

## 第二章 随机变量

### 第一节 随机变量及其分布函数

我们讨论的随机事件中,有些是直接用量来标识的,例如,抽样检验灯泡质量试验中灯泡的寿命;而有些则不是直接用量来标识的,如性别抽查试验中所抽到的性别.为了更深入地研究各种与随机现象有关的理论和应用问题,我们有必要将样本空间的元素与实数对应起来,即将随机试验的每个可能的结果  $e$  都用一个实数  $X$  来表示.例如,在性别抽查试验中用实数“1”表示“出现男性”,用“0”表示“出现女性”.显然,一般来讲此处的实数  $X$  值将随  $e$  的不同而变化,它的值因  $e$  的随机性而具有随机性,我们称这种取值具有随机性的变量为随机变量.

**定义 2.1** 设随机试验的样本空间为  $\Omega$ , 如果对  $\Omega$  中每一个元素  $e$ , 有一个实数  $X(e)$  与之对应, 这样就得到一个定义在  $\Omega$  上的实值单值函数  $X=X(e)$ , 称之为随机变量 (random variable).

随机变量的取值随试验结果而定,在试验之前不能预知它取什么值,只有在试验之后才知道它的确切值,但所有可能的取值是确定的;试验的各个结果的出现有一定的概率,故随机变量取各个值也有一定的概率.这些性质显示了随机变量与普通函数之间有着本质的差异.再者,普通函数是定义在实数集或实数集的一个子集上的,而随机变量是定义在由样本空间的所有子集(事件)所构成的集合(我们也叫这一集合为样本空间)上的(样本空间的元素不一定是实数),这也是二者的差别.

本书中,我们一般以大写字母如  $X, Y, Z, W, \dots$  表示随机变量,而以小写字母如  $x, y, z, w, \dots$  表示实数.

为了研究随机变量的统计规律,并由于随机变量  $X$  的可能取值不一定能逐个列出,因此我们在一般情况下需研究随机变量的取值落在某区间  $(x_1, x_2]$  中的概率,即求  $P\{x_1 < X \leq x_2\}$ , 但由于

$$P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\},$$

由此可见要研究  $P\{x_1 < X \leq x_2\}$  就归结为研究形如  $P\{X \leq x\}$  的概率问题了.不难看出,  $P\{X \leq x\}$  的值常随不同的  $x$  而变化,它是  $x$  的函数,我们称该函数为分布函数.

**定义 2.2** 设  $X$  是随机变量,  $x$  为任意实数, 函数

$$F(x) = P\{X \leq x\}$$

称为  $X$  的分布函数 (distribution function).

对于任意实数  $x_1, x_2 (x_1 < x_2)$ , 有

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\{X \leq x_2\} - P\{X \leq x_1\} \\ &= F(x_2) - F(x_1), \end{aligned} \quad (2-1)$$

因此,若已知  $X$  的分布函数,我们就能知道  $X$  落在任一区间  $(x_1, x_2]$  上的概率.在这个意义上说,分布函数完整地描述了随机变量的统计规律性.

如果将  $X$  看成是数轴上的随机点的坐标,那么,分布函数  $F(x)$  在  $x$  处的函数值就表示  $X$  落在区间  $(-\infty, x]$  上的概率.

分布函数具有如下基本性质:

1°  $F(x)$  为单调不减的函数.

事实上,由 (2-1) 式,对于任意实数  $x_1, x_2 (x_1 < x_2)$ , 有

$$F(x_2) - F(x_1) = P\{x_1 < X \leq x_2\} \geq 0.$$

2°  $0 \leq F(x) \leq 1$ , 且  $\lim_{x \rightarrow +\infty} F(x) = 1$ , 常记为  $F(+\infty) = 1$ .

$$\lim_{x \rightarrow -\infty} F(x) = 0, \text{ 常记为 } F(-\infty) = 0.$$

我们从几何上说明这两个式子. 当区间端点  $x$  沿数轴无限向左移动 ( $x \rightarrow -\infty$ ) 时, 则“ $X$  落在  $x$  左边”这一事件趋于不可能事件, 故其概率  $P\{X \leq x\} = F(x)$  趋于 0; 又若  $x$  无限向右移动 ( $x \rightarrow +\infty$ ) 时, 事件“ $X$  落在  $x$  左边”趋于必然事件, 从而其概率  $P\{X \leq x\} = F(x)$  趋于 1.

3°  $F(x+0) = F(x)$ , 即  $F(x)$  为右连续.

证 略.

反过来可以证明, 任一满足这 3 个性质的函数, 一定可以作为某个随机变量的分布函数.

概率论主要是利用随机变量来描述和研究随机现象的, 而利用分布函数就能很好地表示各事件的概率. 例如,  $P\{X > a\} = 1 - P\{X \leq a\} = 1 - F(a)$ ,  $P\{X < a\} = F(a-0)$ ,  $P\{X = a\} = F(a) - F(a-0)$ , 等等. 在引进了随机变量和分布函数后我们就能利用高等数学的许多结果和方法来研究各种随机现象了, 它们是概率论的两个重要而基本的概念. 下面我们将分别以离散和连续两种类别来更深入地讨论随机变量及其分布函数, 另有一种奇异型随机变量超出本书范围, 就不作介绍了.

## 第二节 离散型随机变量及其分布

如果随机变量所有可能的取值为有限个或可数无穷多个, 则称这种随机变量为离散型随机变量.

容易知道, 要掌握一个离散型随机变量  $X$  的统计规律, 必须且只需知道  $X$  的所有可能取值以及取每一个可能取值的概率.

设离散型随机变量  $X$  所有可能的取值为  $x_k (k=1, 2, \dots)$ ,  $X$  取各个可能取值的概率, 即事件  $\{X=x_k\}$  的概率

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

我们称(2-2)式为离散型随机变量  $X$  的概率分布或分布律. 分布律也常用表格来表示(见表 2-1):

表 2-1

$X$	$x_1$	$x_2$	$x_3$	$\dots$	$x_k$	$\dots$
$p_k$	$p_1$	$p_2$	$p_3$	$\dots$	$p_k$	$\dots$

由概率的性质容易推得, 任一离散型随机变量的分布律  $\{p_k\}$ , 都具有下述两个基本性质:

$$1^\circ \text{ 非负性 } p_k \geq 0, \quad k=1, 2, \dots; \quad (2-3)$$

$$2^\circ \text{ 归一性 } \sum_{k=1}^{\infty} p_k = 1. \quad (2-4)$$

反过来, 任意一个具有以上两个性质的数列  $\{p_k\}$ , 一定可以作为某一个离散型随机变

量的分布律.

为了直观地表达分布律, 我们还可以作类似图 2-1 所示的分布律图. 图 2-1 中  $x_k$  处垂直于  $x$  轴的线段高度为  $p_k$ , 它表示  $X$  取  $x_k$  的概率值.

