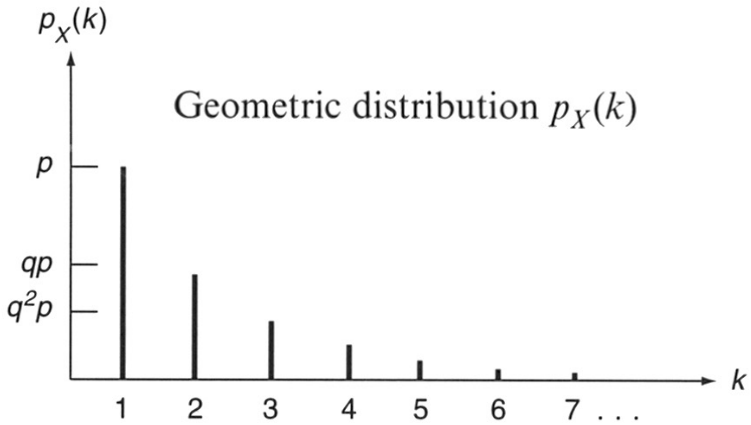
6.1.2 GEOMETRIC DISTRIBUTION 幾何分佈

- Another event of interest arising from Bernoulli trials is the number of trials to (and including) the first occurrence of success.

第一次成功出現，已嘗試的次數

$$\boxed{\begin{aligned} p_X(k) &= P(\underbrace{FF\ldots F}_{k-1}S) = P\underbrace{(F)P(F)\ldots P(F)}_{k-1}P(S) \\ &= q^{k-1}p, \quad k = 1, 2, \ldots . \end{aligned}}$$

Geometric Distribution



m: 最接近 x 的正整數

PDF $$F_X(x) = \sum_{k=1}^{m \le x} p_X(k) = p + qp + \cdots + q^{m-1}p$$

$$= (1-q)(1+q+q^2+\cdots+q^{m-1}) = 1-q^m,$$

$$\frac{1-q^m}{1-q}$$

Geometric Distribution

Mean

$$E\{X\} = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty}\frac{\mathrm{d}}{\mathrm{d}q}q^k$$
$$= p\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty}q^k = p\frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) = \frac{1}{p}.$$

Variance

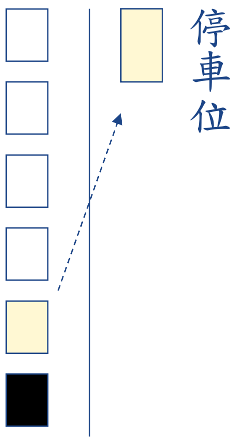
$$\sigma_X^2 = \frac{1-p}{p^2}.$$

Example 6.5. 停車位問題

A driver is eagerly eyeing a precious parking space some distance down the street. There are five cars in front of the driver, each of which having a probability 0.2 of taking the space. What is the probability that the car immediately ahead will enter the parking space?

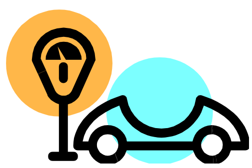
Answer:

a geometric distribution $p_x(k)$ for $k = 5$ and $p = 0.2$.



前面5台車，被第5台停走車位之機率

$$p_X(5) = (0.8)^4(0.2) = 0.82,$$



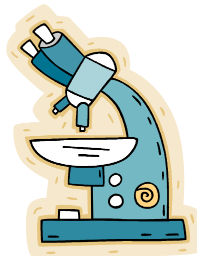
Example 6.6. 標本問題

Assume that the probability of a specimen failing during a given experiment is 0.1.
What is the probability that it will take more than three specimens to have one surviving the experiment?

取超過 3 個以上標本，才有第一個存活之機率

Answer: let X denote the number of trials required for the first specimen to survive. It then has a geometric distribution with $p = 0.9$.
The desired probability is

$$P(X > 3) = 1 - F_X(3) = 1 - (1 - q^3) = (0.1)^3 = 0.001.$$

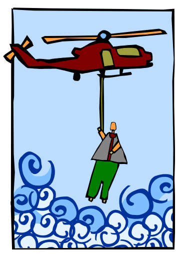


Example 6.7. 洪水頻率問題

let the probability of occurrence of a flood of magnitude greater than a critical magnitude in any given year be 0.01.
Assuming that floods occur independently, determine $E\{N\}$, the average return period.

The *average return period is* defined as the average number of years between floods for which the magnitude is greater than the critical magnitude.

Answer: it is clear that N *is* a random variable with a geometric distribution and $p = 0.01$. The return period is then



$$E\{N\} = \frac{1}{p} = 100 \text{ years.}$$

6.1.3 Negative Binomial Distribution

A natural generalization of the geometric distribution is the distribution of random variable X representing the number of Bernoulli trials necessary for the r-th success to occur, where r is a given positive integer.

