

§ 2.2 离散型随机变量

定义2.3 若随机变量 X 取值有限或可列多个,则称 X 是离散型。设 X 的可能取值为 $x_1, x_2, \dots, x_n, \dots$ 且
 $P(X=x_k)=p_k, (k=1, 2, \dots)$ (2.2.1)

则称(2.2.1)为离散型随机变量 X 的分布律或概率分布或概率函数 可表为表格

X	x_1	x_2	\dots	x_k	\dots
P	p_1	p_2	\dots	p_k	\dots

$$X \sim \begin{pmatrix} x_1, x_2, \dots, x_k, \dots \\ p_1, p_2, \dots, p_k, \dots \end{pmatrix}$$

分布律的性质

1° **非负性**: $p_k \geq 0, k = 1, 2, \dots$

2° **归一性**: $\sum_{k=1}^{\infty} p_k = 1$

3° $P(X \in G) = \sum_{x_k \in G} p_k$

所有属于**G**中的那些 **x_k** 所对应的 **p_k** 求和。

注 1) 性质**1**与**2**是概率分布的特征: 任意满足性质**1**与**2**的 **p_k** 都存在某一个离散型随机变量**X**以及 **$\{x_k\}$** 使得 **$p_k = P(X = x_k), k = 1, 2, \dots$**

注2:若离散型随机变量**X**的分布律为 **$p_k = P(X = x_k), k = 1, 2, \dots$** , 那么其分布函数为 **$F(x) = \sum_{x_k \leq x} p_k$** .

利用分布律的性质计算参数

例：设随机变量 X 的分布律为

$$P\{X = n\} = c \left(\frac{1}{4}\right)^n \quad (n = 2, 3, 4, \dots)$$

试求常数 c .

解：由分布律的归一性得

$$1 = \sum_{n=2}^{\infty} P\{X = n\} = \sum_{n=2}^{\infty} c \left(\frac{1}{4}\right)^n$$

$$\Rightarrow c = 12.$$

常见的离散型随机变量

1、几何分布

我们以例子来说明。

例 对一目标进行射击，直到击中为止。设每次射击的击中率为 p ($0 < p < 1$)。求射击次数 X 的分布律。

解 设 $q = 1 - p$, A_i 表 “第 i ($i = 1, 2, \dots$) 次击中目标”，则 X 的可能取值为全部正整数。对每个正整数 $k \in N$ ，有

$$\begin{aligned}
 P(X = k) &= P(\overline{A_1} \overline{A_2} \text{L} \overline{A_{k-1}} A_k) \\
 &= P(\overline{A_1}) P(\overline{A_2}) \text{L} P(\overline{A_{k-1}}) P(A_k) \\
 &= pq^{k-1},
 \end{aligned}$$

于是X 的分布律为

X	1	2	3	L	k	L
p_k	p	pq	pq^2	L	pq^{k-1}	L

我们称具有如上分布律的随机变量 X 服从参数为 p 的几何分布。

定义 设随机变量 X 一切可能的取值为

$1, 2, 3, \dots$, 并且取这些值的概率为

$$P(X = k) = pq^{n-1}, k = 1, 2, \dots, q = 1 - p.$$

则称 X 服从参数为 p 的几何分布, 记为 $X : G(p)$ 。

2、超几何分布

定义 设 N, n, m 为正整数且

$$n \leq N, m \leq N$$

若随机变量 X 的分布律为

$$P(X = k) = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}, \quad k = 0, 1, 2, \dots, n,$$

则称 X 服从参数为 n, m, N 的超几何分布, 记为

$$X : H(n, m, N).$$

由组合的知识知，若 $X : H(n, m, N)$ ，则

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \frac{C_m^k C_{N-m}^{n-k}}{C_N^n} = 1。$$

分析：上式成立等价于

$$\sum_{k=0}^n C_m^k C_{N-m}^{n-k} = C_m^0 C_{N-m}^n + C_m^1 C_{N-m}^{n-1} + \dots + C_m^n C_{N-m}^0 = C_N^n。$$

上式右端可认为是从 N 个球中取出 n 个球的取法总数，而从 N 个球中取出 n 个球这个事件可被分解为如下事件：取出 0 个红球， n 个白球；取出 1 个红球， $n-1$ 个白球；...；取出 n 个红球，0 个白球。这些取法数的总和即上式的左端。

3、二项分布

定义 (贝努利 Bernoulli 试验 , n 重独立试验, n 重 Bernoulli 试验)