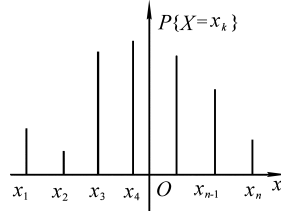


图 2-1

**例 2.1** 设某篮球运动员每次投篮投中的概率为 0.8, 设他在 2 次独立投篮中投中的次数为  $X$ , 求随机变量  $X$  的分布律.

**解** 随机变量  $X$  的可能取值为 0, 1, 2, 记

$$A_i = \{\text{第 } i \text{ 次投篮投中}\}, i=1,2,$$

由假设,  $A_1$  与  $A_2$  相互独立, 记

$$P(A_1) = P(A_2) = p,$$

易知  $X$  的分布律可写为

$$P\{X=0\} = P(\overline{A_1} \overline{A_2}) = (1-p)^2;$$

$$P\{X=1\} = P(A_1 \overline{A_2} \cup \overline{A_1} A_2) = P(A_1 \overline{A_2}) + P(\overline{A_1} A_2) = 2p(1-p);$$

$$P\{X=2\} = P(A_1 A_2) = p^2.$$

$X$  的分布律也可写成如下表 2-2

表 2-2

$X$	0	1	2
$p_k$	$(1-p)^2$	$2p(1-p)$	$p^2$

将  $p=0.8, 1-p=0.2$  代入表 2-2 可得结果如表 2-3 所示.

表 2-3

$X$	0	1	2
$p_k$	0.04	0.32	0.64

下面介绍几种常见的离散型随机变量的概率分布:

### 1. 两点分布

若随机变量  $X$  只可能取  $x_1$  与  $x_2$  两值, 它的分布律是

$$P\{X=x_1\} = 1-p, \quad 0 < p < 1,$$

$$P\{X=x_2\} = p,$$

则称  $X$  服从参数为  $p$  的两点分布.

特别地, 当  $x_1=0, x_2=1$  时两点分布也叫 (0-1) 分布, 记作  $X \sim (0-1)$  分布. 写成分布律表形式如表 2-4 所示.

表 2-4

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

对于一个随机试验, 若它的样本空间只包含两个元素, 即  $\Omega=\{e_1, e_2\}$ , 我们总能在  $\Omega$  上定义一个服从 (0-1) 分布的随机变量

$$X = X(e) = \begin{cases} 0, & \text{当 } e = e_1 \text{ 时,} \\ 1, & \text{当 } e = e_2 \text{ 时,} \end{cases}$$

用它来描述这个试验结果. 因此, 两点分布可以作为描述试验只包含两个基本事件的数学模型. 例如, 在打靶中“命中”与“不中”的概率分布; 产品抽验中“合格品”与“不合格品”的概率分布等. 总之, 一个随机试验如果我们只关心某事件  $A$  出现与否, 则可用一个服从 (0-1) 分布的随机变量来描述.

## 2. 二项分布

若随机变量  $X$  的分布律为

$$P\{X=k\} = C_n^k p^k (1-p)^{n-k}, \quad k=0, 1, \dots, n, \quad (2-5)$$

则称  $X$  服从参数为  $n, p$  的二项分布 (binomial distribution), 记作  $X \sim b(n, p)$ .

易知 (2-5) 式满足 (2-3)、(2-4) 两式. 事实上,  $P(X=k) \geq 0$  是显然的; 再由二项展开式知

$$\sum_{k=0}^n P\{X=k\} = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$$

我们知道,  $P\{X=k\} = C_n^k p^k (1-p)^{n-k}$  恰好是  $[p + (1-p)]^n$  二项展开式中出现  $p^k$  的那一项, 这就是二项分布名称的由来.

回忆  $n$  重伯努利试验中事件  $A$  出现  $k$  次的概率计算公式

$$P_n(k) = C_n^k p^k (1-p)^{n-k}, \quad k=0, 1, \dots, n,$$

可知, 若  $X \sim b(n, p)$ ,  $X$  就可以用来表示  $n$  重伯努利试验中事件  $A$  出现的次数. 因此, 二项分布可以作为描述  $n$  重伯努利试验中事件  $A$  出现次数的数学模型. 例如, 射手射击  $n$  次中, “中的”次数的概率分布; 随机抛掷硬币  $n$  次, 落地时出现“正面”次数的概率分布; 从一批足够多的产品中任意抽取  $n$  件, 其中“废品”件数的概率分布; 等等.

不难看出, (0-1) 分布就是二项分布在  $n=1$  时的特殊情形, 故 (0-1) 分布的分布律也可写成

$$P\{X=k\} = p^k q^{1-k}, \quad k=0, 1; \quad q=1-p.$$

**例 2.2** 某大学的校乒乓球队与数学系乒乓球队举行对抗赛. 校队的实力较系队为强, 当一个校队运动员与一个系队运动员比赛时, 校队运动员获胜的概率为 0.6. 现在校、系双方商量对抗赛的方式, 提出 3 种方案:

- (1) 双方各出 3 人;
- (2) 双方各出 5 人;
- (3) 双方各出 7 人.

3 种方案中均以比赛中得胜人数多的一方为胜利.问: 对系队来说, 哪一种方案有利?

**解** 设系队得胜人数为  $X$ , 则在上述三种方案中, 系队胜利的概率为

$$(1) P\{X \geq 2\} = \sum_{k=2}^3 C_3^k (0.4)^k (0.6)^{3-k} \approx 0.352;$$

$$(2) P\{X \geq 3\} = \sum_{k=3}^5 C_5^k (0.4)^k (0.6)^{5-k} \approx 0.317;$$

$$(3) P\{X \geq 4\} = \sum_{k=4}^7 C_7^k (0.4)^k (0.6)^{7-k} \approx 0.290.$$

因此第一种方案对系队最为有利.这在直觉上是容易理解的, 因为参赛人数越少, 系队侥幸获胜的可能性也就越大.

**例 2.3** 某一大批产品的合格品率为 98%, 现随机地从这批产品中抽样 20 次, 每次抽一个产品, 问抽得的 20 个产品中恰好有  $k$  个 ( $k=1, 2, \dots, 20$ ) 为合格品的概率是多少?

**解** 这是不放回抽样.由于这批产品的总数很大, 而抽出的产品的数量相对于产品总数来说又很小, 那么取出少许几件可以认为并不影响剩下部分的合格品率, 因而可以当作放回抽样来处理, 这样做会有一些误差, 但误差不大.我们将抽检一个产品看其是否为合格品看成一次试验, 显然, 抽检 20 个产品就相当于做 20 次伯努利试验, 以  $X$  记 20 个产品中合格品的个数, 那么  $X \sim b(20, 0.98)$ , 即

$$P\{X=k\} = C_{20}^k (0.98)^k (0.02)^{20-k}, \quad k=1, 2, \dots, 20.$$

若在上例中将参数 20 改为 200 或更大, 显然此时直接计算该概率就显得相当麻烦.为此我们给出一个当  $n$  很大而  $p$  (或  $1-p$ ) 很小时的近似计算公式.