- A : be the event that the first $k-1$ trials yield exactly $r-1$ successes, regardless *of* their order,
- B the event that a success turns up at the k-th trial.
  Then, owing to independence,

第 r 次成功時嘗試的次數 k

$r-1$ 成功 | 成功

$k-1$ trials

$$p_X(k) = P(A \cap B) = P(A)P(B).$$

$$P(A) = \binom{k-1}{r-1} p^{r-1} q^{k-r}, \quad k = r, r+1, \ldots,$$

$$P(B) = p.$$

$$\boxed{p_X(k) = \binom{k-1}{r-1} p^r q^{k-r}, \quad k = r, r+1, \ldots.}$$

Negative Binomial Distribution

- $Y = X - r$ 第 r 次成功時，已經失敗的次數
- The random variable Y is the number of Bernoulli trials *beyond* r needed for the realization of the r-th success, or it can be interpreted as the number of failures before the r-th success.
- replacing k by $m + r$

$\boxed{Y: m \quad X: k}$

$$p_Y(m) = \binom{m+r-1}{r-1} p^r q^m$$

$$= \binom{m+r-1}{m} p^r q^m, \quad m = 0, 1, 2, \ldots.$$

$$\binom{m+r-1}{m} = (-1)^m \binom{-r}{m}.$$

$$\boxed{p_Y(m) = \binom{-r}{m} p^r (-q)^m, \quad m = 0, 1, 2, \ldots,}$$

Negative Binomial Distribution

$$X = X_1 + X_2 + \cdots + X_r,$$

- X_j is the number of trials between the $(j-1)$-th and (including) the j-th successes.

X_j: 第 $j-1 \rightarrow j$ 次成功，嘗試的次數 (Geometric distribution)

$$m_{X_j} = \frac{1}{p},\ \sigma_{X_j}^2 = \frac{1-p}{p^2}$$

$$\Rightarrow \quad m_X = \frac{r}{p}, \quad \sigma_X^2 = \frac{r(1-p)}{p^2}.$$

Example 6.8. 停車塔問題
a curbside parking facility has a capacity for three cars. Determine the probability that it will be full within 10 minutes. It is estimated that 6 cars will pass this parking space within the timespan and, on average, 80% of all cars will want to park there.

第 r 次成功時嘗試的次數 k

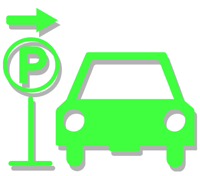
Answer: the desired probability is simply the probability that the number of trials to the third success (taking the parking space) is less than or equal to 6. If X is this number, it has a **negative binomial distribution** with $r = 3$ and $p = 0.8$. Using Equation (6.21), we have

Negative Geometric

$$P(X \le 6) = \sum_{k=3}^{6} p_X(k) = \sum_{k=3}^{6} \binom{k-1}{2} (0.8)^3 (0.2)^{k-3}$$

$$= (0.8)^3[1 + (3)(0.2) + (6)(0.2)^2 + (10)(0.2)^3]$$

$$= 0.983.$$



第3台時滿+第4台時滿+第5台時滿+第6台時滿

6.2 MULTINOMIAL DISTRIBUTION

- Bernoulli trials → only two possible outcomes for each trial.
- Let there be r possible outcomes for each trial, denoted by $E_1, E_2, \ldots, E_r$., and let $P(E_i) = pi, i = 1, \ldots, r$,
- $p_1 + p_2 + \ldots + p_r = 1$.
- A typical outcome of n trials is a succession of symbols such as:

$$E_2 E_1 E_3 E_3 E_6 E_2 \ldots.$$

$$\boxed{p_{X_1 X_2 \ldots X_r}(k_1, k_2, \ldots, k_r) = \frac{n!}{k_1! k_2! \ldots k_r!} p_1^{k_1} p_2^{k_2} \ldots p_r^{k_r},}$$

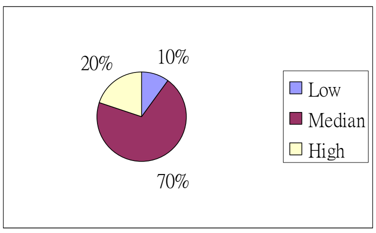
$$m_{X_i} = np_i, \quad \sigma_{X_i}^2 = np_i(1 - p_i)$$

- **Example 6.10.** 收入階級問題
  income levels are classified as low, medium, and high in a study of incomes of a given population. If, on average, 10% of the population belongs to the low-income group and 20% belongs to the high-income group,
  1. What is the probability that, of the 10 persons studied, 3 will be in the low-income group and the remaining 7 will be in the medium-income group?
  2. What is the marginal distribution of the number of persons (out of 10) at the low-income level?