- 1) 只有两个可能结果的试验称为Bernoulli 试验;
- 2) 把一个随机试验重复进行 n 次, 如果试验的结果互不影响, 则称这样的试验为 n 重独立试验;
- 3) 若在 n 重独立试验中, 每次试验只有两个可能的结果: 事件 A 或事件 \overline{A} 发生, 则称这样的试验为 n 重 Bernoulli 试验; 相应的数学模型叫Bernoulli 概型。

例如：抛硬币就是一个Bernoulli 试验；将一个硬币重复地抛100次，就是一个100重Bernoulli试验。

例如：射击，让A表示“击中靶”，则就是一个Bernoulli 试验；将射击10次，就可看着是一个10重Bernoulli试验。

定理：在贝努力试验中，事件A发生的概率 $P(A) = p (0 < p < 1)$ 。设X表示n重贝努力试验中A发生的次数，则

$$P_n(k) @ P(X=k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, L, n.$$

定义 设随机变量 X 的可能取值为 $0,1, 2,...,n$ ， 且

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, k = 0,1,L ,n$$

则称 X 服从参数为 n, p 的**二项分布**，记为 $X \sim B(n,p)$

特别地,当 $n=1$ 时，二项分布 $X \sim B(1,p)$ ，
即为**(0-1)分布**.

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

例题 一学生参加一门由10道四选一的选择题组成的考试。由于该学生平时没有认真学习，因而他做每道选择题只能完全凭猜测。求

- 1) 正好猜对3道题的概率；
- 2) 及格的概率；
- 3) 至多猜对两道题的概率。

解：设 X 为该学生猜对的题数由于是四选一，因此，猜对一道选择题的概率为 $p = 1/4$ 由题意知

$$X : B(10, \frac{1}{4})$$

$$\mathbf{1) \quad } P(X = 3) = C_{10}^3 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^7 \quad \mathbf{B \ 0.2503;}$$

$$\mathbf{2) \quad } P(X \geq 6) = \sum_{k=6}^{10} C_{10}^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10-k} \quad \mathbf{B \ 0.0197;}$$

$$\mathbf{3) \quad } P(X \leq 2) = \sum_{k=0}^2 C_{10}^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{10-k} \quad \mathbf{B \ 0.5256.}$$

我们可以计算出上面例子中 X 的分布律为

X	0	1	2	3	4
p_k	0.0563	0.1877	0.2816	0.2503	0.1460

5	6	7	8	9	10
0.0584	0.0162	0.0031	0.0004	0.0000	0.0000

观察上面 X 的分布情况可见, X 取各种值的概率是不尽相同的, 更仔细地观察可以发现一条规律是: 随着 X 的取值 k 从0开始增加, $P(X = k)$ 先从小变到大, 然后再从大变到小。我们就称使概率 $P(X = k)$ 达到最大的 k 为最可能次数, 相应的事件“ $X = k$ ”称为最可能事件。

一般地，若 $X : B(n, p)$ ，则当 $(n+1)p$ 为整数时， X 有两个最可能次数 $(n+1)p$ 及 $(n+1)p-1$ ；当 $(n+1)p$ 不是整数时，最可能次数为 $(n+1)p$ 的整数部分。

例如，在上面例子中，由于

$$(n+1)p = \frac{11}{4} = 2.75$$

故最可能次数为2。

二项分布的最有可能次数

若 $X \sim B(n, p)$, 则

$$\frac{p_n(k)}{p_n(k-1)} = \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-k+1}} = 1 + \frac{(n+1)p - k}{kq},$$

$$q = 1 - p.$$

故当 $k < (n+1)p$ 时, $p_n(k) > p_n(k-1)$;

当 $k > (n+1)p$ 时, $p_n(k) < p_n(k-1)$.

当 $(n+1)p = k$ 为正整数时, $p_n(k) = p_n(k-1)$,

此时该两项为最大值。

当 $(n+1)p$ 不为正整数时, 取 $k = [(n+1)p]$,

这时 $p_n(k)$ 最大。

超几何分布与二项分布的关系

模型：一袋中装有 N 个小球，其中有 m 个红球，余下的全为白球。现从袋中任意抽取 n ($n \leq N$) 个球，问所取的球中恰有 k 个红球的概率为多少？

分析：如下考虑此摸球模型：抽取 n 次，每次一球。设变量 X 表示“所取的 n 个球中红球的个数”。若抽取是**无放回的**，则 $X \sim H(n, m, N)$ 。若每次抽取是**有放回的**，设 A 表示“抽取是红球”，则 n 次抽球就是 n 重贝努利试验，因此 $X \sim B(n, p)$ ， $p = m/N$ 。

当 N 很大时，可以证明

$$\frac{C_m^k C_{N-m}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}, p = \frac{m}{N}.$$

4、泊松 (Poisson) 分布

定义 若随机变量 \mathbf{X} 可能的取值为 $\mathbf{0,1,2,...}$, 且

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots (\lambda > 0)$$

则称 \mathbf{X} 服从参数为 λ 的泊松分布, $\mathbf{X} \sim \mathbf{P}(\lambda)$.

