# 第 2 章　一维随机变量及其分布

## 本章学习目标

- 掌握随机变量及其概率分布的概念
- 掌握随机变量分布函数的概念及性质，会计算与随机变量有关事件的概率
- 掌握离散型随机变量及其概率分布的概念，掌握 $(0-1)$ 分布、二项分布、泊松分布及其应用
- 掌握连续型随机变量及其概率密度的概念，掌握概率密度与分布函数之间的关系
- 掌握正态分布、均匀分布和指数分布及其应用

上一章我们用样本空间的子集来表示试验的各种结果，对全面讨论随机现象的统计规律性有很大的局限性；为了深入研究和全面掌握随机现象的统计规律性，我们将随机试验的结果与实数对应起来，即将随机试验的结果数量化，为此引入随机变量的概念，用它描述随机现象是近代概率论中最重要的方法，它使概率论从事件及其概率的研究扩大到随机变量及其概率分布的研究，这样就可以应用微积分等数学工具，使概率论成为一门真正的数学学科.

## 2.1　一维随机变量的概念及其分布函数

### 2.1.1　随机变量的概念

在研究随机试验结果的规律性时，我们发现，大多数试验的结果可以直接用一个数来表示，而有些随机试验其可能结果并不是数，但只要将每一个可能结果与一个实数相对应，那么，试验的不同可能结果就可以用一个变量来表示了.

下面给出随机变量的定义.

**定义 1**　设 $E$ 是随机试验，$\Omega$ 是其样本空间. 如果对于每一个样本点 $\omega \in \Omega$，都有一个确定的实数 $X(\omega)$ 与之对应，则称 $\Omega$ 上的实值函数 $X(\omega)$ 为随机变量，简记为 $X$.

通常用大写英文字母 $X, Y, Z$ 或希腊字母 $\xi, \eta, \zeta$ 等表示随机变量，用相应的小写英文字母 $x, y, z$ 表示随机变量的取值.

引入随机变量后，就可以用含随机变量 $X$ 的等式或不等式来表示随机事件. 例如，$\{X=2\}$，$\{X<5\}$，$\{0 \leqslant X < 1\}$ 等，因此引入随机变量概念后，可把对事件的研

究转化为对随机变量的研究，从而可用微积分等数学工具来深入研究和处理．随机变量是我们今后主要的研究对象，但要注意这种变量与微积分中的变量是有区别的，主要有以下三个特点：

（1）随机变量 $X$ 是定义在样本空间上的实单值函数 $X(\omega)$，即 $X(\omega)$ 的定义域是样本空间，而微积分中的函数 $y=f(x)$ 的定义域是实数集合；

（2）取值的随机性，即 $X$ 取何值，在试验前是无法确定的；

（3）取值的统计规律性，即 $X$ 取某值的概率是确定的．

从随机试验可能出现的结果来看可分为两大类：

（1）离散型随机变量：主要特征是它所有可能取值为有限个或无限可列（像自然数一样可依次一一列出）个；

（2）非离散型随机变量：它包含的范围很广，情况比较复杂，本书中只研究其中最重要也是最常见的连续型随机变量，其主要特征是它所有可能取值充满某个区间 $(a,b)$（其中 $a$ 可以为 $-\infty$，$b$ 可以为 $+\infty$）．

### 2.1.2　随机变量的分布函数及其基本性质

1．分布函数的定义

**定义 2**　设 $X$ 是一个随机变量，$x$ 为任意实数，则称函数

$$F(x)=P\{X\leqslant x\},\ x\in(-\infty,+\infty)$$

为 $X$ 的分布函数．

分布函数 $F(x)$ 是定义在全体实数上的一个普通实函数，同时也具有明确的概率意义：若将 $X$ 看成是数轴上的随机点的坐标，则分布函数 $F(x)$ 在 $x$ 处的函数值就是 $X$ 落在区间 $(-\infty,x]$ 上的概率，因而可以用高等数学的方法来研究随机变量．

由此定义，若已知随机变量 $X$ 的分布函数 $F(x)$，则 $X$ 落入任一区间 $(x_1,x_2]$ 的概率等于 $F(x)$ 在此区间上的增量，即

$$\begin{aligned}P\{x_1<X\leqslant x_2\}&=P\{X\leqslant x_2\}-P\{X\leqslant x_1\}\\&=F(x_2)-F(x_1)\ .\end{aligned}$$

因此，知道了随机变量的分布函数也就掌握了该随机变量的统计规律性．

2．分布函数的性质

设 $F(x)$ 为随机变量 $X$ 的分布函数，则 $F(x)$ 具有下列性质：

（1）单调不减性：若 $x_1<x_2$，则 $F(x_1)\leqslant F(x_2)$；

（2）归一性：对任意实数 $x$，$0\leqslant F(x)\leqslant 1$，且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=\lim_{x\to-\infty}P\{X\leqslant x\}=P\{X\leqslant-\infty\}=0\ ,$$

$$F(+\infty)=\lim_{x\to+\infty}F(x)=\lim_{x\to+\infty}P\{X\leqslant x\}=P\{X\leqslant+\infty\}=1\ ;$$

（3）右连续性：即 $F(x+0)=F(x)$，若 $X$ 为连续型随机变量，则 $F(x)$ 处处连续．

具有以上三个性质的实函数，必是某个随机变量的分布函数，故这三个性质也是分布函数的充分必要条件.

由于 $F(x)$ 是随机变量取 $X$ 取小于等于 $x$ 的诸值 $x_k$ 的概率之和，故又称 $F(x)$ 为累积概率函数.

**例 1** 掷一枚均匀硬币 3 次，以 $X$ 表示出现正面的次数，求 $X$ 的分布函数.

**解** 由题意知 $X$ 的可能取值为 $0,1,2,3$，$X=0$ 表示三次都是反面，所以

$$P\{X=0\}=\left(\frac{1}{2}\right)^3=\frac{1}{8},$$

$X=1$ 表示 3 次中恰有一次为正面，所以

$$P\{X=1\}=C_3^1\times\frac{1}{2}\times\left(\frac{1}{2}\right)^2=\frac{3}{8},$$

同理可得

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

当 $x<0$ 时，

$$F(x)=P\{X\leqslant x\}=P(\varnothing)=0,$$

当 $0\leqslant x<1$ 时，

$$F(x)=P\{X\leqslant x\}=P\{X=0\}=\frac{1}{8},$$

当 $1\leqslant x<2$ 时，

$$F(x)=P\{X\leqslant x\}=P\{X=0\}+P\{X=1\}=\frac{1}{8}+\frac{3}{8}=\frac{1}{2},$$

当 $2\leqslant x<3$ 时，

$$F(x)=P\{X\leqslant x\}=1-P\{X=3\}=1-\frac{1}{8}=\frac{7}{8},$$

显然，当 $x\geqslant 3$ 时，

$$F(x)=P\{X\leqslant x\}=1.$$

所以 $X$ 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ 1/8, & 0\leqslant x<1,\\ 1/2, & 1\leqslant x<2,\\ 7/8, & 2\leqslant x<3,\\ 1, & x\geqslant 3.\end{cases}$$

**例 2** 设随机变量 $X$ 的分布函数为

$$F(x)=\begin{cases}A+Be^{-x}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

（1）试确定系数 $A$、$B$；（2）求 $P\{-1<X\leqslant 1\}$ .