---

(1)

$$p_{X_1 X_2 X_3}(3, 7, 0) = \frac{10!}{3!7!0!}(0.1)^3(0.7)^7(0.2)^0 \cong 0.01.$$



(2)

The marginal distribution of X_1 is binomial with $n = 10$ and $p = 0.1$.

6.3 POISSON DISTRIBUTION

- It is used in mathematical models for describing, in a specific interval of time, such events as
  - the emission of a particles from a radioactive substance, 放射物質
  - passenger arrivals at an airline terminal, 旅客到達數
  - the distribution of dust particles reaching a certain space, 灰塵掉落數
  - car arrivals at an intersection, and many other similar phenomena. 十字路口車輛數

Poisson Distribution

$$p_k(0,t) = P[X(0,t) = k], \quad k = 0, 1, 2, \ldots,$$

$$\boxed{p_k(0,t) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \ldots.}$$

$$\boxed{\nu = \lambda t} \qquad \boxed{p_k(0,t) = \frac{\nu^k e^{-\nu}}{k!}, \quad k = 0, 1, 2, \ldots.}$$

Poisson Distribution

Mean $E\{X(0,t)\} = \sum_{k=0}^{\infty} kp_k(0,t) = \mathrm{e}^{-\lambda t}\sum_{k=0}^{\infty}\frac{k(\lambda t)^k}{k!}$

$$= \lambda t\mathrm{e}^{-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t\mathrm{e}^{-\lambda t}\mathrm{e}^{\lambda t} = \lambda t.$$

$$\boxed{e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}}$$

Variance $\sigma^2_{X(0,t)} = \lambda t.$

λ : the average number of arrivals per unit interval of time
mean rate of arrival

Example 6.12.
let X_1 and X_2 be two independent random variables, both having Poisson distributions with parameters ν_1 and ν_2 respectively, and let $Y = X_1 + X_2$. Determine the distribution of Y.

Answer: we proceed by determining first the characteristic functions of X_1 and X_2. They are

$$\boxed{\sum_{k=0}^{\infty}\frac{(e^{jt}\nu_1)^k}{k!} = \exp\left(e^{jt}\nu_1\right)}$$

$$\phi_{X_1}(t) = E\{\mathrm{e}^{\mathrm{j}tX_1}\} = \mathrm{e}^{-\nu_1}\sum_{k=0}^{\infty}\frac{\mathrm{e}^{\mathrm{j}tk}\nu_1^k}{k!}$$

$$= \exp[\nu_1(\mathrm{e}^{\mathrm{j}t}-1)]$$

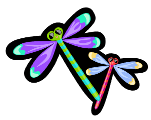
同理 $\phi_{X_2}(t) = \exp[\nu_2(\mathrm{e}^{jt}-1)].$

$$\phi_Y(t) = \phi_{X_1}(t)\phi_{X_2}(t) = \exp[(\nu_1+\nu_2)(\mathrm{e}^{jt}-1)].$$

$$p_Y(k) = \frac{(\nu_1+\nu_2)^k\exp[-(\nu_1+\nu_2)]}{k!}, \quad k = 0,1,2,\ldots.$$

Theorem 6.2: the Poisson distribution generates itself under addition of independent random variables.

Example 6.13. Problem: suppose that the probability of an insect laying r eggs is $v^r e^{-v}/r!$, $r = 0, 1, \dots$, and that the probability of an egg developing is p. Assuming mutual independence of individual developing processes, show that the probability of a total of k survivors is $(pv)^k e^{-pv}/k!$.



昆蟲孵蛋問題

Answer:
X be the number of eggs laid by the insect 孵蛋數目
Y be the number of eggs developed. 成功數
Then, given $X = r$, the distribution of Y is binomial with parameters r and p. Thus,

$$P(Y = k | X = r) = \binom{r}{k} p^k (1-p)^{r-k}, k = 0, 1, \dots, r.$$

$$P(Y=k) = \sum_{r=k}^{\infty} P(Y=k|X=r)P(X=r) = \sum_{r=k}^{\infty} \binom{r}{k} \frac{p^k(1-p)^{r-k}\nu^r e^{-\nu}}{r!}.$$

If we let $\boxed{r = k + n,}$ 改變下界

$$P(Y=k) = \sum_{n=0}^{\infty} \binom{n+k}{k} \frac{p^k(1-p)^n \nu^{n+k} \mathrm{e}^{-\nu}}{(n+k)!}$$

$$= \frac{(p\nu)^k \mathrm{e}^{-\nu}}{k!} \sum_{n=0}^{\infty} \frac{(1-p)^n \nu^n}{n!}$$

$$= \frac{(p\nu)^k \mathrm{e}^{-\nu} \mathrm{e}^{(1-p)\nu}}{k!} = \frac{(p\nu)^k \mathrm{e}^{-p\nu}}{k!}, \quad k = 0, 1, 2, \ldots.$$

- END -