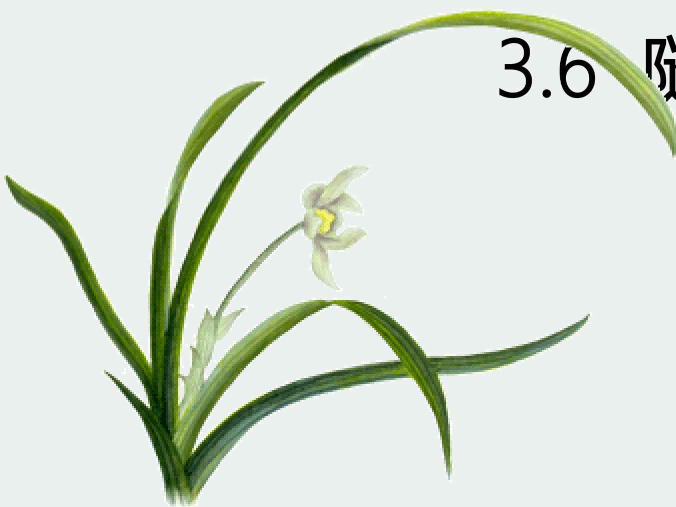
第3章 随机变量及其分布



在前面的学习中,我们用字母A、B、C、… 表示随机事件，并视之为样本空间Ω的子集，研究了随机事件的关系和运算；引进了随机事件的概率的概念和性质，并针对古典概型，主要研究了用加法原理、乘法原理及排列组合等手段计算事件的概率。

本节，我们将引进随机变量的概念，并用随机变量表示随机事件，以便采用高等数学的方法描述、研究随机现象。

3.1 随机变量的概念

- 随机变量的定义
- 随机变量的特征
- 随机变量的类型

3.1 随机变量的概念

引例 抛掷硬币两次，观察抛掷的结果.

$$\Omega = \{++, +-, -+, --\}$$ **抛掷 n 次？**

或者说 抛掷结果为：两次正面；一正一反；两次反面.

如果用 X 表示试验中正面出现的次数，则 X 的取值 0, 1, 2，此时，

✓ 数量化表示
✓ 刻画全面系统

“两次正面” = “X取到值2”，可记为$\{X=2\}$

“一正一反” = “X取到值1”， 记为$\{X=1\}$,

“两次反面” = “X取到值0”， 记为$\{X=0\}$

$$P(X=0)=\frac{1}{4}, P(X=1)=\frac{1}{2}, P(X=2)=\frac{1}{4}.$$

□ 随机变量的定义

定义 设随机试验的样本空间为Ω，如果对于每一个样本点 $\omega \in \Omega$，均有唯一的实数 $X(\omega)$ 与之对应，则称 $X = X(\omega)$ 为样本空间Ω上的随机变量。

例

- **某种个灯泡的使用寿命X**
  X 的可能取值为 [0,N]
- **某电话总机在一分钟内收到的呼叫次数Y.**
  Y 的可能取值为 0，1，2，3，...,
- **在[0，1]区间上随机取点的点的坐标Z.**
  Z 的可能取值为 [0，1]上的全体实数。

□ 随机变量的特征

1) 它是从样本空间 Ω 到实数集合上的映射;

2) 它的取值随试验结果而改变，具有随机性;

3）随机变量在某一范围内取值，表示一个随机事件.

如引例中，***X*=0，1，2，**事件

$$\{X \leq 1\} = \text{“最多出现一次正面”}$$

$$\{X \geq 1\} = \text{“至少出现一次正面”}$$

随机变量的类型

- **离散型**

  随机变量的所有取值是有限个或可列个.

如：某电话总机在一分钟内收到的呼叫次数Y.

- 非离散型

  随机变量的取值有无穷多个，且非可列个.

如：某种个灯泡的使用寿命X.

其中连续型随机变量是非离散型的一种重要类型

3.2 离散随机变量

- 离散随机变量及其分布列
- 分布列的性质
- 几种常见的离散分布

3.2 离散随机变量

离散随机变量及其分布列

设随机变量 X 的所有可能取值是

$$x_1, x_2, \cdots, x_n, \cdots$$

而取值 x_k 的概率为 p_k， 即

$$\boxed{P\left\{X = x_k\right\} = p_k, k = 1, 2, \cdots, n, \cdots}$$

称 X为***离散随机变量，*** **上**式为X的分布律（列）

或概率质量函数（Probability distribution）.