**解**　（1）根据 $F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$ ，得

$$\lim_{x\to+\infty}(A+B\mathrm{e}^{-x})=A=1,$$

又由 $F(x)$ 在 $x=0$ 处右连续，有

$$F(0+0)=F(0)=0,$$

即

$$A+B=0,$$

则

$$B=-A=-1;$$

（2）由 $F(x)=\begin{cases}1-\mathrm{e}^{-x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$ 得

$$P\{-1<X\leqslant 1\}=F(1)-F(-1)=1-\mathrm{e}^{-1}-0=1-\mathrm{e}^{-1}.$$

## 2.2　离散型随机变量及其概率分布

### 2.2.1　离散型随机变量及其概率分布

1．概率分布

如前面所述，只能取有限个或无穷可列个值的随机变量 $X$ 称为**离散型随机变量**．掌握一个离散型随机变量 $X$ 的统计规律，只要知道 $X$ 的所有可能取值以及 $X$ 取每一个可能值的概率是多少，对这两点的全面描述被称作离散型随机变量的概率分布．

**定义 3**　设离散型随机变量 $X$ 的全部可能取值为 $x_k\ (k=1,2,\cdots)$，且取 $x_k$ 的概率为 $p_k$，即

$$P\{X=x_k\}=p_k\quad(k=1,2,\cdots).\tag{2.1}$$

称（2.1）式为离散型随机变量 $X$ 的概率分布或分布列，简称分布．

分布列也可以用表格的形式表示，即：

| $X$ | $x_1$ | $x_2$ | $\cdots$ | $x_k$ | $\cdots$ |
|---|---|---|---|---|---|
| $P\{X=x_k\}$ | $p_1$ | $p_2$ | $\cdots$ | $p_k$ | $\cdots$ |

**例 3**　袋中有 5 个球，其中 2 个黑球，3 个白球，从中随机地一次抽取 3 个球，求取得黑球数的概率分布．

**解**　令 $X$ 表示“取得的黑球数”，则 $X$ 的可能取值为 $0,1,2$ ，可以求得 $X$ 的分布列为

$$P\{X=0\}=\frac{C_3^3}{C_5^3}=\frac{1}{10},$$

$$P\{X=1\}=\frac{C_3^2C_2^1}{C_5^3}=\frac{6}{10},$$

$$P\{X=2\}=\frac{C_3^1C_2^2}{C_5^3}=\frac{3}{10}.$$

$X$的分布列的表格形式为：

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $p_k$ | $\frac{1}{10}$ | $\frac{6}{10}$ | $\frac{3}{10}$ |

2. 分布列的性质

由概率的基本性质可知，分布列满足如下两条性质：

（1）$p_k \geqslant 0\ (k=1,2,\cdots)$；

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

反之，满足上述两个性质的数列必可作为某个离散型随机变量的分布列.

**例 4** 若随机变量$X$的分布列如下：

| $X$ | $-2$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|---|
| $p_k$ | $a$ | $3a$ | $\frac{1}{8}$ | $a$ | $2a$ |

求：（1）常数 $a$；（2）写出 $X$ 的分布列；（3）$P\{X<1\}$；$P\{-2<X\leqslant 0\}$和$P\{X\geqslant 2\}$.

**解** 根据分布列的性质，必有$a>0$，且$\sum\limits_{i=1}^{5}p_i=1$，所以

$$a+3a+\frac{1}{8}+a+2a=1,$$

解得

$$a=\frac{1}{8},$$

故$X$的分布列为：

| $X$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| $p_k$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{4}$ |

$$P\{X<1\}=P\{X=-2\}+P\{X=-1\}+P\{X=0\}=\frac{5}{8},$$

$$P\{-2<X\leqslant 0\}=P\{X=-1\}+P\{X=0\}=\frac{1}{2},$$

$$P\{X\geqslant 2\}=P\{X=2\}=\frac{1}{4}.$$

若离散型随机变量$X$的分布列为

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

则

$$F(x)=P\{X\leqslant x\}=\sum_{x_k\leqslant x}p_k=\sum_{x_k\leqslant x}P\{X=x_k\}.$$

**注：**用离散型随机变量 $X$ 的分布函数求概率时，要考虑所取区间的端点是否包含在内.

**例 5** 已知离散型随机变量$X$的分布函数为

$$F(x)=\begin{cases}0, & x<3,\\ 1/10, & 3\leqslant x<4,\\ 2/5, & 4\leqslant x<5,\\ 1, & x\geqslant 5.\end{cases}$$

（1）求$X$的分布列；（2）求$P\{3.5<X\leqslant 4.5\}$.

**解** （1）由题意可知，$X$的可能取值为$3,4,5$，故有

$$P\{X=3\}=F(3)-F(3-0)=\frac{1}{10},$$

$$P\{X=4\}=F(4)-F(4-0)=\frac{2}{5}-\frac{1}{10}=\frac{3}{10},$$

$$P\{X=5\}=F(5)-F(4)=1-\frac{2}{5}=\frac{3}{5}.$$

则$X$的分布列为：

| $X$ | $3$ | $4$ | $5$ |
|---|---|---|---|
| $p_k$ | $\frac{1}{10}$ | $\frac{3}{10}$ | $\frac{3}{5}$ |

（2）$P\{3.5<X\leqslant 4.5\}=F(4.5)-F(3.5)=\frac{2}{5}-\frac{1}{10}=\frac{3}{10}$.

由此例可知，离散型随机变量 $X$ 的分布列和分布函数是两个等价的工具，但

使用分布列要方便得多.

### 2.2.2 几种常见的离散型随机变量的分布

1. 两点 $(0-1)$ 分布

**定义 4** 若随机变量 $X$ 的分布列为

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

其概率分布表为：

| $X$ | 0 | 1 |
|---|---|---|
| $p_k$ | $1-p$ | $p$ |

则称 $X$ 服从参数为 $p$ 的两点 $(0-1)$ 分布.

如果一个随机试验的样本空间只包含两个样本点，如抛硬币出现正面或反面；检查一产品的质量合格或不合格；登记一新生婴儿的性别；通信中线路畅通或中断等等，均可用服从 $(0-1)$ 分布的随机变量来描述，但对不同的问题，参数 $p$ 的取值不同.

2. 二项分布

**定义 5** 若随机变量 $X$ 可能取值为 $0,1,2,\cdots,n$，且

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

其中 $0<p<1,\ p+q=1$，$n$ 为非负整数，则称 $X$ 服从参数为 $n$、$p$ 的二项分布或伯努利分布，记为 $X\sim B(n,p)$.

容易看出，当 $n=1$ 时，二项分布正是 $(0-1)$ 分布，故当 $X$ 服从 $(0-1)$ 分布时，常记为 $X\sim B(1,p)$.

可以证明二项分布满足：

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

（2）$\sum\limits_{k=0}^{n}P\{X=k\}=\sum\limits_{k=0}^{n}C_n^k p^k q^{n-k}=(p+q)^n=1$.

由（2）可知，随机变量 $X$ 取 $k$ 值的概率

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

恰好是 $(p+q)^n$ 的展开式的第 $k+1$ 项，所以称 $X$ 服从二项分布.

二项分布来自于 $n$ 重伯努利试验模型，在 $n$ 重伯努利试验中，若事件 $A$ 在每次试验中发生的概率为 $p\ (0<p<1)$，则在 $n$ 次试验中事件 $A$ 发生的可能次数为 $0,1,2,\cdots,n$，事件 $A$ 发生 $k$ 次的概率为 $C_n^k p^k q^{n-k} \quad (k=0,1,2,\cdots,n)$.

由此可见，“事件 $A$ 发生的次数” $X$ 这个随机变量服从二项分布 $B(n,p)$.

