概率论与数理统计

第八讲

离散型随机变量的分布律

1. 定义与性质

定义

一个随机变量如果取至多可列个值（即有限或可列），则称这个随机变量是离散（discrete）型的.

- 实际问题中, 许多现象可用离散型随机变量描述, 观察下列事例

a) 掷一枚硬币三次, 记X为出现正面的次数;

b) 抽查一个人的血型, 以X表示他可输血的血型数;

c) 重复独立做某个试验, 设成功率为一固定值, 记X为首次成功时的试验次数;

d) X 为某网站在一段时间内被点击的次数;

e) X 为某产品的使用寿命;

上述事例, a), b)中的X取有限个值, c), d)中的 X 取可数个值, 这些 X 均为离散型随机变量; e)中的 X 则连续取值, 不属离散型随机变量.

如何研究离散型随机变量的分布？

注意：1)任何一个形如$\{X \le x\}$的事件只涉及至多可列个取值，故可表成形如$\{X = x_k\}$的至多可列和；2)不同取值的事件$\{X = x_k\}$是互斥的.

只须研究：$P\{X=x_k\}=?$

定义

设离散型随机变量取值为 $X=x_k$ ，$(k=1,2,\cdots)$，则称 $P\{X=x_k\}\triangleq P_k$ 为 X 的**分布律**（distribution law）.

第8讲 离散型随机变量的分布律

分布律的表示方法

解析法 $P\{X = x_k\} = p_k \quad (k = 1, 2, \cdots)$

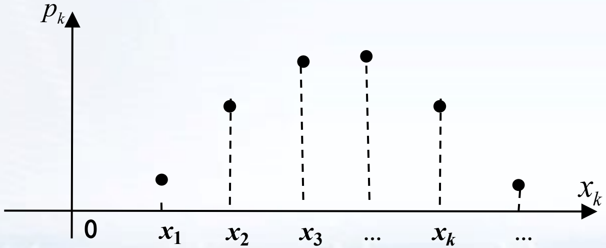
列表法

X	x_1	x_2	$\cdots$	x_k	$\cdots$
p_k	p_1	p_2	$\cdots$	p_k	$\cdots$

矩阵法

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p_1 & p_2 & \cdots & p_k & \cdots \end{pmatrix}$$

图示法



例 (掷硬币问题) 将一枚硬币连抛三次，观察正、反面出现的情况，记 X为正面出现的次数,求 X 的分布律

解：X 的取值为 $0,1,2,3$,其样本空间为

$$S=\{TTT,TTH,THT,HTT,THH,HTH,HHT,HHH\}$$

故X的分布律为 $P\{X=0\}=\frac{1}{8}$, $P\{X=1\}=\frac{3}{8}$,

$$P\{X=2\}=\frac{3}{8},\ P\{X=3\}=\frac{1}{8}.$$

- **问** 分布律有什么特点?

显然分布律满足：

1. $p_k \geq 0, \quad k = 1, 2, \cdots$
2. $\sum_{k=1}^{\infty} p_k = 1.$

提示：

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} P\{X = x_k\} = P\{\bigcup_{k=1}^{\infty}\{X = x_k\}\} = P\{X\text{的所有取值}\}$$

常用上式来验证所给出的随机变量的分布律是否正确.

例 (输血问题, 续)随机抽取一人观察血型, X表示他可输血的血型数. 据有关资料, 中国人O, A, B, AB血型的人分别占41%, 28%, 24%, 7%.

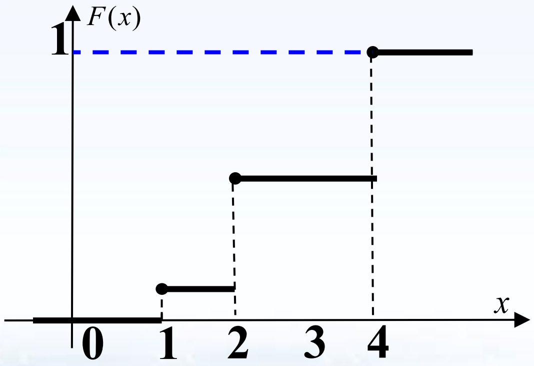
a) 写出 X 的分布律;

b) 写出其分布函数, 并作图.