**定理 2.1 (泊松(Poisson)定理)** 设  $np_n = \lambda$  ( $\lambda > 0$  是一常数,  $n$  是任意正整数), 则对任意一固定的非负整数  $k$ , 有

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

**证** 由  $p_n = \lambda/n$ , 有

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

对任意固定的  $k$ , 当  $n \rightarrow \infty$  时,

$$\begin{aligned} \left[ 1 \cdot \left(1-\frac{1}{n}\right) \left(1-\frac{2}{n}\right) \cdots \left(1-\frac{k-1}{n}\right) \right] &\rightarrow 1, \\ \left(1-\frac{\lambda}{n}\right)^n &\rightarrow e^{-\lambda}, \quad \left(1-\frac{\lambda}{n}\right)^{-k} \rightarrow 1 \end{aligned}$$

故

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

由于  $\lambda=np_n$  是常数, 所以当  $n$  很大时  $p_n$  必定很小, 因此, 上述定理表明当  $n$  很大  $p$  很小时, 有以下近似公式

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}, \quad (2-6)$$

其中  $\lambda=np$ .

从表 2-5 可以直观地看出 (2-6) 式两端的近似程度.

表 2-5

$k$	按二项分布公式直接计算				按泊松近似公式 (2.6) 计算
	$n=10$ $p=0.1$	$n=20$ $p=0.05$	$n=40$ $p=0.025$	$n=100$ $p=0.01$	$\lambda=1(=np)$
0	0.349	0.358	0.363	0.366	0.368
1	0.385	0.377	0.372	0.370	0.368
2	0.194	0.189	0.186	0.185	0.184
3	0.057	0.060	0.060	0.061	0.061
4	0.011	0.013	0.014	0.015	0.015
...	...	...	...	...	...

由上表可以看出, 两者的结果是很接近的. 在实际计算中, 当  $n \geq 20, p \leq 0.05$  时近似效果

颇佳, 而当  $n \geq 100, np \leq 10$  时效果更好.  $\frac{\lambda^k e^{-\lambda}}{k!}$  的值有表可查 (见本书附表 3)

二项分布的泊松近似, 常常被应用于研究稀有事件 (即每次试验中事件  $A$  出现的概率  $p$  很小), 当伯努利试验的次数  $n$  很大时, 事件发生的频数的分布.

**例 2.4** 某十字路口有大量汽车通过, 假设每辆汽车在这里发生交通事故的概率为 0.001, 如果每天有 5000 辆汽车通过这个十字路口, 求发生交通事故的汽车数不少于 2 的概率.

**解** 设  $X$  表示“发生交通事故的汽车数”, 则  $X \sim b(n, p)$ , 此处  $n=5000, p=0.001$ , 令  $\lambda=np=5$ ,

$$\begin{aligned} P\{X \geq 2\} &= 1 - P\{X < 2\} = 1 - \sum_{k=0}^1 P\{X = k\} \\ &= 1 - (0.999)^{5000} - 5(0.999)^{4999} \\ &\approx 1 - \frac{5^0 e^{-5}}{0!} - \frac{5e^{-5}}{1!}. \end{aligned}$$

查表可得

$$P\{X \geq 2\} = 1 - 0.00674 - 0.03369 = 0.95957.$$

**例 2.5** 某人进行射击, 设每次射击的命中率为 0.02, 独立射击 400 次, 试求至少击中两次的概率.

**解** 将一次射击看成是一次试验. 设击中次数为  $X$ , 则  $X \sim b(400, 0.02)$ , 即  $X$  的分布律为

$$P\{X=k\} = C_{400}^k (0.02)^k (0.98)^{400-k}, \quad k=0, 1, \dots, 400.$$

故所求概率为

$$\begin{aligned}P\{X \geq 2\} &= 1 - P\{X=0\} - P\{X=1\} \\&= 1 - (0.98)^{400} - 400(0.02)(0.98)^{399} \\&\approx 0.9972.\end{aligned}$$

这个概率很接近 1, 我们从两方面来讨论这一结果的实际意义. 其一, 虽然每次射击的命中率很小 (为 0.02), 但如果射击 400 次, 则击中目标至少两次是几乎可以肯定的. 这一事实说明, 一个事件尽管在一次试验中发生的概率很小, 但只要试验次数很多, 而且试验是独立进行的, 那么这一事件的发生几乎是肯定的. 这也告诉人们决不能轻视小概率事件. 其二, 如果在 400 次射击中, 击中目标的次数竟不到两次, 由于  $P\{X < 2\} \approx 0.003$  很小, 根据实际推断原理, 我们将怀疑“每次射击的命中率为 0.02”这一假设, 即认为该射手射击的命中率达不到 0.02.

### 3. 泊松分布

若随机变量  $X$  的分布律为

$$P\{X=k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0, 1, 2, \dots, \quad (2-7)$$

其中  $\lambda > 0$  是常数, 则称  $X$  服从参数为  $\lambda$  的泊松分布 (poisson distribution), 记为  $X \sim P(\lambda)$ .

易知 (2-7) 式满足 (2-3)、(2-4) 两式, 事实上, 显然  $P\{X=k\} \geq 0$ ; 再由

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

可知

$$\sum_{k=0}^{\infty} P\{X=k\} = 1.$$

由泊松定理可知, 泊松分布可以作为描述大量试验中稀有事件出现的次数  $k=0, 1, \dots$  的概率分布情况的一个数学模型. 例如, 大量产品中抽样检查时得到的不合格品数; 一个集团中员工生日是元旦的人数; 一页中印刷错误出现的数目; 数字通讯中传输数字时发生误码的个数等, 都近似服从泊松分布. 除此之外, 理论与实践都说明, 一般说来它也可作为下列随机变量的概率分布的数学模型: 在任给一段固定的时间间隔内, ① 由某块放射性物质放射出的经过计数器  $\alpha$  粒子; ② 某地区发生交通事故的次数; ③ 来到某公共设施要求给予服务的顾客数 (这里的公共设施的意义可以是极为广泛的, 诸如售货员、机场跑道、电话交换台、医院等, 在机场跑道的例子中, 顾客可以相应地想象为飞机). 泊松分布是概率论中一种很重要的分布.

**例 2.6** 由某商店过去的销售记录知道, 某种商品每月的销售数可以用参数  $\lambda = 5$  的泊松分布来描述. 为了以 95% 以上的把握保证不脱销, 问商店在月底至少应进该种商品多少件?

**解** 设该商店每月销售这种商品数为  $X$ , 月底进货为  $a$  件, 则为了以 95% 以上的把握保证不脱销, 应有

$$P\{X \leq a\} \geq 0.95.$$

由于  $X \sim P(5)$ , 因此上式即为

$$\sum_{k=0}^a \frac{e^{-5} 5^k}{k!} \geq 0.95.$$

查表可知

$$\sum_{k=0}^8 \frac{e^{-5} 5^k}{k!} \approx 1 - 0.0681 = 0.9319 < 0.95,$$

$$\sum_{k=0}^9 \frac{e^{-5} 10^k}{k!} \approx 1 - 0.0318 = 0.9682 > 0.95$$

于是, 这家商店只要在月底进货这种商品 9 件 (假定上个月没有存货), 就可以 95% 以上的把握保证这种商品在下个月不会脱销.

下面我们就一般的离散型随机变量讨论其分布函数. 设离散型随机变量  $X$  的分布律如表 2-1 所示, 由分布函数的定义, 可知

$$F(x) = P\{X \leq x\} = \sum_{x_k \leq x} P\{X = x_k\} = \sum_{x_k \leq x} p_k,$$

此处的  $\sum_{x_k \leq x}$  和式表示对所有满足  $x_k \leq x$  的  $k$  求和, 形象地讲就是对那些满足  $x_k \leq x$  所对应的  $p_k$  的累加.