“$A$ 发生 $k$ 次”即“ $X=k$ ”，由于

$$\bigcup_{k=0}^{n}\{X=k\}=\bigcup_{k=0}^{n}\{\text{事件}A\text{发生}k\text{次}\}=\Omega,$$

故有 $\sum_{k=0}^{n} C_n^k p^k q^{n-k} = P\{\Omega\}=1$，同时也说明 $\sum_{k=0}^{n}P\{X=k\}=1$．

**例 6** 设某工厂产品的次品率为 0.02，从该厂生产的一大批产品中随机抽取 100 件进行检测，求：

（1）求次品数的的分布列；

（2）恰有 2 件次品的概率；

（3）次品数不超过 3 件的概率．

**解** 由于产品非常多，无论是有放回抽样还是无放回抽样，都可以作为有放回抽样来处理．令 $X$ 表示 100 件产品中的次品数，则 $X\sim B(100,0.02)$．

（1）其分布列为

$$P\{X=k\}=C_{100}^k(0.02)^k(0.98)^{100-k}\quad(k=0,1,2,\cdots,100);$$

（2）恰有 2 件次品的概率

$$P\{X=2\}=C_{100}^2(0.02)^2(0.98)^{98};$$

（3）次品数不超过 3 件的概率

$$\begin{aligned}P\{X\leqslant 3\}&=C_{100}^0\times 0.02^0\times 0.98^{100}+C_{100}^1\times 0.02^1\times 0.98^{99}\\&\quad+C_{100}^2\times 0.02^2\times 0.98^{98}+C_{100}^3\times 0.02^3\times 0.98^{97}.\end{aligned}$$

**例 7** 某篮球运动员投篮 3 次，每次投中的概率为 0.6，求投中次数的分布列．

**解** 令 $X$ 表示投中的次数，则 $X\sim B(3,0.6)$，$X$ 的可能取值为 $0,1,2,3$，相应的概率分别为

$$P\{X=0\}=C_3^0(0.6)^0(0.4)^3=0.064,$$
$$P\{X=1\}=C_3^1(0.6)^1(0.4)^2=0.288,$$
$$P\{X=2\}=C_3^2(0.6)^2(0.4)^1=0.432,$$
$$P\{X=3\}=C_3^3(0.6)^3(0.4)^0=0.216.$$

即 $X$ 的概率分布．其概率分布图如图 2.1．

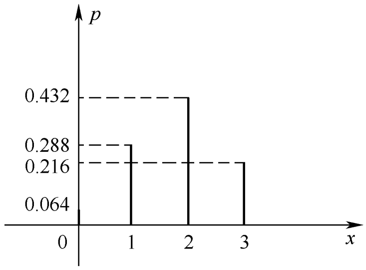


图 2. 1

从图 2.1 中看到，$P\{X=k\}$ 的概率先是随着 $k$ 的增大而增加，直到达到最大值，而后单调减少．一般的二项分布 $B(n,p)$ 都具有这一性质．

3．泊松（Poisson）分布

**定义 6** 若随机变量 $X$ 的分布列为

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

则称 $X$ 服从参数为 $\lambda$ 的泊松分布，记为 $X\sim P(\lambda)$．

泊松分布满足两个基本性质：

（1）$P\{X=k\}\geqslant 0 \quad (k=0,1,2,\cdots)$；

（2）$\sum\limits_{k=0}^{\infty}P\{X=k\}=1$．

证明（2）式，只需用到 $\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!}=\mathrm{e}^{\lambda}$．

泊松分布是一种常见的重要分布之一，服从泊松分布的随机现象特别集中在社会生活和物理学领域，在社会生活中，又尤其适用于各种服务的需求现象或排队现象．如纺织厂生产的一批布匹上的疵点个数；某种昆虫产卵个数；在一段时间间隔里放射性物质发出的经过计数器的 $\alpha$ 粒子数；某一地区一段时间内发生的交通事故的次数；某市级医院一天内的急诊病人数；某公共汽车终点站一段时间内的乘客数；一本书一页中的印刷错误数等等，都服从或近似服从泊松分布；同时泊松分布与二项分布具有密切的联系．

泊松分布可以看作二项分布的近似，从而产生了二项分布的近似公式，法国数学家泊松（S. D. Poisson）在 1837 年引入了一个著名的定理．

**定理 2.1（泊松定理）** 设随机变量 $X\sim B(n,p)$，如果

$$\lim_{n\to\infty}np=\lambda\ (\lambda>0),$$

则

$$\lim_{n\to\infty}C_n^k p^k(1-p)^{n-k}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \quad (k=0,1,2,\cdots).$$

**证** 依题意有

$$\begin{aligned}C_n^k p^k(1-p)^{n-k}&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\frac{\lambda^k}{k!}\frac{n}{n}\cdot\frac{n-1}{n}\cdot\cdots\cdot\frac{n-k+1}{n}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}\\&=\frac{\lambda^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}.\end{aligned}$$

对任意固定的非负整数 $k$，有

$$\lim_{n\to\infty}\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]=1,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=\mathrm{e}^{-\lambda},$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1,$$

从而

$$\lim_{n\to\infty}C_n^k p^k(1-p)^{n-k}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\qquad(k=0,1,2,\cdots).$$

当$n$很大，$p$很小时，有近似公式

$$C_n^k p^k q^{n-k}\approx\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\quad(k=0,1,2,\cdots,n),$$

其中$\lambda=np$.

在实际计算中，当$n\geqslant 10$，$p\leqslant 0.1$时，就可用泊松分布来近似二项分布．泊松分布的值可从书后附表中查得.

**例 8** 在例 6（2),（3）中，

（2）恰有 2 件次品的概率为

$$P\{X=2\}=C_{100}^2(0.02)^2(0.98)^{98}\approx\frac{2^2}{2!}\mathrm{e}^{-2}=2\mathrm{e}^{-2};$$

（3）次品数不超过 3 件的概率为

$$\begin{aligned}P\{X\leqslant 3\}&=C_{100}^0\times 0.02^0\times 0.98^{100}+C_{100}^1\times 0.02^1\times 0.98^{99}\\&\quad+C_{100}^2\times 0.02^2\times 0.98^{98}+C_{100}^3\times 0.02^3\times 0.98^{97}\\&\approx\mathrm{e}^{-2}+\frac{2^1}{1!}\mathrm{e}^{-2}+\frac{2^2}{2!}\mathrm{e}^{-2}+\frac{2^3}{3!}\mathrm{e}^{-2}=\frac{19}{3}\mathrm{e}^{-2}.\end{aligned}$$

**例 9** 设每分钟通过某交叉路口的汽车流量$X$服从泊松分布，且已知在一分钟内无车辆通过与恰有一辆车通过的概率相同，求在一分钟内至少有两辆车通过的概率.

**解** 设$X$服从参数为$\lambda$的泊松分布，由题意知

$$P\{X=0\}=P\{X=1\},$$

即

$$\frac{\lambda^0}{0!}\mathrm{e}^{-\lambda}=\frac{\lambda^1}{1!}\mathrm{e}^{-\lambda}.$$

解得

$$\lambda=1,$$

因此，至少有两辆车通过的概率为

$$P\{X \geqslant 2\} = 1 - P\{X=0\} - P\{X=1\}$$
$$= 1 - \frac{\lambda^0}{0!}e^{-1} - \frac{1^1}{1!}e^{-1} = 1 - 2e^{-1}.$$

