河海大学

Ch2　离散型随机变量



• 随机变量的概念

定义 设随机试验 E 的样本空间是 Ω，若对每个 $\omega\in\Omega$，有定义在 Ω 上的一个实数 $X(\omega)$ 与之对应，称这样一个定义在 Ω 上的单值实函数 $X=X(\omega)$ 为随机变量(*Random Variable*)，简记为 *r.v.* X。随机变量一般用英文大写字母 X、Y、Z 等表示，也可用希腊字母 ξ、η、ζ 等表示。

$$\therefore\ r.v.\ X:\Omega\to R$$

• 一维离散型随机变量的分布律

定义　全部可能取值为有限个或可列无限个的随机变量为**离散型随机变量**。

即全部可能取值至多为可列无限个的随机变量为**离散型随机变量**。

若X为离散型随机变量，其取值为$x_1, x_2, \cdots, x_n, \cdots$，

X取每个可能值的概率为$P\{\omega \mid X(\omega) = x_k\}, k = 1, 2, \cdots$

记为 $p_k \overset{\Delta}{=} P\{X = x_k\}, k = 1, 2, \cdots$

称 $p_k = P\{X = x_k\}$，$k=1,2,\cdots$为$r.v.X$ 的**分布律**或**分布列**或**概率分布**。

$r.v.\ X$的分布律 $p_k = P\{X = x_k\}, k = 1, 2, \cdots$

也可表为

$$X \sim P\{X = x_k\} = p_k,\ (k = 1, 2, \ldots),$$

或

$$X \sim \begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_n & \cdots \\ \hline P & p_1 & p_2 & \cdots & p_n & \cdots \end{array}$$ 概率分布表

分布律的性质

(1)(非负性) $p_k \geq 0, k = 1, 2, \cdots;$

(2)(规范性) $\sum_{k=1}^{\infty} p_k = 1.$

例 设袋中有5只球，编号为1、2、3、4、5，在袋中同时取3只球，以X表示取出的3只球中的最大号码。试写出X的分布律。

例 设X的分布律为：

$$P(X=k)= a\, e^{-k+2},\quad k=0,1,2,\cdots$$

求常数a。

• 几个常见的离散型分布

1. 退化分布(单点分布)

$X \sim P\{X=a\}=1$，其中a为常数。

即 $X \sim \begin{array}{c|c} X & a \\ \hline P & 1 \end{array}$

2. (0—1)分布(两点分布)