**例 2.7** 求例 2.1 中  $X$  的分布函数  $F(x)$ .

**解** 由例 2.1 的分布律知

当  $x < 0$  时,

$$F(x) = P\{X \leq x\} = 0;$$

当  $0 \leq x < 1$  时,

$$F(x) = P\{X \leq x\} = P\{X=0\} = 0.04;$$

当  $1 \leq x < 2$  时,

$$F(x) = P\{X \leq x\} = P(\{X=0\} \cup \{X=1\}) = P\{X=0\} + P\{X=1\} = 0.04 + 0.32 = 0.36;$$

当  $x \geq 2$  时,

$$\begin{aligned} F(x) &= P\{X \leq x\} = P(\{X=0\} \cup \{X=1\} \cup \{X=2\}) \\ &= P\{X=0\} + P\{X=1\} + P\{X=2\} \\ &= 0.04 + 0.32 + 0.64 \\ &= 1. \end{aligned}$$

$$F(x) = P\{X \leq x\} = \begin{cases} 0, & x < 0, \\ 0.04, & 0 \leq x < 1, \\ 0.36, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

$F(x)$  的图形是一条阶梯状右连续曲线, 在  $x=0, 1, 2$  处有跳跃, 其跳跃高度分别为 0.04, 0.32, 0.64, 这条曲线从左至右依次从  $F(x)=0$  逐步升级到  $F(x)=1$ .

对表 2-1 所示的一般的分布律, 其分布函数  $F(x)$  表示一条阶梯状右连续曲线, 在  $X=x_k (k=1, 2, \dots)$  处有跳跃, 跳跃的高度恰为  $p_k = P\{X=x_k\}$ , 从左至右, 由水平直线  $F(x)=0$ , 分别按阶高  $p_1, p_2, \dots$  升至水平直线  $F(x)=1$ .

以上是已知分布律求分布函数. 反过来, 若已知离散型随机变量  $X$  的分布函数  $F(x)$ , 则  $X$  的分布律也可由分布函数所确定:

$$p_k = P\{X=x_k\} = F(x_k) - F(x_{k-1}).$$



### 第三节 连续型随机变量及其分布

上一节我们研究了离散型随机变量,这类随机变量的特点是它的可能取值及其相对应的概率能被逐个地列出.这一节我们将要研究的连续型随机变量就不具有这样的性质了.连续型随机变量的特点是它的可能取值连续地充满某个区间甚至整个数轴.例如,测量一个工件长度,因为在理论上说这个长度的值  $X$  可以取区间  $(0, +\infty)$  上的任何一个值.此外,连续型随机变量取某特定值的概率总是零(关于这点将在以后说明).例如,抽检一个工件其长度  $X$  丝毫不差刚好是其固定值(如 1.824cm)的事件  $\{X=1.824\}$  几乎是不可能的,应认为  $P\{X=1.824\}=0$ .因此讨论连续型随机变量在某点的概率是毫无意义的.于是,对于连续型随机变量就不能用对离散型随机变量那样的方法进行研究了.为了方便论述,我们先来看一个例子.

**例 2.8** 一个半径为 2 米的圆盘靶,设击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比,并设射击都能中靶,以  $X$  表示弹着点与圆心的距离,试求随机变量  $X$  的分布函数.

**解** (1) 若  $x < 0$ , 因为事件  $\{X \leq x\}$  是不可能事件, 所以

$$F(x) = P\{X \leq x\} = 0.$$

(2) 若  $0 \leq x \leq 2$ , 由题意  $P\{0 \leq X \leq x\} = kx^2$ ,  $k$  是常数, 为了确定  $k$  的值, 取  $x=2$ , 有  $P\{0 \leq X \leq 2\} = 2^2k$ , 但事件  $\{0 \leq X \leq 2\}$  是必然事件, 故  $P\{0 \leq X \leq 2\} = 1$ , 即  $2^2k = 1$ , 所以  $k = 1/4$ , 即

$$P\{0 \leq X \leq x\} = x^2/4.$$

于是

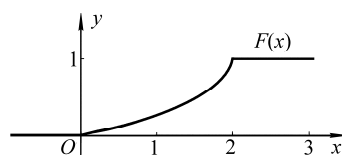
$$F(x) = P\{X \leq x\} = P\{X < 0\} + P\{0 \leq X \leq x\} = x^2/4.$$

(3) 若  $x \geq 2$ , 由于  $\{X \leq 2\}$  是必然事件, 于是

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

综上所述

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{4}x^2, & 0 \leq x < 2, \\ 1, & x \geq 2, \end{cases}$$



它的图形是一条连续曲线, 如图 2-2 所示.

图 2-2

另外, 容易看到本例中  $X$  的分布函数  $F(x)$  还可写成如下形式:

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

其中

$$f(t) = \begin{cases} \frac{1}{2}t, & 0 < t < 2, \\ 0, & \text{其他.} \end{cases}$$

这就是说  $F(x)$  恰好是非负函数  $f(t)$  在区间  $(-\infty, x]$  上的积分, 我们称这种随机变量  $X$  为连续型随机变量.一般地有如下定义.

**定义 2.3** 若对随机变量  $X$  的分布函数  $F(x)$ , 存在非负函数  $f(x)$ , 使对于任意实数  $x$ , 有

$$F(x) = \int_{-\infty}^x f(t)dt, \quad (2-8)$$

则称  $X$  为连续型随机变量, 其中  $f(x)$  称为  $X$  的概率密度函数, 简称概率密度或密度函数 (density function).

由 (2-8) 式可知连续型随机变量  $X$  的分布函数  $F(x)$  是连续函数. 由分布函数的性质  $F(-\infty)=0$ ,  $F(+\infty)=1$  及  $F(x)$  单调不减, 知  $F(x)$  是一条位于直线  $y=0$  与  $y=1$  之间的单调不减的连续 (但不一定光滑) 曲线.

由定义 2.3 知道,  $f(x)$  具有以下性质:

$$1^\circ f(x) \geq 0;$$

$$2^\circ \int_{-\infty}^{+\infty} f(x)dx = 1;$$

$$3^\circ P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx \quad (x_1 \leq x_2);$$

$$4^\circ \text{ 若 } f(x) \text{ 在 } x \text{ 点处连续, 则有 } F'(x) = f(x).$$

由性质  $2^\circ$  可知, 介于曲线  $y=f(x)$  与  $y=0$  之间的面积为 1; 由性质  $3^\circ$  可知,  $X$  落在区间  $(x_1, x_2]$  内的概率  $P\{x_1 < X \leq x_2\}$  等于区间  $(x_1, x_2]$  上曲线  $y=f(x)$  之下曲边梯形的面积. 由性质  $4^\circ$  可知, 在  $f(x)$  的连续点  $x$  处有

$$f(x) = \lim_{\Delta x \rightarrow 0^+} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0^+} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x}.$$

这种形式恰与物理学中线密度定义相类似, 这也正是为什么称  $f(x)$  为概率密度的原因. 同样我们也要指出, 反过来, 任一满足以上  $1^\circ, 2^\circ$  两个性质的函数  $f(x)$ , 一定可以作为某个连续型随机变量的密度函数.