离散随机变量分布列的表示法

- 公式法 $P\left\{X=x_k\right\}=p_k, k=1,2,\cdots,n,\cdots$
- 表格法

X	$x_1,$	$x_2,$	$\cdots$	$x_k,$	$\cdots$
P	p_1 ,	p_2 ,	$\cdots$	p_K	$\cdots$

- 图像法

离散随机变量X的分布列全面表达了X的所有可能取值以及取各个值的概率情况.

□ 分布列的性质

$$P\{X = x_k\} = p_k, k = 1, 2, \cdots, n, \cdots$$

1) $p_k \geq 0 \qquad k = 1, 2, \cdots$

2) $\sum\limits_{k=1}^{\infty} p_k = 1$

反之，满足上述两条性质的数列$\{p_k\}$也可以作为某一离散随机变量X的分布列.

分布列确定概率

例1　设X的分布列为

X	-1	1	2
P	**1/3**	**1/2**	**1/6**

求 **P(0<*X*≤2)**

解　$P(0<X\leq 2)=P(X=1)+P(X=2)$

$=1/2+1/6=2/3$

例2 设有一批产品20件，其中有3件次品，从中任意抽取2件，如果用X表示取得的次品数，求随机变量X的分布律及事件“至少抽得一件次品”的概率。

解： X的可能取值为 0，1，2

$$P\{X=0\}=\frac{C_{17}^{2}}{C_{20}^{2}}=\frac{136}{190}=P(\text{两件全为正品})$$

$$P\{X=1\}=\frac{C_{3}^{1}C_{17}^{1}}{C_{20}^{2}}=\frac{51}{190}=P(\text{只有一件为次品})$$

$$P\{X=2\}=\frac{C_{3}^{2}}{C_{20}^{2}}=\frac{3}{190}=P(\text{两件全为次品})$$

故X的分布列为

X	0	1	2
p_k	$\frac{136}{190}$	$\frac{51}{190}$	$\frac{3}{190}$

而“至少抽得一件次品”$=\{X\geq 1\}$ $=\{X=1\}\cup\{X=2\}$

注意：$\{X=1\}$与$\{X=2\}$是互不相容的！

故 $P\{X\geq 1\}=P\{X=1\}+P\{X=2\}=\frac{51}{190}+\frac{3}{190}=\frac{54}{190}=\frac{27}{95}$

实际上，这仍是古典概型的计算题，只是表达事件的方式变了.

例3 设随机变量*X*的分布列为

$$P\{X=k\}=b\left(\frac{2}{3}\right)^k, k=1,2,3,\cdots$$

试确定常数***b***.

解 由分布列的性质,有

$$\sum_{k=1}^{\infty} P(X=k)=\sum_{k=1}^{\infty} b\left(\frac{2}{3}\right)^k=\frac{b\,\frac{2}{3}}{1-\frac{2}{3}}$$

$$=b\,\frac{2/3}{1/3}=2b=1 \Longrightarrow b=\frac{1}{2}.$$

□ 几种常见的离散分布

■ 0-1分布(贝努利分布、二点分布)

定义：若随机变量X只可能取0,1两个值，X的分布列为

X	0	1
P	$1-p$	p

则称X服从参数为 p $(0< p <1)$ 的 (0-1)分布或二点分布或称X为伯努利随机变量.

注：若试验E只有两个试验结果$A,\overline{A}$，则称E为伯努利(Bernoulli)试验，伯努利试验都可以用两点分布描述。

如 示性随机变量：$I_A(\omega)=\begin{cases}1, & \omega\in A\\ 0, & \omega\notin A\end{cases}$, $\quad p=P(A)$

■ 二项分布 *B*inomial distribution

在*n*重伯努利试验中,设每次试验中*A*发生的概率为 *p* (0< *p* <1)，若以*X*表示事件*A*发生的次数，则*X*可能的取值为0,1,2,3,…,*n*， *P*(*X*=*k*)=?