$$X \sim \begin{array}{c|cc} X & 0 & 1 \\ \hline P & 1-p & p \end{array}$$

或 $X \sim P\{X=k\}=p^k(1-p)^{1-k}$，$(0< p <1)$ $k=0$，1

3. 几何分布

一次试验中只考虑某事件A出现或不出现，设$P(A)=p$， $P(\bar{A})=1-p$。现重复独立地做试验，一旦A发生就立即停止试验。

以X表示A首次发生所需的试验次数，则其分布率为：

$$X \sim P\{X=k\}=(1-p)^{k-1}p,\ (0<p<1)\quad k=1,\ 2,\ \ldots$$

称 X 服从参数为 p的**几何分布**，记为$X\sim G(p)$。

4. 二项分布$B(n, p)$

以 X 记 n重贝努里试验中 A发生的次数，则其分布律为：

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0,1,2,\cdots,n$$

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

例 一大楼装有5个同类型的供水设备，调查表明在任一时刻 t 每个设备被使用的概率为0.01。问在同一时刻：

（1）恰有两个设备被使用的概率是多少；

（2）至多有三个设备被使用的概率是多少？

5. 泊松(*Poisson*)分布 $P(\lambda)$

若随机变量X的所有取值为一切非负整数，且其分布律为：

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

其中$\lambda>0$为常数，称 X 服从参数为λ的**泊松(*Poisson*)分布**，记为$X\sim P(\lambda)$。

若$X(t)$表示在时间区间$[0, t]$中某服务台到达的顾客数。若$X(t)$满足：

（1）在不重叠的时间区间内到达的顾客数相互独立（无后效性）；

（2）在时间区间$[a, a+t]$内到达的顾客数只与时间长度有关，而与区间起点a无关（平稳性）；

（3）当t充分小时，在区间（$a, a+t$）内到达两个或两个以上的顾客不可能（普通性）；

（4）在有限区间中只到达有限个顾客且不可能始终没有顾客到达（非平凡性）。

定理

在上述条件下，在长度为 t 的时间区间上到达的顾客数 $X(t)$服从参数为λt的*Poisson*分布，其中$\lambda > 0$是一个常数。

6. 负二项分布

以 X 记可列重贝努里试验中A恰好发生 r次所需的试验次数，则其分布律为：

$$P(X=k)=C_{k-1}^{r-1}(1-p)^{k-r}p^r,\ \ k=r,r+1,r+2,\cdots$$

称X服从参数为(r, p)的**负二项分布**，记为

$X \sim NB(r, p)$

负二项分布又叫**帕斯卡(Pascal)**分布

7. 超几何分布

设N个元素分为两类，其中M个属于第一类，$N-M$个属于第二类。现从中按不重复抽样取n个，以X记这n个中属于第一类元素的个数。则X的分布律为：

$$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\quad k=0,1,2,\cdots,\min(n,M)$$

称X服从参数为(N，M，n)的**超几何分布**。

- 常见分布律之间的关系

1. (0－1)分布和二项分布的关系

(0－1)分布是二项分布$B(n, p)$中$n＝1$时的特款。

2. 几何分布和负二项分布的关系

几何分布是负二项分布$NB(r, p)$中$r＝1$时的特款。

3. 超几何分布和二项分布的关系

定理 设在超几何分布中，n是一个取定的正整数，

而 $\lim\limits_{N\to\infty}\dfrac{M}{N}=p$，则

$$\lim_{N\to\infty}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=C_N^k p^k(1-p)^{n-k},$$

$$k=0,1,2,\dots,n$$

即 当N充分大时，超几何分布趋向于二项分布。

事实上： 超几何分布用来描述不放回抽样的情况；

而二项分布则用来描述放回抽样的情况；

当N充分大时，两种抽样方式的差别很小。

4. 二项分布和泊松分布的关系

定理 设随机变量 $X_n \sim B(n, p_n)$, $(n=0, 1, 2, \cdots)$, 且

$\lim\limits_{n\to\infty} np_n = \lambda > 0$ ，λ为常数，则

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

$k = 0, 1, 2, \cdots$

该定理也称为泊松定理。

泊松定理表明，泊松分布是二项分布的极限分布，当n很大，p很小时，二项分布就可近似地看成是泊松分布，即

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

其中$\lambda = np$.

一般的，当$n \geq 10$，$p \leq 0.1$时就**可用泊松分布近似代替二项分布。**

例 某人射击的命中率为0.02，他独立射击400次，试求其命中次数不少于2的概率。

解 设X表示400次独立射击中命中的次数，则$X \sim B(400, 0.02)$，故

$$P\{X \geq 2\} = 1 - P\{X=0\} - P\{X=1\}$$

$$= 1 - 0.98^{400} - (400)(0.02)(0.98^{399}) = 0.997165$$

另解(用泊松分布) 由于$\lambda \approx np = (400)(0.02) = 8$，

故 $X \sim P\{X=k\} = \dfrac{8^k}{k!}e^{-8}, k \geq 0$ 近似地有

$$P\{X \geq 2\} = 1 - P\{X=0\} - P\{X=1\}$$

$$= 1 - e^{-8} - 8e^{-8} = 0.996981$$

二维离散型随机变量

1. 联合分布律

若二维随机变量(X, Y)只能取至多可列个值(x_i, y_j), $(i, j=1, 2,\cdots)$，则称(X, Y)为**二维离散型随机变量**。

若二维离散型随机变量(X, Y) 取 (x_i, y_j)的概率为p_{ij}, 即 $P\{X=x_i, Y=y_j\}=p_{ij}$，$(i, j=1, 2,\cdots)$ 则称 p_{ij} 为二维离散型随机变量(X, Y)的**分布律**, 或随机变量(X, Y)的**联合分布律。**