## 习题 2.2

1．下列给出的数列，哪些是随机变量的分布列，并说明理由．

（1）$p_k = \frac{k}{15}$，$k = 0, 1, 2, 3, 4, 5$；

（2）$p_k = \frac{5-k^2}{6}$，$k = 0, 1, 2, 3$；

（3）$p_k = \frac{1}{4}$，$k = 2, 3, 4, 5$；

（4）$p_k = \frac{k+1}{25}$，$k = 1, 2, 3, 4, 5$．

2．某射手每次射击的命中率为 $p$，现射手一次接一次地射击直到击中为止，以 $X$ 表示命中时所耗费的子弹数，求 $X$ 的分布列．

3．设随机变量 $X$ 的分布列为 $P\{X=k\} = \frac{C}{2^k}\ (k = 0, 1, 2, 3, 4)$，试求：

（1）常数 $C$；

（2）$P\{X \leqslant 2\}$ 和 $P\left\{\frac{1}{2} < X < \frac{5}{2}\right\}$．

4．从五个数 $1, 2, 3, 4, 5$ 中任取三个数 $x_1, x_2, x_3$．

（1）求 $X = \max\{x_1, x_2, x_3\}$ 的分布列及 $P\{X \leqslant 4\}$；

（2）求 $Y = \min\{x_1, x_2, x_3\}$ 的分布列及 $P\{Y > 3\}$．

5．在 5 件产品中有 2 件次品，从中任取 2 件．用随机变量 $X$ 表示其中的次品数，写出 $X$ 的分布列和分布函数．

6．一汽车沿一街道行驶，需要通过 3 个均设有红绿信号灯的路口，每个路口信号灯为红或绿与其他路口信号灯为红或绿相互独立，且红、绿两种信号显示的时间相等，以 $X$ 表示该汽车首次遇到红灯前已通过的路口个数，求 $X$ 的分布列．

7．设 $X \sim B(1, p)$，若 $X$ 取 1 的概率为它取 0 的概率的两倍．试求 $X$ 的分布列和分布函数．

8．一大楼装有 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 0.1，问在同一时刻：

（1）恰有 2 个设备被使用的概率；

（2）至少有 3 个设备被使用的概率；

（3）至少有 1 个设备被使用的概率．

9．设有 80 台同类型设备，各台工作相互独立，发生故障的概率都是 0.01，且一台设备的故障一个人能维修．考虑两种配备维修工人的方案：

（1）由 4 个人维修，每人承包 20 台；

（2）由 3 个人共同维护 80 台.

试比较两种方案的优劣.

10．设 $X \sim P(\lambda)$，且 $P\{X=1\}=P\{X=2\}$，求 $P\{X=4\}$.

## 2.3　一维连续型随机变量

### 2.3.1　一维连续型随机变量及其概率密度

1. 概率密度函数

**定义 7**　设 $F(x)$ 是随机变量 $X$ 的分布函数，若存在一个非负函数 $f(x)$，有

$$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t \ (-\infty<x<+\infty), \tag{2.2}$$

则称 $X$ 为连续型随机变量，同时称 $f(x)$ 为 $X$ 的概率密度函数，简称密度函数或概率密度.

2. 密度函数的性质

由密度函数的定义可知，$f(x)$ 具有以下性质：

（1）$f(x)\geqslant 0$；

（2）$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$.

若某个函数满足性质（1）、（2），则此函数可作为某个随机变量的密度函数.

由定义和积分学知识，可得以下结论：

（1）在整个实数轴上，$F(x)$ 是连续的，即连续型随机变量的分布函数一定是连续的；

（2）在密度函数 $f(x)$ 的连续点 $x_0$ 处有

$$F'(x_0)=f(x_0), \tag{2.3}$$

即密度函数为分布函数的导数.

式（2.2）与式（2.3）表示了分布函数与概率密度之间的两个关系，利用这些关系，可以根据分布函数和概率密度中的一个推出另一个.

如果 $X$ 为连续型随机变量，那么

$$\begin{aligned} P\{a<X\leqslant b\}&=P\{X\leqslant b\}-P\{X\leqslant a\}=F(b)-F(a)\\ &=\int_{-\infty}^{b} f(x)\mathrm{d}x-\int_{-\infty}^{a} f(x)\mathrm{d}x=\int_{a}^{b} f(x)\mathrm{d}x . \end{aligned} \tag{2.4}$$

下面计算 $P\{X=a\}$，因为

$$P\{X=a\}\leqslant P\{a\leqslant X<a+\Delta x\}=\int_{a}^{a+\Delta x} f(t)\mathrm{d}t ,$$

而

$$0 \leqslant P\{X=a\} \leqslant \lim_{\Delta x \to 0} \int_a^{a+\Delta x} f(t)\mathrm{d}t = 0,$$

所以

$$P\{X=a\}=0.$$

即连续型随机变量取任何固定值的概率都为零，这与离散型随机变量的情形完全相反．但是$P\{X=a\}=0$，并不意味着事件$\{X=a\}$是不可能事件，因为$x=a$毕竟是实数轴上实实在在的一点，倘若你能无限次地进行重复试验的话，$X$是不会永远跳过$a$的，$P\{X=a\}=0$只说明在大量重复试验中，$\{X=a\}$发生的次数非常的少．

同理，概率为 1 的事件未必是必然事件．

因为$P\{X=a\}=0$，所以对于连续型随机变量有

$$\begin{aligned} p\{a<X<b\} &= P\{a<X\leqslant b\} = P\{a\leqslant X<b\} \\ &= P\{a\leqslant X\leqslant b\} = \int_a^b f(x)\mathrm{d}x. \end{aligned}$$

**例 10**　设连续型随机变量$X$的密度函数为

$$f(x)=\begin{cases} kx+1, & 0\leqslant x\leqslant 2, \\ 0, & \text{其他}. \end{cases}$$

试确定常数$k$及分布函数$F(x)$，并求$P\{1\leqslant X\leqslant 2\}$．

**解**　由$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x=1$，有

$$\int_0^2 (kx+1)\mathrm{d}x = 2k+2=1,$$

解得

$$k=-\frac{1}{2},$$

当$x<0$时，

$$F(x)=0,$$

当$0\leqslant x<2$时，

$$F(x)=\int_{-\infty}^x f(t)\mathrm{d}t = \int_0^x \left(-\frac{t}{2}+1\right)\mathrm{d}t = -\frac{x^2}{4}+x,$$

当$x\geqslant 2$时，

$$F(x)=1.$$

所以分布函数$F(x)$为

$$F(x)=\begin{cases} 0, & x<0, \\ -\dfrac{x^2}{4}+x, & 0\leqslant x<2, \\ 1, & x\geqslant 2. \end{cases}$$

而

$$P\{1 \leqslant X \leqslant 2\} = F(2) - F(1) = 1 - \frac{3}{4} = \frac{1}{4}.$$

### 2.3.2　几种常见的连续型随机变量的分布

1. 均匀分布

**定义 8**　若连续型随机变量 $X$ 的密度函数为

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b, \\ 0, & \text{其他}, \end{cases} \tag{2.5}$$

则称 $X$ 在区间 $[a,b]$ 上服从均匀分布，记为 $X \sim U[a,b]$.

其相应的分布函数为

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leqslant x < b, \\ 1, & x \geqslant b. \end{cases} \tag{2.6}$$

均匀分布的密度函数满足：

（1）$f(x) \geqslant 0$（显然）；

（2）$\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_a^b \frac{1}{b-a}\mathrm{d}x = 1$.