前面我们曾指出对连续型随机变量  $X$  而言, 取任一特定值  $a$  的概率为零, 即  $P\{X=a\}=0$ , 事实上, 令  $\Delta x > 0$ , 设  $X$  的分布函数为  $F(x)$ , 则由

$$\{X=a\} \subset \{a - \Delta x < X \leq a\},$$

得

$$0 \leq P\{X=a\} \leq P\{a - \Delta x < X \leq a\} = F(a) - F(a - \Delta x).$$

由于  $F(x)$  连续, 因此

$$\lim_{\Delta x \rightarrow 0} F(a - \Delta x) = F(a).$$

当  $\Delta x \rightarrow 0$  时, 由夹逼定理, 得

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

由此很容易推导出

$$P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} = P\{a < X < b\}.$$

即在计算连续型随机变量落在某区间上的概率时, 可不必区分该区间端点的情况. 此外还要说明的是, 事件  $\{X=a\}$  “几乎不可能发生”, 但并不保证绝不会发生, 它是“零概率事件”而不是不可能事件.

**例 2.9** 设连续型随机变量  $X$  的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ Ax^2, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

试求:

- (1) 系数  $A$ ;  
 (2)  $X$  落在区间  $(0.3, 0.7)$  内的概率;  
 (3)  $X$  的密度函数.

**解** (1) 由于  $X$  为连续型随机变量, 故  $F(x)$  是连续函数, 因此有

$$1 = F(1) = \lim_{x \rightarrow 1^-} F(x) = \lim_{x \rightarrow 1^-} Ax^2 = A,$$

即  $A=1$ , 于是有

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

$$(2) P\{0.3 < X < 0.7\} = F(0.7) - F(0.3) = (0.7)^2 - (0.3)^2 = 0.4;$$

(3)  $X$  的密度函数为

$$f(x) = F'(x) = \begin{cases} 2x, & 0 \leq x < 1; \\ 0, & \text{其他.} \end{cases}$$

由定义 2.3 知, 改变密度函数  $f(x)$  在个别点的函数值, 不影响分布函数  $F(x)$  的取值, 因此, 并不在乎改变密度函数在个别点上的值 (比如在  $x=0$  或  $x=1$  上  $f(x)$  的值).

**例 2.10** 设随机变量  $X$  具有密度函数

$$f(x) = \begin{cases} kx, & 0 \leq x < 3, \\ 2 - \frac{x}{2}, & 3 \leq x \leq 4, \\ 0, & \text{其他.} \end{cases}$$

- (1) 确定常数  $k$ ;  
 (2) 求  $X$  的分布函数  $F(x)$ ;  
 (3) 求  $P\{1 < X \leq \frac{7}{2}\}$ .

**解** (1) 由  $\int_{-\infty}^{\infty} f(x) dx = 1$ , 得

$$\int_0^3 kx dx + \int_3^4 (2 - \frac{x}{2}) dx = 1,$$

解得  $k=1/6$ , 故  $X$  的密度函数为

$$f(x) = \begin{cases} \frac{x}{6}, & 0 \leq x < 3, \\ 2 - \frac{x}{2}, & 3 \leq x \leq 4, \\ 0, & \text{其他.} \end{cases}$$

$$(2) \text{ 当 } x < 0 \text{ 时, } F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt = 0;$$

$$\text{当 } 0 \leq x < 3 \text{ 时, } F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt = \int_{-\infty}^0 f(t) dt + \int_0^x f(t) dt = \int_0^x \frac{t}{6} dt = \frac{x^2}{12};$$

$$\text{当 } 3 \leq x < 4 \text{ 时, } F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t) dt = \int_{-\infty}^0 f(t) dt + \int_0^3 f(t) dt + \int_3^x f(t) dt$$

$$= \int_0^3 \frac{t}{6} dt + \int_3^x (2 - \frac{t}{2}) dt = -\frac{x^2}{4} + 2x - 3;$$

$$\begin{aligned} \text{当 } x \geq 4 \text{ 时, } F(x) &= P\{X \leq x\} = \int_{-\infty}^x f(t) dt = \int_{-\infty}^0 f(t) dt + \int_0^3 f(t) dt + \int_3^4 f(t) dt + \int_4^x f(t) dt \\ &= \int_0^3 \frac{t}{6} dt + \int_3^4 (2 - \frac{t}{2}) dt = 1. \end{aligned}$$

即

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{12}, & 0 \leq x < 3, \\ -\frac{x^2}{4} + 2x - 3, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

$$(3) \quad P\{1 < X \leq 7/2\} = F(7/2) - F(1) = 41/48.$$

下面介绍 3 种常见的连续型随机变量.

### 1. 均匀分布

若连续型随机变量  $X$  具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{其他.} \end{cases} \quad (2-9)$$

则称  $X$  在区间  $(a, b)$  上服从均匀分布 (uniform distribution), 记为  $X \sim U(a, b)$ .

易知  $f(x) \geq 0$  且

$$\int_{-\infty}^{\infty} f(x) dx = \int_a^b \frac{1}{b-a} dx = 1.$$

由 (2-9) 式可得

$$1^\circ \quad P\{X \geq b\} = \int_b^{\infty} 0 dx = 0, \quad P\{X \leq a\} = \int_{-\infty}^a 0 dx = 0,$$

即

$$P\{a < X < b\} = 1 - P\{X \geq b\} - P\{X \leq a\} = 1;$$

2° 若  $a \leq c < d \leq b$ , 则

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

因此, 在区间  $(a, b)$  上服从均匀分布的随机变量  $X$  的物理意义是:  $X$  以概率 1 在区间  $(a, b)$  内取值, 而以概率 0 在区间  $(a, b)$  以外取值, 并且  $X$  值落入  $(a, b)$  中任一子区间  $(c, d)$  中的概率与子区间的长度成正比, 而与子区间的位置无关.

由 (2-8) 式易得  $X$  的分布函数为

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases} \quad (2-10)$$

密度函数  $f(x)$  和分布函数  $F(x)$  的图形分别如图 2-3 和图 2-4 所示.

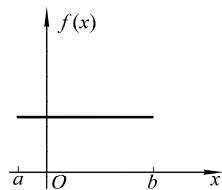


图 2-3

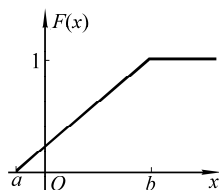


图 2-4

在数值计算中, 由于四舍五入, 小数点后第一位小数所引起的误差  $X$  一般可以看作是一个在  $[-0.5, 0.5]$  上服从均匀分布的随机变量; 又如在  $(a, b)$  中随机掷质点, 则该质点的坐标  $X$  一般也可看做是一个在  $(a, b)$  上服从均匀分布的随机变量.