可记为

$$(X, Y)\sim P\{X=x_i, Y=y_j\}=p_{ij}，(i, j=1, 2,\cdots)$$

$P\{X=x_i, Y=y_j\}=p_{ij}$，$(i, j=1, 2, \cdots)$

联合分布律的性质

(1) (非负性) $p_{ij} \geq 0$，$i, j=1, 2, \cdots$；

(2)(规范性) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij}=1$

二维离散型随机变量的联合分布律也可列表表示如下：

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

例 盒子里装有3只黑球，2只红球，2只白球，今在其中任取4只球，以 X 表示取到黑球的数目，以 Y 表示取到红球的数目。试写出 X 和 Y 的联合分布律；求 $P\{X \le 1，Y \ge 2\}$。

- **边缘分布律**

若随机变量 X 与 Y 的联合分布律为

$$(X, Y) \sim P\{X = x_i, Y = y_j\} = p_{ij}, \quad (i, j = 1, 2, \cdots)$$

则称

$$P\{X = x_i\} = p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij}, \quad i = 1, 2, \cdots$$

为(X, Y)关于X的边缘分布律；

同理 $$P\{Y = y_j\} = p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij}, \quad j = 1, 2, \cdots$$

称为(X, Y)关于Y的边缘分布律。

$$P\{X=x_i\}=p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij},\quad i=1,2,\cdots$$

$$P\{Y=y_j\}=p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij},\quad j=1,2,\cdots$$

边缘分布律自然也满足分布律的性质：

(1) $p_{i\cdot}\ge 0;\ (p_{\cdot j}\ge 0)$

(2) $\sum_{i=1}^{\infty}p_{i\cdot}=1.\ (\sum_{j=1}^{\infty}p_{\cdot j}=1)$

二维离散型随机变量的边缘分布律也可列表表示如下：

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

$$p_{i\cdot} = \sum_{j=1}^{\infty} p_{ij}, \quad p_{\cdot j} = \sum_{i=1}^{\infty} p_{ij}$$

例 设(X, Y)的联合分布律如右表。试求

（1）X和Y的边缘分布律；

（2）$P\{X \le 1\}$和$P\{Y \ge 1\}$。

$X \backslash Y$	-1	1	2
0	$\frac{1}{12}$	0	$\frac{3}{12}$
$\frac{3}{2}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{3}{12}$	$\frac{1}{12}$	0

$X \backslash Y$	-1	1	2	$p_{i\cdot}$
0	$1/12$	0	$3/12$	$1/3$
$\frac{3}{2}$	$2/12$	$1/12$	$1/12$	$1/3$
2	$3/12$	$1/12$	0	$1/3$
$p_{\cdot j}$	$1/2$	$1/6$	$1/3$	1

（2）$P\{X \le 1\}$和$P\{Y \ge 1\}$。

• 条件分布律

设随机变量X与Y的联合分布律为

$$(X, Y) \sim P\{X=x_i, Y=y_j\}=p_{ij}, \quad (i, j=1, 2, \cdots)$$

X和Y的边缘分布律分别为

$$P\{X=x_i\}=p_{i\cdot}, \quad i=1, 2, \cdots$$

和 $P\{Y=y_j\}=p_{\cdot j}, \quad j=1, 2, \cdots$

若对固定的 j, $p_{\cdot j}>0$, 则称

$$P\{X=x_i \mid Y=y_j\}=\frac{P\{X=x_i, Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},$$

$$i=1, 2, \cdots$$

为$Y=y_j$的条件下，X的条件分布律。

记为 $p_{i|j}=P\{X=x_i \mid Y=y_j\}=\dfrac{p_{ij}}{p_{\cdot j}}$.

同理，若对固定的 i，$p_{i\cdot}>0$, 则称

$$P_{j|i}=P\{Y=y_j \mid X=x_i\}=\frac{p_{ij}}{p_{i\cdot}},\ j=1,2,\cdots$$

为$X=x_i$的条件下，Y的条件分布律。

条件分布律也满足分布律的性质。

例　一射手进行射击，命中目标的概率为p $(0< p <1)$，射击进行到命中目标两次为止，现用 X 表示首次命中目标所进行的射击次数，用Y 表示总共进行的射击次数。试求 X 和Y 的联合分布律及条件分布律。