若 $a \leqslant c < d \leqslant b$，易得

$$P\{c < X < d\} = \int_c^d \frac{1}{b-a}\mathrm{d}x = \frac{d-c}{b-a}.$$

上式说明在 $[a,b]$ 上服从均匀分布的随机变量 $X$ 落入 $[a,b]$ 中任一子区间 $(c,d)$ 内的概率与该子区间的长度 $d-c$ 成正比，而与子区间在 $[a,b]$ 上的具体位置无关．即它落入区间 $[a,b]$ 中任意等长度的子区间内的概率是相同的，所以均匀分布也称为等概率分布．由此，均匀分布和第 1 章第 2 节中的几何概型有密切的联系．

在实际问题中，公共汽车站乘客的候车时间、近似计算中的舍入误差等都服从均匀分布．

**例 11**　试用均匀分布求解第 1 章例 7 的概率：

某公共汽车站从上午 7 时起，每隔 15 分钟来一趟车，一乘客在 7:00 到 7:30 之间随机到达该车站，求：

（1）该乘客等候不到 5 分钟乘上车的概率；

（2）该乘客等候时间超过 10 分钟才乘上车的概率．

**解**　设乘客于上午 7 点过 $X$ 分到达车站，由于乘客在 7:00 到 7:30 之间随机到达，因此 $X$ 是服从区间 $(0,30)$ 上的均匀分布，即 $X$ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{30}, & 0\leqslant x\leqslant 30,\\ 0, & \text{其他}.\end{cases}$$

（1）该乘客等候时间不到 5 分钟，必须且只需在 7:10 到 7:15 之间或在 7:25 到 7:30 之间到达车站，因此所求概率为

$$P\{10<X<15\}+P\{25<X<30\}=\int_{10}^{15}\frac{1}{30}\mathrm{d}x+\int_{25}^{30}\frac{1}{30}\mathrm{d}x=\frac{1}{3}.$$

（2）同（1）的分析方法类似可得所求概率为

$$P\{0<X<5\}+P\{15<X<20\}=\frac{1}{3}.$$

2. 指数分布

**定义 9** 若连续型随机变量 $X$ 的密度函数为

$$f(x)=\begin{cases}\lambda\mathrm{e}^{-\lambda x}, & x\geqslant 0,\\ 0, & x<0.\end{cases} \tag{2.7}$$

其中 $\lambda>0$ 为常数，则称随机变量 $X$ 服从参数为 $\lambda$ 的指数分布，记为 $X\sim E(\lambda)$.

指数分布的密度函数满足：

（1）$f(x)\geqslant 0$（显然）；

（2）$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\int_{0}^{+\infty}\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x=1$.

其相应的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ 1-\mathrm{e}^{-\lambda x}, & x\geqslant 0.\end{cases} \tag{2.8}$$

在实际问题中，电子元件的寿命和随机服务系统中的服务时间等都服从指数分布.

**例 12** 假设某元件的寿命 $X$ 是一个连续型随机变量，其概率密度为

$$f(x)=\begin{cases}C\mathrm{e}^{-\frac{x}{100}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

（1）确定常数 $C$；

（2）求寿命超过 100 小时的概率.

**解** （1）由密度函数性质 2 知

$$\int_{0}^{+\infty}C\mathrm{e}^{-\frac{x}{100}}\mathrm{d}x=\left(-100C\mathrm{e}^{-\frac{x}{100}}\right)\Bigg|_{0}^{+\infty}=100C=1,$$

由此得 $C=0.01$.

所以

$$X\sim E(0.01);$$

（2）寿命超过 100 小时的概率为

$$P\{X>100\}=1-F(100)=1-(1-\mathrm{e}^{-0.01\times100})=\mathrm{e}^{-1}\approx0.3679\ .$$

3．正态分布

若连续型随机变量 $X$ 的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<+\infty)\ ,\qquad（2.9）$$

其中 $\mu,\sigma>0$ 为常数，则称 $X$ 服从参数为 $\mu$ 和 $\sigma$ 的正态分布，或高斯（Gauss）分布，记为 $X\sim N(\mu,\sigma^2)$．其分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t\quad(-\infty<x<+\infty)\ .\qquad（2.10）$$

正态分布的密度函数满足：

（1）$f(x)\geqslant0\quad(-\infty<x<+\infty)$；

（2）$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$．

性质（1）显然；验证性质（2），在（2.9）式中，只需令 $t=\dfrac{x-\mu}{\sigma}$，则 $\mathrm{d}x=\sigma\mathrm{d}t$，故有

$$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=1\quad\left(\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\sqrt{2\pi}\right).$$

当 $\mu=0$，$\sigma^2=1$ 时，$X$ 的密度函数记为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\quad(-\infty<x<+\infty)\ ,\qquad（2.11）$$

称 $X$ 服从**标准正态分布**．记为 $X\sim N(0,1)$．其分布函数记为

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\quad(-\infty<x<+\infty)\ .\qquad（2.12）$$

由正态分布的密度函数可以画出它的图像，称其为正态曲线．易知正态分布的密度函数还具有下列性质：

（1）$f(x)$ 在 $(-\infty,+\infty)$ 内处处连续；

（2）$f(x)$ 的图形关于直线 $x=\mu$ 对称；

（3）$f(x)$ 在 $x=\mu$ 处取得最大值为 $\dfrac{1}{\sqrt{2\pi}\sigma}$；

（4）$f(x)$ 在点 $x=\mu\pm\sigma$ 处对应有拐点；

（5）$f(x)$ 在 $x$ 轴上方，且以 $x$ 轴为水平渐近线；

（6）若固定 $\sigma$ 而改变 $\mu$ 值，则曲线左右位置不同，但形状不变，即此时 $f(x)$ 图形沿着 $x$ 轴平行移动，参数 $\mu$ 决定曲线 $f(x)$ 的位置，故称 $\mu$ 为位置参数；

（7）若固定 $\mu$ 值而改变 $\sigma$ 值，则曲线形状改变而位置不变，曲线的峰顶为 $f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$，$\sigma$ 值越大，曲线越平缓，即分布越分散；$\sigma$ 值越小，曲线越陡峭，即分布越集中．从几何直观上可看出，参数 $\sigma$ 决定曲线 $f(x)$ 的形状，故称 $\sigma$ 为形状参数，它反映了 $X$ 所取值的离散程度．

正态分布是概率论中最常见的分布之一，它在概率统计的理论与应用中都占有头等重要的地位．自然界和工程技术中的很多随机变量都是服从正态分布．例如，测量的误差、一批产品的质量指标、人体的身高或体重、作物的单位面积产量、弹着点的分布、气象中的月平均气温、湿度、降水量等都服从或近似服从正态分布．此外，它在产品检验、无线电噪声理论和自动控制等领域也有着广泛的应用．

只要令 $s=\dfrac{t-\mu}{\sigma}$（称为标准化），就可把正态随机变量的分布函数（2.10）式化为用标准正态随机变量的分布函数 $\Phi(x)$ 表示的形式，即

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}}\mathrm{e}^{-\frac{s^2}{2}}\mathrm{d}s=\Phi\left(\frac{x-\mu}{\sigma}\right). \tag{2.13}$$

从而一般正态分布的随机变量落在某区域内的概率可以利用标准正态分布来计算，即

$$P\{a<X\leqslant b\}=F(b)-F(a)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.14}$$

上式右端的值可查书后附表．

$\Phi(x)$ 还具有如下性质：

（1）$\Phi(-x)=1-\Phi(x)$；（因此 $x<0$ 时，可先求 $\Phi(-x)$ 的取值）

（2）$\Phi(0)=\dfrac{1}{2}$；

（3）$P\{|X|\leqslant x\}=2\Phi(x)-1$；

（4）$P\{|X|\geqslant x\}=2[1-\Phi(x)]$．

**例 13** 已知 $X\sim N(0,1)$，求：

（1）$P\{X\leqslant 2.35\}$；（2）$P\{X\leqslant -1.24\}$；（3）$P\{|X|\leqslant 1.54\}$．