**例 2.11** 某公共汽车站从上午 7:00 开始, 每 15min 来一辆车, 如某乘客到达此站的时间是 7:00—7:30 之间的服从均匀分布的随机变量, 试求他等车少于 5min 的概率.

**解** 设乘客于 7:00 过  $X$  min 到达车站, 由于  $X$  在  $[0, 30]$  上服从均匀分布, 即有

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

显然, 只有乘客在 7:10—7:15 之间或 7:25—7:30 之间到达车站时, 他(或她)等车的时间才少于 5min, 因此所求概率为

$$P\{10 < X \leq 15\} + P\{25 < X \leq 30\} = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = 1/3.$$

## 2. 指数分布

若随机变量  $X$  的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (2-11)$$

其中  $\lambda > 0$  为常数, 则称  $X$  服从参数为  $\lambda$  的指数分布(exponentially distribution), 记作  $X \sim E(\lambda)$ .

显然  $f(x) \geq 0$ , 且  $\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1$ .

容易得到  $X$  的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

指数分布最常见的一个场合是寿命分布. 指数分布具有“无记忆性”, 即对于任意  $s, t > 0$ , 有

$$P\{X > s+t | X > s\} = P\{X > t\}. \quad (2-12)$$

如果用  $X$  表示某一元件的寿命, 那么上式表明, 在已知元件已使用了  $s$  h 的条件下, 它还能再使用至少  $t$  h 的概率, 与从开始使用时算起它至少能使用  $t$  h 的概率相等. 这就是说该元件对它已使用过  $s$  h 没有记忆. 当然, 指数分布描述的是无老化时的寿命分布, 但“无老化”

是不可能的,因而只是一种近似.对一些寿命长的元件,在初期阶段老化现象很小,在这一阶段,指数分布比较确切地描述了其寿命分布情况.

(2-12) 式是容易证明的.事实上,

$$\begin{aligned} P\{X > s+t | X > s\} &= \frac{P\{X > s, X > s+t\}}{P\{X > s\}} = \frac{P\{X > s+t\}}{P\{X > s\}} \\ &= \frac{1-F(s+t)}{1-F(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P\{X > t\}. \end{aligned}$$

### 3.正态分布

若连续型随机变量  $X$  的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty, \quad (2-13)$$

其中  $\mu, \sigma$  ( $\sigma > 0$ ) 为常数,则称  $X$  服从参数为  $\mu, \sigma$  的正态分布 (normal distribution), 记为  $X \sim N(\mu, \sigma^2)$ .

显然  $f(x) \geq 0$ , 下面来证明  $\int_{-\infty}^{+\infty} f(x) dx = 1$ . 令  $\frac{x-\mu}{\sigma} = t$ , 得到

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt.$$

记  $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$ , 则有

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{t^2+s^2}{2}} dt ds.$$

作极坐标变换:  $s = r \cos \theta, t = r \sin \theta$ , 得到

$$I^2 = \int_{-\infty}^{+\infty} \int_0^{2\pi} r e^{-\frac{r^2}{2}} dr d\theta = 2\pi,$$

而  $I > 0$ , 故有  $I = \sqrt{2\pi}$ , 即有

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}.$$

于是

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1.$$

正态分布是概率论和数理统计中最重要的分布之一.在实际问题中大量的随机变量服从或近似服从正态分布.只要某一个随机变量受到许多相互独立随机因素的影响,而每个个别因素的影响都不能起决定性作用,那么就可以断定随机变量服从或近似服从正态分布.例如,因为人的身高、体重受到种族、饮食习惯、地域、运动等因素影响,但这些因素又不能对身高、体重起决定性作用,所以我们可以认为身高、体重服从或近似服从正态分布.

参数  $\mu, \sigma$  的意义将在第四章中说明.  $f(x)$  的图形如图 2-5 所示,它具有如下性质:

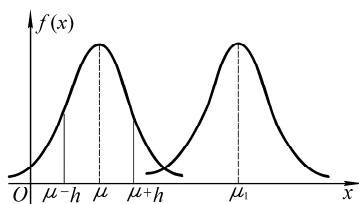


图 2-5

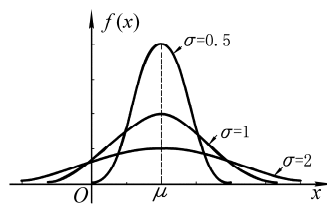


图 2-6

- 1° 曲线关于  $x = \mu$  对称;
- 2° 曲线在  $x = \mu$  处取到最大值,  $x$  离  $\mu$  越远,  $f(x)$  值越小. 这表明对于同样长度的区间, 当区间离  $\mu$  越远,  $X$  落在这个区间上的概率越小;
- 3° 曲线在  $\mu \pm \sigma$  处有拐点;
- 4° 曲线以  $x$  轴为渐近线;
- 5° 若固定  $\mu$ , 当  $\sigma$  越小时图形越尖陡 (见图 2-6), 因而  $X$  落在  $\mu$  附近的概率越大; 若固定  $\sigma$ ,  $\mu$  值改变, 则图形沿  $x$  轴平移, 而不改变其形状. 故称  $\sigma$  为精度参数,  $\mu$  为位置参数.

由 (2-13) 式得  $X$  的分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (2-14)$$

特别地, 当  $\mu=0$ ,  $\sigma=1$  时, 称  $X$  服从标准正态分布  $N(0, 1)$ , 其密度函数和分布函数分别用  $\varphi(x)$ ,  $\Phi(x)$  表示, 即有

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (2-15)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt. \quad (2-16)$$

易知,  $\Phi(-x) = 1 - \Phi(x)$ .

人们已事先编制了  $\Phi(x)$  的函数值表 (见本书附录 2).

一般地, 若  $X \sim N(\mu, \sigma^2)$ , 则有  $\frac{X-\mu}{\sigma} \sim N(0,1)$ .

事实上,  $Z = \frac{X-\mu}{\sigma}$  的分布函数为

$$\begin{aligned} P\{Z \leq x\} &= P\left\{\frac{X-\mu}{\sigma} \leq x\right\} = P\{X \leq \mu + \sigma x\} \\ &= \int_{-\infty}^{\mu + \sigma x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \end{aligned}$$

令  $\frac{t-\mu}{\sigma} = s$ , 得

$$P\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{s^2}{2}} ds = \Phi(x),$$

由此知  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ .

因此, 若  $X \sim N(\mu, \sigma^2)$ , 则可利用标准正态分布函数  $\Phi(x)$ , 通过查表求得  $X$  落在任一区间  $(x_1, x_2]$  内的概率, 即

$$\begin{aligned} P\{x_1 < X \leq x_2\} &= P\left\{\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right\} \\ &= P\left\{\frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right\} - P\left\{\frac{X - \mu}{\sigma} \leq \frac{x_1 - \mu}{\sigma}\right\} \\ &= \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right). \end{aligned}$$

例如, 设  $X \sim N(1.5, 4)$ , 可得

$$\begin{aligned} P\{-1 \leq X \leq 2\} &= P\left\{\frac{-1 - 1.5}{2} \leq \frac{X - 1.5}{2} \leq \frac{2 - 1.5}{2}\right\} \\ &= \Phi(0.25) - \Phi(-1.25) \\ &= \Phi(0.25) - [1 - \Phi(1.25)] \\ &= 0.5987 - 1 + 0.8944 = 0.4931. \end{aligned}$$

设  $X \sim N(\mu, \sigma^2)$ , 由  $\Phi(x)$  函数表可得

$$\begin{aligned} P\{\mu - \sigma < X < \mu + \sigma\} &= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.6826, \\ P\{\mu - 2\sigma < X < \mu + 2\sigma\} &= \Phi(2) - \Phi(-2) = 0.9544, \\ P\{\mu - 3\sigma < X < \mu + 3\sigma\} &= \Phi(3) - \Phi(-3) = 0.9974. \end{aligned}$$

我们看到, 尽管正态变量的取值范围是  $(-\infty, +\infty)$ , 但它的值落在  $(\mu - 3\sigma, \mu + 3\sigma)$  内几乎是肯定的事, 因此在实际问题中, 基本上可以认为有  $|X - \mu| < 3\sigma$ . 这就是人们所说的“ $3\sigma$ 原则”.