由乘法原理，随机变量*X*的分布列为

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$$

$$k=0,1,2\ldots,n;$$

称*X*服从参数为 *n*, *p* 的二项分布(也称Bernoulli 分布),记为 $X\sim B(n,p)$.

例4 考生凭猜测答四选一的选择题，问在猜答3道这样的题目中，

(1) 猜对2道题的概率有多大?

(2) 至少猜对2道题的概率有多大?

解 猜答3道题可看做3重伯努利试验，设猜对的题目数为X，则X是参数为 $n=3, p=\frac{1}{4}$, 的二项分布：

$$X \sim B(3, \frac{1}{4})$$

$$(1) P(X=2)=C_3^2 p^2(1-p) \approx 0.14$$

$$(2) P(X \geq 2)=P(X=3)+P(X=3)$$

$$=C_3^2 p^2(1-p)+C_3^3 p^3 \approx 0.16$$

例5 设有N件产品，其中有M件次品，现进行n次有放回的抽样，每次抽取一件，求这n次中共抽到的次品数X的概率分布。

解 由于抽样是有放回的，因此这是n重伯努利试验，若以A表示一次抽样中抽到次品这一事件，则

$$p = P(A) = \frac{M}{N},$$

故 $X \sim B(n, \frac{M}{N})$，即

$$P(X = k) = C_n^k (\frac{M}{N})^k (1 - \frac{M}{N})^{n-k}, (k = 0, 1, 2, \cdots, n)$$

二项分布的最可能取值

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

因为 $\dfrac{P(X=k)}{P(X=k-1)}=\dfrac{C_n^k p^k q^{n-k}}{C_n^{k-1}p^{k-1}q^{n-k+1}}=\dfrac{(n-k+1)p}{kq}$

$$=1+\frac{(n+1)p-k}{kq}\begin{cases}>1,k<(n+1)p\\<1,k<(n+1)p\end{cases},(k=1,2,\cdots,n)$$

故，二项分布当 $k=(n+1)p$为正整数时 $P(X=k)$在 $k=(n+1)p$ 和 $k=(n+1)p-1$时都取得最大值；而当 $k=(n+1)p$不是整数时，$P(X=k)$在 $k_0=[(n+1)p]$ 时取得最大值.

■ 泊松分布 Poisson distribution

若随机变量 X 的分布律为:

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

其中 $\lambda>0$, 则称 X 服从参数为 λ 的泊松分布,

记为 $X\sim P(\lambda)$

$$\sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}=1$$

附表1 是泊松分布累积概率值表.

实际问题中如下X是服从或近似服从Poisson分布的.

- **服务台在某时间段内接待的服务次数;**
- **交换台在某时间段内接到呼叫的次数;**
- **矿井在某段时间发生事故的次数;**
- **显微镜下相同大小的方格内微生物的数目;**
- **单位体积空气中含有某种微粒的数目.**

体积相对小的物质在较大的空间内的稀疏分布，都可以看作泊松分布, 其参数λ可以由观测值的平均值求出。

例6 已知某电话交换台每分钟接到的呼唤次数X服从 $\lambda=4$ 的泊松分布，分别求(1)每分钟内恰好接到3次呼唤的概率,(2)每分钟不超过4次呼唤的概率.

解

$$\boxed{P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}}$$

$\lambda=4, k=3$

$$P(X=3)=\frac{4^3}{3!}e^{-4}=0.19563$$

$$P(X\le 4)=P(X=0)+P(X=1)+P(X=2)$$
$$+P(X=3)+P(X=4)=0.628838$$

■几何分布 geometric distribution

从一批次品率为 p 的产品中，有放回抽样直到抽到次品为止。求抽到次品时，已抽取次数X 的分布列.