**解** （1）$P\{X\leqslant 2.35\}=\Phi(2.35)=0.9906$；

（2）$P\{X\leqslant -1.24\}=\Phi(-1.24)=1-\Phi(1.24)=0.1075$；

（3）$P\{|X|\leqslant 1.54\}=2\Phi(1.54)-1=2\times 0.9382-1=0.8764$．

**例 14** 设随机变量 $X\sim N(3,4)$，求 $P\{-1<X<4\}$ 和 $P\{X\geqslant 2\}$．

**解** $P\{-1<X<4\}=F(4)-F(-1)=\Phi\left(\dfrac{4-3}{2}\right)-\Phi\left(\dfrac{-1-3}{2}\right)$

$$=\Phi\left(\frac{1}{2}\right)-\Phi\left(-2\right)=\Phi\left(\frac{1}{2}\right)-[1-\Phi(2)]$$

$$=0.6915+0.9772-1=0.6687.$$

$$P\{X\geqslant 2\}=1-P\{X<2\}=1-\Phi\left(\frac{2-3}{2}\right)=1-\Phi\left(-\frac{1}{2}\right)$$

$$=1-\left[1-\Phi\left(\frac{1}{2}\right)\right]=\Phi\left(\frac{1}{2}\right)=0.6915.$$

**例 15** 设随机变量 $X\sim N(\mu,\sigma^2)$，求：$P\left\{\left|X-\mu\right|<k\sigma\right\}$.

**解** $P\left\{\left|X-\mu\right|<k\sigma\right\}=P\left\{\left|\dfrac{X-\mu}{\sigma}\right|<k\right\}=2\Phi(k)-1$，所以

$$P\left\{\left|X-\mu\right|<\sigma\right\}=2\Phi(1)-1=0.6826,$$

$$P\left\{\left|X-\mu\right|<2\sigma\right\}=2\Phi(2)-1=0.9544,$$

$$P\left\{\left|X-\mu\right|<3\sigma\right\}=2\Phi(3)-1=0.9974.$$

由此例可看出，若随机变量 $X\sim N(\mu,\sigma^2)$，则随机变量 $X$ 落在区间 $[\mu-3\sigma,\mu+3\sigma]$ 内的概率几乎等于 1，即 $X$ 值落入以 $\mu$ 为中心，$3\sigma$ 为半径的区间内几乎是必然的，而事件 $\left\{\left|X-\mu\right|>3\sigma\right\}$ 是一个小概率事件．这就是著名的“$3\sigma$ 原则”，它在工业生产中常用来作为质量控制的依据．

## 习题 2.3

1．设随机变量 $X$ 的概率密度为

$$f(x)=\begin{cases}\dfrac{A}{\sqrt{1-x^2}}, & |x|\leqslant 1,\\ 0, & |x|>1.\end{cases}$$

试求：（1）系数 $A$；（2）$P\left\{-\dfrac{1}{2}<X<\dfrac{1}{2}\right\}$；（3）$X$ 的分布函数．

2．设随机变量 $X$ 的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

试求：（1）$P\{X\leqslant 2\}$；（2）$P\{X>3\}$；（3）密度函数 $f(x)$．

3．已知随机变量 $X$ 的分布函数为

$$F(x)=A+B\arctan x，-\infty<x<+\infty，$$

试求：（1）常数 $A$ 与 $B$ 的值；（2）$P\{-1<X<1\}$；（3）密度函数 $f(x)$．

4．设 $X \sim U[2,4]$，试求（1）$P\{1 \leqslant X < 3\}$；（2）$P\left\{(X-3)^2 < 0.25\right\}$．

5．设 $X$ 在 $[1,6]$ 上服从均匀分布，求方程 $x^2+x+1=0$ 有实根的概率．

6．设电子管寿命 $X$ 的概率密度为

$$f(x)=\begin{cases}\dfrac{100}{x^2}, & x>100,\\ 0, & \text{其他}.\end{cases}$$

若一架收音机上装有 3 个这种管子，试求：

（1）使用的最初 150 小时内至少有 2 个电子管被烧坏的概率；

（2）在使用的最初 150 小时内烧坏的电子管数 $Y$ 分布列；

（3）$Y$ 的分布函数．

7．设顾客在某银行窗口等待服务的时间 $X$（单位：分）服从参数为 $\dfrac{1}{5}$ 的指数分布，若等待时间超过 10 分钟，他就离开．设他一个月内要来银行 5 次，以 $Y$ 表示一个月内他没有等到服务而离开窗口的次数，求 $Y$ 的分布列及 $P\{Y \geqslant 1\}$．

8．设 $X \sim N(2.5, 4)$，试求（1）$P\{X>5\}$；（2）$P\{X<-1\}$；（3）$P\left\{|X-2|<3\right\}$．

9．设 $X \sim N(5,4)$，确定 $C$，使 $P\{X<C\}=P\{X>C\}$．

10．设 $X \sim N(2,\sigma^2)$，且 $P\{2<X<4\}=0.3$，试求 $P\{X<0\}$．

11．$X \sim N(108,9)$，试求：（1）$a$，使 $P\{X \leqslant a\}=0.9$；（2）$b$，使 $P\left\{|X-b|>b\right\}=0.1$．

12．某系一年级概率统计的考试成绩近似服从正态分布 $N(75,10^2)$，如果 88 分以上优秀，问概率统计成绩优秀的学生占全系一年级总学生数的百分之几？

## 2.4 一维随机变量函数的分布

在许多实际问题中，所考虑的随机变量常常依赖于另一个随机变量．一般地，设 $X$ 是一个随机变量，$g(x)$ 是一个实函数，则函数 $Y=g(X)$ 也是一个随机变量，称之为随机变量 $X$ 的函数．利用 $X$ 的分布可求出 $Y=g(X)$ 的分布，下面分两种情况进行讨论．

### 2.4.1 离散型随机变量函数的分布列

设离散型随机变量 $X$ 的分布列为

| $X$ | $x_1$ | $x_2$ | $\cdots$ | $x_k$ | $\cdots$ |
|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | $\cdots$ | $p_k$ | $\cdots$ |

则 $Y=g(X)$ 也是一个离散型随机变量，其所有可能取值为 $y_k=g(x_k)$ $(k=1,2,\cdots)$．

如果 $g(x_i)$ 值全不相等，即当 $x_i \neq x_j$ 时，$g(x_i) \neq g(x_j)$ $(i, j = 1, 2, \cdots, k, \cdots)$，这时有

$$P\{Y = y_i\} = P\{Y = g(x_i)\} = P\{X = x_i\} \ (i = 1, 2, \cdots, k, \cdots),$$

即随机变量 $Y = g(X)$ 的分布列为

| $Y$ | $g(x_1)$ | $g(x_2)$ | $\cdots$ | $g(x_k)$ | $\cdots$ |
|---|---|---|---|---|---|
| $P$ | $p_1$ | $p_2$ | $\cdots$ | $p_k$ | $\cdots$ |

如果 $y_k = g(x_k)$ $(k = 1, 2, \cdots)$ 中有相等的值时，则应将那些相等的值分别合并，由于事件

$$\{Y = y_k\} = \sum_{g(x_i) = y_k} \{X = x_i\},$$

则由概率的可加性可得事件 $\{Y = y_k\}$ 的概率为

$$P\{Y = y_k\} = \sum_{g(x_i) = y_k} P\{X = x_i\} = \sum_{g(x_i) = y_k} p_i,$$

从而得出 $Y = g(X)$ 的分布列.

**例 16** 设离散型随机变量 $X$ 的分布列为

| $X$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|
| $P$ | 0.2 | 0.3 | 0.4 | 0.1 |

试分别求 $Y = 2X$，$Z = X^2 - 1$ 的分布列.