**例 2.12** 公共汽车车门的高度是按成年男子与车门顶碰头的机会在 1% 以下来设计的. 设男子身高  $X$  服从  $\mu = 170(\text{cm})$ ,  $\sigma = 6(\text{cm})$  的正态分布, 即  $X \sim N(170, 6^2)$ , 问车门高度应如何确定?

**解** 设车门高度为  $h(\text{cm})$ , 按设计要求  $P\{X \geq h\} \leq 0.01$  或  $P\{X < h\} \geq 0.99$ , 因为  $X \sim N(170, 6^2)$ , 故

$$P\{X < h\} = P\left\{\frac{X - 170}{6} < \frac{h - 170}{6}\right\} = \Phi\left(\frac{h - 170}{6}\right) \geq 0.99,$$

查表得  $\Phi(2.33) = 0.9901 > 0.99$ .

故取  $\frac{h - 170}{6} = 2.33$ , 即  $h = 184$ . 设计车门高度为 184 (cm) 时, 可使成年男子与车门碰头的机会不超过 1%.

**例 2.13** 测量到某一目标的距离时发生的随机误差  $X$  (单位: m) 具有密度函数

$$f(x) = \frac{1}{40\sqrt{2\pi}} e^{-\frac{(x-20)^2}{3200}}.$$

试求在三次测量中至少有一次误差的绝对值不超过 30m 的概率.

**解**  $X$  的密度函数为



$$f(x) = \frac{1}{40\sqrt{2\pi}} e^{-\frac{(x-20)^2}{3200}} = \frac{1}{40 \times \sqrt{2\pi}} e^{-\frac{(x-20)^2}{2 \times 40^2}},$$

即  $X \sim N(20, 40^2)$ , 故一次测量中随机误差的绝对值不超过 30m 的概率为

$$\begin{aligned} P\{|X| \leq 30\} &= P\{-30 \leq X \leq 30\} = \Phi\left(\frac{30-20}{40}\right) - \Phi\left(\frac{-30-20}{40}\right) \\ &= \Phi(0.25) - \Phi(-1.25) = 0.5981 - (1 - 0.8944) = 0.4931. \end{aligned}$$

设  $Y$  为三次测量中误差的绝对值不超过 30m 的次数, 则  $Y$  服从二项分布  $b(3, 0.4931)$ , 故

$$P\{Y \geq 1\} = 1 - P\{Y=0\} = 1 - (0.5069)^3 = 0.8698.$$

为了便于今后应用, 对于标准正态变量, 我们引入了  $\alpha$  分位点的定义.

设  $X \sim N(0, 1)$ , 若  $z_\alpha$  满足条件:

$$P\{X > z_\alpha\} = \alpha, \quad 0 < \alpha < 1, \quad (2-17)$$

则称点  $z_\alpha$  为标准正态分布的上  $\alpha$  分位点, 例如, 由查表可得  $z_{0.05} = 1.645, z_{0.001} = 3.01$ . 故 1.645 与 3.01 分别是标准正态分布的上 0.05 分位点与上 0.001 分位点.

## 第四节 随机变量函数的分布

我们常常遇到一些随机变量, 它们的分布往往难于直接得到 (如测量轴承滚珠体积值  $Y$  等), 但是与它们有函数关系的另一些随机变量, 其分布却是容易知道的 (如滚珠直径测量值  $X$ ). 因此, 要研究随机变量之间的函数关系, 从而通过这种关系由已知的随机变量的分布求出与其有函数关系的另一个随机变量的分布.

**例 2.14** 设随机变量  $X$  具有表 2-6 所示的分布律, 试求  $X^2$  的分布律.

表 2-6

$X$	-1	0	1	1.5	3
$P_k$	0.2	0.1	0.3	0.3	0.1

**解** 由于在  $X$  的取值范围内, 事件 “ $X=0$ ” “ $X=1.5$ ” “ $X=3$ ” 分别与事件 “ $X^2=0$ ” “ $X^2=2.25$ ” “ $X^2=9$ ” 等价, 所以

$$\begin{aligned} P\{X^2=0\} &= P\{X=0\} = 0.1, \\ P\{X^2=2.25\} &= P\{X=1.5\} = 0.3, \\ P\{X^2=9\} &= P\{X=3\} = 0.1. \end{aligned}$$

事件 “ $X^2=1$ ” 是两个互斥事件 “ $X=-1$ ” 及 “ $X=1$ ” 的和, 其概率为这两事件概率和, 即

$$P\{X^2=1\} = P\{X=-1\} + P\{X=1\} = 0.2 + 0.3 = 0.5.$$

于是得  $X^2$  的分布律如表 2-7 所示.

表 2-7

$X^2$	0	1	2.25	9
$p$	0.1	0.5	0.3	0.1

**例 2.15** 设连续型随机变量  $X$  具有概率密度  $f_X(x)$ ,  $-\infty < x < +\infty$ , 求  $Y=g(X)=X^2$  的概率密度.

**解** 先求  $Y$  的分布函数  $F_Y(y)$ , 由于  $Y=g(X)=X^2 \geq 0$ , 因此当  $y \leq 0$  时, 事件 “ $Y \leq y$ ” 的概率为 0, 即  $F_Y(y) = P\{Y \leq y\} = 0$ , 当  $y > 0$  时, 有

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx. \end{aligned}$$

将  $F_Y(y)$  关于  $y$  求导, 即得  $Y$  的概率密度为

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

例如, 当  $X \sim N(0, 1)$ , 其概率密度为 (2-15) 式, 则  $Y=X^2$  的概率密度为

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

此时称  $Y$  服从自由度为 1 的  $\chi^2$  分布.

上例中关键的一步在于将事件 “ $Y \leq y$ ” 由其等价事件 “ $-\sqrt{y} \leq X \leq \sqrt{y}$ ” 代替, 即将事件 “ $Y \leq y$ ” 转换为有关  $X$  的范围所表示的等价事件. 下面我们仅对  $Y=g(X)$ , 其中  $g(x)$  为严格单调函数, 写出一般结论.

**定理 2.2** 设随机变量  $X$  具有概率密度  $f_X(x)$ ,  $-\infty < x < +\infty$ , 又设函数  $g(x)$  处处可导且  $g'(x) > 0$  (或  $g'(x) < 0$ ), 则  $Y=g(X)$  是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X[h(y)]|h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他.} \end{cases} \quad (2-18)$$

其中  $\alpha = \min(g(-\infty), g(+\infty))$ ,  $\beta = \max(g(-\infty), g(+\infty))$ ,  $h(y)$  是  $g(x)$  的反函数.