- **离散型随机变量的相互独立性**

设随机变量X与Y的联合分布律为

$$(X, Y) \sim P\{X=x_i, Y=y_j\}=p_{ij}, \quad (i, j=1, 2, \cdots)$$

若对**任意的** i、j，有 $p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$，

即 $P\{X=x_i, Y=y_j\}=P\{X=x_i\}\cdot P\{Y=y_j\}$

则称随机变量**X与Y相互独立**。

例 设随机变量(X, Y)的分布律为

$X \backslash Y$	1	2
1	$\frac{1}{8}$	b
2	a	$\frac{1}{4}$
3	$\frac{1}{24}$	$\frac{1}{8}$

且 X与Y相互独立，试求a、b的值。

上述独立的概念不难推广到n维离散型随机变量的情形。

设X_1，X_2，…，X_n为一个n维离散型随机变量，若对任意的x_1，x_2，…，x_n有:

$$P\{X_1=x_1, X_2=x_2, \cdots, X_n=x_n\}$$

$$=P\{X_1=x_1\}\cdot P\{X_2=x_2\}\cdots P\{X_n=x_n\}$$

则称随机变量X_1，X_2，…，X_n**相互独立**。

以 X 记 n 重贝努里试验中A发生的次数，则

$X\sim B(n, p)$

若记 $X_i=\begin{cases}1, & \text{若在第 } i \text{ 次试验中 } A \text{ 发生;}\\ 0, & \text{若在第 } i \text{ 次试验中 } A \text{ 不发生。}\end{cases}$

$i=1,2,\cdots,n$

$X_i \sim \begin{array}{c|cc} X & 0 & 1 \\ \hline P & 1-p & p \end{array}$ 且 $X_1, X_2, \cdots, X_n$ 相互独立

于是有：$X = X_1 + X_2 + \cdots + X_n$

若r.v. $X_1, X_2, \cdots, X_n$相互独立且服从同一(0–1)分布，则 $\sum_{i=1}^{n} X_i \sim B(n, p)$。

• 离散型随机变量函数的分布律

1. 一维离散型随机变量函数的分布律

定理 设X一个随机变量，若 $y=g(x)$是一元单值实函数，则$Y=g(X)$也是一个随机变量。

若 $X\sim P\{X=x_k\}=p_k$， $k=1, 2, \cdots$ 则

$$Y=g(X)\sim P\{Y=g(x_k)\}=p_k, \quad k=1, 2, \cdots$$

其中$g(x_k)$有相同的，其对应概率合并。

显然，Y的分布律也满足分布律的性质。

例 设 *r.v.* X的分布律为：

X	-2	-1	0	1	3
p	0. 2	0.1	0.3	0.3	0.1

求$Y=X^2$及 $Z=2X+3$的分布律。

2. 多维离散型随机变量函数的分布律

定理 设X_1，X_2，…，X_n一个 n 维随机变量，若$y=g(x_1, x_2, \cdots, x_n)$是一个 n 元实值函数，则$Y=g(X_1, X_2, \cdots, X_n)$也是一个随机变量。

写出$Y=g(X_1, X_2, \cdots, X_n)$的所有可能取值，然后求其取每一个值的概率。

例 设随机变量(X, Y)的分布律为

$X \backslash Y$	0	1	2	3	4
0	0.1	0.05	0.01	0.02	0.01
1	0.04	0.06	0.02	0.03	0.04
2	0.13	0.08	0.01	0.05	0.03
3	0.08	0.11	0.05	0.06	0.02

(1) 求$P\{X=2|Y=2\}$;

(2) 求$S=X+Y$的分布律;

(3) 求$T=2Y-X$的分布律;

(4) 求$V=\max(X, Y)$的分布律;

(5) 求$U=\min(X, Y)$的分布律。

例 设 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，且 X 与 Y 相互独立，求 $Z = X + Y$ 的分布律。

推论 设 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，且 X 与 Y 相互独立，则 $X + Y \sim P(\lambda_1 + \lambda_2)$。

推论 设 $X_i \sim P(\lambda_i)$，$i = 1, 2, \cdots, n$，且 $X_1, X_2, \cdots, X_n$ 相互独立，则 $\sum X_i \sim P(\sum \lambda_i)$。