**解**

| $X$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|
| $P$ | 0.2 | 0.3 | 0.4 | 0.1 |
| $Y = 2X$ | $-2$ | $0$ | $2$ | $4$ |
| $Z = X^2 - 1$ | $0$ | $-1$ | $0$ | $3$ |

故 $Y$ 的分布列为

| $Y$ | $-2$ | $0$ | $2$ | $4$ |
|---|---|---|---|---|
| $P$ | 0.2 | 0.3 | 0.4 | 0.1 |

$Z$ 的分布列为

| $Z$ | $-1$ | $0$ | $3$ |
|---|---|---|---|
| $P$ | 0.3 | 0.6 | 0.1 |

## 2.4.2 连续型随机变量函数的分布

设连续型随机变量 $X$ 的密度函数为 $f_X(x)\,(-\infty < x < +\infty)$，又设 $g(x)$ 是严格单调且可导的函数，则 $Y = g(X)$ 也是一个连续性随机变量，其分布函数为

$$F_Y(y) = P\{Y \leqslant y\} = P\{g(X) \leqslant y\} = \int\limits_{g(x)\leqslant y} f_X(x)\mathrm{d}x\,.$$

$Y$ 的密度函数 $f_Y(y)$ 可由 $f_Y(y) = \dfrac{\mathrm{d}}{\mathrm{d}y}F_Y(y)$ 求得.

具体方法（分布函数法）：

（1）写出 $Y$ 的分布函数的关系式：

$$F_Y(y) = P\{Y \leqslant y\} = P\{g(X) \leqslant y\}\,.$$

（2）用 $X$ 的分布函数来表述 $P\{g(X) \leqslant y\}$.

先将“$g(X) \leqslant y$”表示成等价的 $X$ 的取值范围，再依分布函数的定义或性质作相应运算即可.

（3）求得 $f_Y(y)$.

实际上，在（2）中已得到了用 $X$ 的分布函数来表述 $Y$ 的分布函数 $F_Y(y)$，对 $F_Y(y)$ 关于 $y$ 求导，即可得到 $f_Y(y)$.

**例 17** 随机变量 $X$ 的概率密度为

$$f_X(x) = \begin{cases} \dfrac{x}{8}, & 0 < x < 4, \\ 0, & \text{其他}. \end{cases}$$

求 $Y = 2X - 1$ 的概率密度.

**解** 要得到 $Y = 2X - 1$ 的概率密度，先求它的分布函数 $F_Y(y)$，根据题意有

$$\begin{aligned} F_Y(y) &= P\{Y \leqslant y\} = P\{2X - 1 \leqslant y\} \\ &= P\left\{X \leqslant \frac{y+1}{2}\right\} = F_X\left(\frac{y+1}{2}\right), \end{aligned}$$

两边对 $y$ 求导，得

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}F_X\left(\frac{y+1}{2}\right) = f_X\left(\frac{y+1}{2}\right)\frac{\mathrm{d}\left(\dfrac{y+1}{2}\right)}{\mathrm{d}y} = \frac{1}{2}f_X\left(\frac{y+1}{2}\right),$$

当 $0 < x < 4$ 时，$f_X(x) \neq 0$，即 $-1 < y < 7$ 时，$F_Y(y) \neq 0$，此时

$$f_Y(y) = \frac{1}{2}f_X\left(\frac{y+1}{2}\right) = \frac{y+1}{32},$$

当 $x \leqslant 0$ 或 $x \geqslant 4$ 时，$f_X(x) = 0$，即 $y \leqslant -1$ 或 $y \geqslant 7$ 时 $F_Y(y) = 0$，

所以 $Y = 2X - 1$ 的概率密度为

$$f_Y(y)=\begin{cases}\dfrac{y+1}{32}, & -1<y<7,\\ 0, & \text{其他}.\end{cases}$$

**例 18**　已知 $X\sim N(\mu,\sigma^2)$，求 $Y=\dfrac{X-\mu}{\sigma}$ 的概率密度.

**解**　先求 $F_Y(y)=P\{Y\leqslant y\}=P\left\{\dfrac{X-\mu}{\sigma}\leqslant y\right\}=P\{X\leqslant\sigma y+\mu\}$.

将上式两边对 $y$ 求导并注意到

$$f_X(x)=F_X'(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

不难看出 $F_X'(y)$ 处处存在且连续，所以

$$f_Y(y)=F_Y'(y)=\sigma f_X(\sigma y+\mu)=\sigma\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\mu+\sigma y-\mu)^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}}.$$

即 $Y=\dfrac{X-\mu}{\sigma}\sim N(0,1)$，称 $Y$ 为 $X$ 的**标准化**.

**例 19**　设 $X$ 具有密度函数 $f_X(x)$，$x\in\mathbf{R}$，求随机变量 $Y=X^2$ 的密度函数.

**解**　设 $Y$ 的密度函数为 $f_Y(y)$，分布函数为 $F_Y(y)$，并设 $X$ 的分布函数为 $F_X(x)$.

当 $y>0$ 时，$Y$ 的分布函数为

$$\begin{aligned}F_Y(y)&=P\{Y\leqslant y\}=P\{X^2\leqslant y\}\\&=P\left\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\right\}\\&=F_X(\sqrt{y})-F_X(-\sqrt{y}),\end{aligned}$$

上式两边对 $y$ 求导数，得

$$\begin{aligned}f_Y(y)=F_Y'(y)&=\frac{1}{2\sqrt{y}}f_X(\sqrt{y})+\frac{1}{2\sqrt{y}}f_X(-\sqrt{y})\\&=\frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y})+f_X(-\sqrt{y})\right],\end{aligned}$$

当 $y\leqslant 0$ 时，由于 $Y=X^2\geqslant 0$，于是 $Y$ 的分布函数 $F_Y(y)=0$，故此时

$$f_Y(y)=F_Y'(y)=0.$$

综上所述，$Y$ 的分布函数为

$$f_Y(y)=\begin{cases}\dfrac{1}{2\sqrt{y}}\left[f_X(\sqrt{y})+f_X(-\sqrt{y})\right], & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

### 习题 2.4

1．设 $X$ 的分布列为

| $X$ | $-2$ | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|---|
| $P$ | 0.15 | 0.20 | 0.30 | 0.25 | 0.10 |

求 $Y=2X-1$ 及 $Z=X^2+1$ 的分布列.

2．设 $X$ 的分布列为

| $X$ | $\frac{\pi}{4}$ | $\frac{\pi}{2}$ | $\frac{3\pi}{4}$ |
|---|---|---|---|
| $P$ | 0.2 | 0.5 | 0.3 |

求 $Y=\sin X$ 的分布列.

3．设 $X$ 的密度函数为 $f(x)=\begin{cases}2x, & 0<x<1, \\ 0, & 其他.\end{cases}$ 求下列各随机变量函数的密度函数：

（1）$Y=2X$；（2）$Y=-X+1$；（3）$Y=X^2$.

4．设 $X\sim U[0,1]$，求下列各随机变量函数的密度函数：

（1）$Y=3X+1$；（2）$Y=\ln X$；（3）$Y=\mathrm{e}^X$.

5．设 $X\sim E(1)$，求随机变量的函数 $Y=\mathrm{e}^X$ 的密度函数.

6．设 $X\sim N(0,1)$，求下列各随机变量函数的密度函数：

（1）$Y=\mathrm{e}^X$；（2）$Y=|X|$；（3）$Y=2X^2+1$.

## 本章小结

本章详细介绍了随机变量及其分布，用随机变量来描述随机事件是概率论中最重要的方法，其思路是将样本数量化，即用实数来标识随机事件.

一、随机变量的分类

$$随机变量\begin{cases}离散型\\ 非离散型\begin{cases}连续型\\ 其他\end{cases}\end{cases}$$

若随机变量的所有可能取值为有限个或可列无限个，就称此随机变量为离散型随机变量，否则称为非离散型随机变量.