二项分布以泊松分布为极限分布

定理2.2(泊松定理) 设随机变量 $X_n \sim B(n, p_n)$,

且满足 $np_n = \lambda$ 则

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda},$$

证: $C_n^k p_n^k (1-p_n)^{n-k}$

$$= \frac{1}{k!} \cdot n(n-1)\dots(n-k+1) \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\dots(n-k+1)}{n^k} \cdot \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n})^k}$$

Diagram annotations: A red box highlights the fraction $\frac{n(n-1)\dots(n-k+1)}{n^k}$ with an arrow pointing to a red **1**. A red circle highlights the term $(1 - \frac{\lambda}{n})^k$ in the denominator with an arrow pointing to a red **1**.

$\therefore \lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda}$ 得证.

Poisson定理表明，当 n 很大而 p 很小时，二项分布 $B(n, p)$ 的概率函数与Poisson分布 $P(\lambda)(\lambda = np)$ 的概率函数近似地相等，即

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{where } \lambda = np.$$

因此，当 n 充分大 (≥ 50)， p 很小 (≤ 0.1) 时，可用参数为 $\lambda = np$ 的Poisson分布来近似地描述 $B(n, p)$ 的规律。利用这一近似公式，通过查表，可以比较容易地计算二项分布中的概率。

关于查表的问题

Poisson分布表见书上的附表2，它表示的是

$$1 - F(x-1) = \sum_{r=x}^{r=\infty} \frac{e^{-\lambda} \lambda^r}{r!} = P(X \geq x).$$

例如 设 $X : P(3.5)$ ，查表知

$$P(X \geq 2) = 0.864112; \quad P(X \geq 5) = 0.27455;$$

则

$$\begin{aligned} P(1.7 \leq X \leq 4.7) &= P(X \geq 2) - P(X \geq 5) \\ &= 0.864112 - 0.27455 = 0.589577. \end{aligned}$$

相应地，若 $X : P(4.5)$ ，同样的方法计算

$$P(0.3 \leq X \leq 8.5).$$

例2.12

某网吧有**300**台电脑,每台电脑的上网人因各种原因需要网管帮助的概率为**0.01**,现在有两种方式配备网管:

A:配备**10**名网管,每人负责**30**台电脑;

B:配备**8**名网管,共同负责**300**台电脑;

(1)证明:方式**B**比方式**A**效果好;

(2)若只需要方式**B**下有上网人得不到及时帮助的概率小于**0.02**,则**8**名网管可减少至几名?

例2.12

证明: 设 p_1, p_2 分别为两种方式下有人得不到帮助的概率, 则只需证 $p_1 \geq p_2$

X_i 为方式A下第*i*名网管负责的30台电脑中任意时刻需要帮助的人数, $X_i \sim B(30, 0.01)$

设 A_i 为方式A下第*i*名网管负责的30台电脑中有人得不到及时帮助, $i=1, 2, \dots, 10$

$$P(A_i) = P(X_i \geq 2)$$

$$= 1 - P(X_i = 0) - P(X_i = 1) = 0.0361$$

例2.12

注意 A_1, A_2, \dots, A_{10} 相互独立, 于是

$$\begin{aligned} p_1 &= P(A_1 \cup A_2 \cup \dots \cup A_{10}) \\ &= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_{10}) \\ &= 1 - P(\bar{A}_1) \times P(\bar{A}_2) \times \dots \times P(\bar{A}_{10}) = 0.3077 \end{aligned}$$

例2.12

Y 为方式B下300台电脑中任一时刻需要帮助的人数, $Y \sim B(300, 0.01)$

由于 $np=3$,近似地有 $Y \sim P(3)$ 于是,查泊松分布表,有

$$p_2 = P(Y > 8) = P(Y \geq 9) = 0.0038$$

(2)设 N 为使得

$$P(Y > N) = P(Y \geq N + 1) < 0.02$$

的最小的 N ,查泊松分布表,得 $N+1=8$