解： a)

X	1	2	4
p_k	0.07	$0.28+0.24$	0.41

b) $$F(x)=\begin{cases}0, & x<1\\ 0.07, & 1\le x<2\\ 0.59, & 2\le x<4\\ 1, & x\ge 4\end{cases}$$



问 分布函数有什么特点?

第8讲 离散型随机变量的分布律

- 离散型随机变量分布函数的特征：
  阶梯函数，阶梯数为有限或可列个.
  设 X 的分布律为 $P\{X = x_k\} = p_k \quad (k = 1, 2, \cdots)$
  则

$$F(x) = \begin{cases} 0, & x < x_1 \\ p_1, & x_1 \le x < x_2 \\ p_1 + p_2, & x_2 \le x < x_3 \\ \quad \cdots & \\ \sum_{j=1}^{k-1} p_j, & x_{k-1} \le x < x_k \\ \quad \cdots & \end{cases}$$

- **注** 分布函数又称累积(cumulative)分布函数.

注 (掷两颗骰子)掷两颗均匀的骰子，观测其点数，令 X 为两骰子点数之和，求

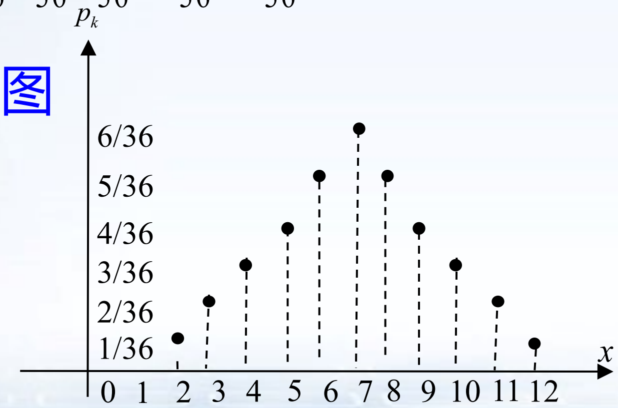
a) X 的分布律，并作出分布律图；

b) X 为奇数的概率.

答：a)

X	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

分布律图



b) $\frac{18}{36}=\frac{1}{2}$

注 **(首次成功)**重复独立地做某个试验. 设成功率 p 为一定值. 记 X 为首次试验成功时试验次数. 求 X 的分布律, 并验证它确实为一分布律.

解：

X	1	2	$\cdots$	k	$\cdots$
p_k	p	$p(1-p)$	$\cdots$	$p(1-p)^{k-1}$	$\cdots$

注意到 $\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} p(1-p)^{k-1} = \dfrac{p}{1-(1-p)} = 1.$

这说明只要 $p>0$, 试验下去总会成功.

X 的这种分布由此等比数列(几何级数)表达, 故称为参数为p的几何分布(**geometric distribution**).

例 (摸球问题)设袋中有r个红球及b个蓝球, 从中摸出n个 $(n \le r+b)$ 球, 记 X 为其中红球的数量. 求X的分布律.

分析 其分布律适用古典概率模型n个球总的摸法有

$\binom{r+b}{n}$ 种, 有利场合为 $\binom{r}{x}\binom{b}{n-x}$.

故分布律为

$$P\{X=x\} = \frac{\binom{r}{x}\binom{b}{n-x}}{\binom{r+b}{n}}, \quad \max\{0, n-b\} \le x \le \min\{r, n\}$$

- **注** 上述 X 的分布称为参数为 $m(=r+b)$, r, n的超几何分布.

2 X 可用于无放回抽样时, 抽出的某种事物数(如红球数, 次品数…).

3 顺便证明了一个数学恒等式

$$\sum_{k=0}^{r}\frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}}=1$$

4 还有更多常见的分布将在下一讲专门介绍.