解 X的所有可能取值为 1,2,3,… ,k,… $P(X=k)=?$

记 A_i="第i次取到正品"，$i=1,2,3,\ldots$

则A_i, $i=1,2,3,\ldots$是相互独立的！且$(X=k)$对应着事件

$$A_1A_2\cdots A_{k-1}\overline{A_k}$$

$$P(X=k)=P(A_1A_2\cdots A_{k-1}\overline{A_k})=(1-p)^{k-1}p,\ k=1,2,\ldots$$

定义 设每次独立试验成功的概率为 p, 其中 $0< p <1$, 若以 X 表示直到第一次成功所作的试验次数，则 X 可能的取值为 $1,2,3,\ldots,$

随机变量 X 的分布列为

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

称 X 服从参数为 p 的几何分布，记为

$$X \sim G(p)$$

$$\sum_{k=1}^{\infty} P\{X=k\}=\sum_{k=1}^{\infty}(1-p)^{k-1}p=1$$

几何分布的无记忆性

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

设 $X\sim G(p)$，n,m为任意两个正整数，则有

$$P\{X>n+m\mid X>n\}=P\{X>m\}$$

事实上，$P\{X>n+m\mid X>n\}=\dfrac{P\{X>n+m,X>n\}}{P\{X>n\}}$

$$=\frac{P\{X>n+m\}}{P\{X>n\}}=\frac{\sum\limits_{k=n+m+1}^{\infty}(1-p)^{k-1}p}{\sum\limits_{k=n+1}^{\infty}(1-p)^{k-1}p}=(1-p)^m$$

$$P\{X>m\}=\sum_{k=m+1}^{\infty}(1-p)^{k-1}p=(1-p)^m$$

例7 某血库急需*AB*型血，需从献血者中获得，根据经验，每100个献血者中只能获得2名身体合格的*AB*型血的人，今对献血者一个接一个进行化验，用*X*表示在第一次找到合格的*AB*型血时，献血者已被化验的人数，求*X*的概率分布。

解 由已知， $X \sim G(p)$，$p=2\%$，则

$$P\{X=k\}=(1-p)^{k-1}p$$

$$=0.98^{k}\times 0.02, \qquad k=1,2,\cdots$$

注意，由几何分布的无记忆知，

若已化验了*n*个人，没有获得合格的*AB*型血，则再化验*m*个人找不到合格的*AB*型血的概率与已知的信息（即 前*n*个人的不是*AB*型血）无关.

■ 负二项分布

负二项分布是几何分布的一般化。

设每次试验成功的概率为 p，连续独立试验直到成功 r 次为止，若以 X 表示试验次数，则 X 可能的取值为：$r, r+1, r+2, \ldots,$

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

$$k=r, r+1, r+2, \cdots$$

则称 X 服从参数为 p 的负二项分布(或帕斯卡分布).

记为 $X \sim NB(r, p)$

注：负二项随机变量可以表示为 r 个独立的几何随机变量之和.

■ 超几何分布 hyper geometric distribution

假设一批产品有N个，其中有 M 个次品，从中抽取n个，若以X表示抽到的次品数，则X的分布列为

$$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

$$k=0,1,2,\cdots,\min\{M,n\}$$

称X服从超几何分布.

注：二项分布可用来描述有放回抽样，而超几何分布可以用来描述不放回抽样. 在实际的质量控制中，由于人力物力或时间的关系，仅能有一小部分产品被抽检，或若抽检具有破坏性，不能放回。

定理 设超几何分布中，n是一个取定的正整数，而

$$\lim_{N\to\infty}\frac{M}{N}=p, 0<p<1.$$

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

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

由此可见，对于固定的$n\geq 1$,当N充分大时，有

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}\approx C_n^k\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}$$

$$k=0,1,2,\cdots,\min(M,n).$$

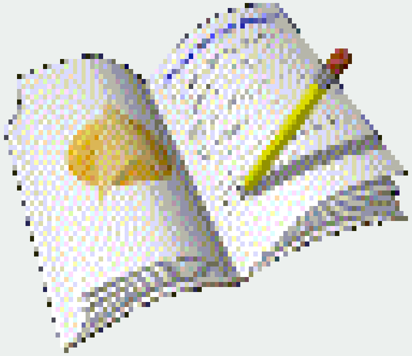
主观题 10分

设置

若某人做某事的成功率为1%，他重复努力400次，求至少成功一次的概率.

正常使用主观题需2.0以上版本雨课堂

作答



课后作业

习题3

1，2，3，4，5，6，7.