二、分布函数

无论是离散型的还是非离散型的随机变量 $X$，都可利用分布函数

$$F(x)=P\{X\leqslant x\}\quad(-\infty<x<+\infty),$$

来描述．若已知随机变量 $X$ 的分布函数，就可知道 $X$ 落在任一区间 $(x_1,x_2]$ 上的概率

$$P\{x_1 < X \leqslant x_2\} = F(x_2) - F(x_1)\ .$$

可见，分布函数完整地描述了随机变量取值的统计规律性，它是研究随机变量的重要工具，并为利用高等数学来研究随机变量提供了方便.

三、常见的三种离散型随机变量

1．$(0-1)$ 分布的分布列为

| $X$ | 0 | 1 |
| --- | --- | --- |
| $p_k$ | $1-p$ | $p$ |

其中 $p \geqslant 0$；

2．二项分布 $X \sim B(n,p)$ 的分布列为

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

3．泊松分布 $X \sim P(\lambda)$ 的分布列为

$$P\{X=k\} = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \quad (k=0,1,2,\cdots)\text{，其中 } \lambda > 0\ .$$

四、连续型随机变量

设随机变量 $X$ 的分布函数为 $F(x)$，若存在非负函数 $f(x)$，使得对于任意 $x$，有

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t\ ,$$

则称 $X$ 为连续型随机变量，其中 $f(x)\ (\geqslant 0)$ 称为 $X$ 的概率密度.

常见的三种连续型随机变量

1．均匀分布 $X \sim U[a,b]$，其密度函数和分布函数分别为

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b, \\ 0, & \text{其他；} \end{cases}$$

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leqslant x < b, \\ 1, & x \geqslant b. \end{cases}$$

2．指数分布的密度函数和分布函数分别为

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0; \end{cases} \quad \text{其中 } \lambda > 0\ ,$$

$$F(x) = \begin{cases} 1-\mathrm{e}^{-\lambda x}, & x \geqslant 0, \\ 0, & \text{其他.} \end{cases}$$

3．一般正态分布的密度函数和分布函数分别为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\quad(-\infty<x<+\infty),$$

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t\quad(-\infty<x<+\infty);$$

标准正态分布的密度函数和分布函数分别为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\quad(-\infty<x<+\infty),$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\quad(-\infty<x<+\infty).$$

五、随机变量函数

一般地，

$$f(x)=\begin{cases}\sum\limits_{x_k\leqslant x}p_k, & X\text{为离散型随机变量,}\\ \int_{-\infty}^{x}f(t)\mathrm{d}t, & X\text{为连续型随机变量.}\end{cases}$$

随机变量 $X$ 的函数 $Y=g(X)$ 也是一个随机变量，掌握利用 $X$ 的分布（$X$ 的分布列或概率密度）求 $Y=g(X)$ 的分布（$Y$ 的分布列或概率密度）的方法.

# 复习题 2

**一、填空题**

1．设离散型随机变量 $X$ 的分布函数为 $F(x)=\begin{cases}0, & x<-1,\\ a, & -1\leqslant x<1,\\ \frac{2}{3}-a, & 1\leqslant x<2,\\ a+b, & x\geqslant 2,\end{cases}$ 且 $P\{X=2\}=\frac{1}{2}$，则 $a=$________，$b=$________，$X$ 的分布列为________；

2．设连续型随机变量 $X$ 的概率密度为 $f(x)=\begin{cases}k\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$ 则 $k=$________，$P\{1<X\leqslant 2\}=$________，$P\{X=2\}=$________，$P\{X<2\}=$________；

3．设 5 个晶体管中有 2 个次品，3 个正品．如果每次从中任取 1 个进行测试，测试后的产品不放回，直到把 2 个次品都找到为止，需要进行的测试次数 $X$ 是一个随机变量，则 $P\{X=5\}=$________，$P\{X\leqslant 2\}=$________；

4．设随机变量 $X$ 服从参数为 $\lambda$ 的泊松分布，且 $P\{X=0\}=0.5\times P\{X=2\}$ ，则 $\lambda=$ ________；

5．设 $F(x)$ 是离散型随机变量的分布函数，若 $P\{X=b\}=$ ________，则 $P\{a<X<b\}=F(b)-F(a)$ 成立．

二、选择题

1．已知随机变量 $X$ 只能取 $-1,0,1,2,3$ 五个数值，其相应的概率依次为 $\frac{1}{2c},\frac{1}{4c},\frac{1}{8c},\frac{1}{16c},\frac{1}{16c}$ ，则 $c=$（　　）．

A．2　　B．3　　C．4　　D．1

2．若函数 $y=f(x)$ 是一随机变量 $X$ 的概率密度，则（　　）一定成立．

A． $f(x)$ 的定义域为 $[0,1]$　　B． $f(x)$ 的值域为 $[0,1]$

C． $f(x)$ 非负　　D． $f(x)$ 在 $(-\infty,+\infty)$ 内连续

3．如果 $F(x)$ 是（　　），则 $F(x)$ 一定不可以是连续型随机变量的分布函数．

A．非负函数　　B．连续函数

C．有界函数　　D．单调减少函数

4．下面各式中，（　　）可以作为某一随机变量 $X$ 的概率密度．

A． $f_1(x)=\begin{cases}\sin x，& x\in[0,\pi]\\ 0， & 其他\end{cases}$　　B． $f_2(x)=\begin{cases}\sin x，& x\in\left[0,\frac{3}{2}\pi\right]\\ 0， & 其他\end{cases}$

C． $f_3(x)=\begin{cases}\sin x，& x\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]\\ 0， & 其他\end{cases}$　　D． $f_4(x)=\begin{cases}\sin x，& x\in\left[0,\frac{\pi}{2}\right]\\ 0， & 其他\end{cases}$

5．随机变量 $P\{X<0\}=$ $X\sim N(2,\sigma^2)$ ， $P\{0<x<4\}=0.3$ ，则 $P\{X<0\}=$（　　）．

A．0.5　　B．0.3　　C．0.35　　D．0.7

三、计算题

1．设在 15 个同类型的零件中有 2 个是次品，在其中取 3 次，每次任取 1 个，做不放回抽样，以 $X$ 表示取出次品的次数，求 $X$ 的分布列，并计算 $P\{X\leqslant 1\}$ ．

2．设连续型随机变量 $X$ 的概率密度为 $f(x)=\begin{cases}a\cos x， & -\frac{\pi}{2}\leqslant x\leqslant\frac{\pi}{2}，\\ 0， & 其他.\end{cases}$

求（1）系数 $a$ ；（2） $X$ 的分布函数；（3） $P\{X\leqslant 1\}，P\{X=1\}，P\{-1<X<1\}$ ．

3．设连续型随机变量 $X$ 的分布函数为 $F(x)=\begin{cases}0， & x\leqslant 0，\\ Ax^3， & 0<x<2，\\ 1， & x\geqslant 2.\end{cases}$

求：（1）系数 $A$；（2） $X$ 的概率密度；（3） $P\{0<X<1\}$ ， $P\{1.5<X\leqslant 2\}$ ， $P\{2\leqslant X\leqslant 3\}$ ．

4．设 $k$ 在 $[0,5]$ 上服从均匀分布，求方程 $4x^2+4xk+k+2=0$ 有实根的概率．

5．某种型号的电灯泡使用时间（单位：小时）为一随机变量 $X$，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{5000}\mathrm{e}^{-\frac{x}{5000}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

求：3 个这种型号的电灯泡使用了 1000 小时后至少有 2 个仍可继续使用的概率．

6．某科统考成绩近似服从正态分布 $N(70,10^2)$，在参加统考的人中，及格者 100 人（及格分数为 60），计算：

（1）不及格人数；

（2）成绩前 10 名的人数在考生中所占的比例；

（3）估计第 10 名考生的成绩．