我们只证  $g'(x) > 0$  的情况. 由于  $g'(x) > 0$ , 因此  $g(x)$  在  $(-\infty, +\infty)$  上严格单调递增, 它的反函数  $h(y)$  存在, 且在  $(\alpha, \beta)$  上严格单调递增、可导. 我们先求  $Y$  的分布函数  $F_Y(y)$ , 再通过求  $F_Y(y)$  求导求出  $f_Y(y)$ .

由于  $Y=g(X)$  在  $(\alpha, \beta)$  上取值, 因此

当  $y \leq \alpha$  时,  $F_Y(y) = P\{Y \leq y\} = 0$ ;

当  $y \geq \beta$  时,  $F_Y(y) = P\{Y \leq y\} = 1$ ;

当  $\alpha < y < \beta$  时,

$$F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq h(y)\} = \int_{-\infty}^{h(y)} f_X(x) dx.$$

于是得概率密度

$$f_Y(y) = \begin{cases} f_X[h(y)]h'(y), & \alpha < y < \beta, \\ 0, & \text{其他.} \end{cases}$$

对于  $g'(x) < 0$  的情况可以同样证明, 即

$$f_Y(y) = \begin{cases} f_X[h(y)][-h'(y)], & \alpha < y < \beta, \\ 0, & \text{其他.} \end{cases}$$

将上面两种情况合并得

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & \alpha < y < \beta, \\ 0, & \text{其他.} \end{cases}$$

注: 若  $f(x)$  在  $[a, b]$  之外为零, 则只需假设在  $(a, b)$  上恒有  $g'(x) > 0$  (或恒有  $g'(x) < 0$ ), 此时

$$\alpha = \min\{g(a), g(b)\}, \quad \beta = \max\{g(a), g(b)\}.$$

**例 2.16** 设随机变量  $X \sim N(\mu, \sigma^2)$ . 试证明  $X$  的线性函数  $Y = aX + b$  ( $a \neq 0$ ) 也服从正态分布.

**证** 设  $X$  的概率密度

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

再令  $y = g(x) = ax + b$ , 得  $g(x)$  的反函数

$$x = h(y) = \frac{y - b}{a}.$$

所以  $h'(y) = 1/a$ .

由 (2-18) 式, 得  $Y = g(X) = aX + b$  的概率密度为

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right), \quad -\infty < y < +\infty,$$

即

$$f_Y(y) = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{[y - (b + a\mu)]^2}{2(a\sigma)^2}}, \quad -\infty < y < +\infty,$$

即有

$$Y = aX + b \sim N(a\mu + b, (a\sigma)^2).$$

**例 2.17** 由统计物理学知分子运动速度的绝对值  $X$  服从麦克斯韦 (Maxwell) 分布, 其概率密度为

$$f(x) = \begin{cases} \frac{4x^2}{a^3\sqrt{\pi}} e^{-\frac{x^2}{a^2}}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

其中  $a > 0$  为常数, 求分子动能  $Y = \frac{1}{2}mX^2$  ( $m$  为分子质量) 的概率密度.

**解** 已知  $y = g(x) = \frac{1}{2}mx^2$ ,  $f(x)$  只在区间  $(0, +\infty)$  上非零且  $g'(x)$  在此区间恒单调递增,

由 (2-18) 式, 得  $Y$  的概率密度为

$$\psi(y) = \begin{cases} \frac{4\sqrt{2y}}{m^{3/2}a^3\sqrt{\pi}} e^{-\frac{2y}{ma^2}}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

## 小 结

随机变量  $X=X(e)$  是定义在样本空间  $\Omega=\{e\}$  上的实值单值函数, 它的取值随试验结果而定, 是不能预先确定的, 且它的取值有一定的概率, 因而它与普通函数是不同的. 引入随机变量, 就可以用微积分的理论和方法对随机试验与随机事件的概率进行数学推理与计算, 从而完成对随机试验结果的规律性的研究.

分布函数

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

反映了随机变量  $X$  的取值不大于实数  $x$  的概率. 随机变量  $X$  落入实轴上任意区间  $(x_1, x_2]$  上的概率也可用  $F(x)$  来表示, 即

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

因此掌握了随机变量  $X$  的分布函数, 就了解了随机变量  $X$  在  $(-\infty, \infty)$  上的概率分布, 可以说分布函数完整地描述了随机变量的统计规律性.

本书只讨论了两类重要的随机变量. 一类是离散型随机变量. 对于离散型随机变量, 我们需要知道它可能取哪些值, 以及它取每个可能值的概率, 常用分布律

$$P\{X=x_k\}=p_k, \quad k=1, 2, \dots$$

或用表 2-8 表示它取值的统计规律性. 要掌握已知分布律求分布函数  $F(x)$  的方法以及已知分布函数  $F(x)$  求分布律的方法. 分布律与分布函数是一一对应的.

表 2-8

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

另一类是连续型随机变量, 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

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

则称  $X$  是连续型随机变量, 其中  $f(x)$  称为  $X$  的概率密度函数. 连续型随机变量的分布函数是连续的, 但不能认为凡是分布函数为连续函数的随机变量就是连续型随机变量. 判别一个随机变量是不是连续型的, 要看符合定义条件的  $f(x)$  是否存在 (事实上存在分布函数  $F(x)$  连续, 但又不能以非负函数的变上限的定积分表示的随机变量).

要掌握已知  $f(x)$  求  $F(x)$  的方法, 以及已知  $F(x)$  求  $f(x)$  的方法. 由连续型随机变量定义可知, 改变  $f(x)$  在个别点的函数值, 并不改变  $F(x)$  的值, 因此改变  $f(x)$  在个别点的值是无关紧要的.

读者要掌握分布函数、分布律、密度函数的性质.

本章还介绍了几种重要的随机变量的分布: (0-1) 分布、二项分布、泊松分布、均匀分布、指数分布、正态分布. 读者必须熟练掌握这几种分布的分布律或密度函数, 还须知道每一种分布的概率意义, 对这几种分布的理解不能仅限于知道它们的分布律或密度函数.

随机变量  $X$  的函数  $Y=g(X)$  也是一个随机变量. 求  $Y$  的分布时, 首先要准确界定  $Y$  的取值范围 (在离散型时要注意相同值的合并), 其次要正确计算  $Y$  的分布, 特别是当  $Y$  为连续型随机变量时的情形. 当  $y=g(x)$  单调或分段单调时, 可按定理写出  $Y$  的密度函数  $f_Y(y)$ , 否则应先按分布函数定义求出  $F_Y(y)$ , 再对  $y$  求导, 得到  $f_Y(y)$  (即使是当  $y=g(x)$  单调或分段单调时, 也应掌握先求出  $F_Y(y)$ , 再求出  $f_Y(y)$  的一般方法).

### 重要术语及主题

随机变量	分布函数	离散型随机变量及其分布律
连续型随机变量及其密度函数	(0-1) 分布	

二项分布	泊松分布	均匀分布
指数分布	正态分布	随机变量函数的分布

## 习 题 二

1. 一袋中有 5 只乒乓球, 编号为 1, 2, 3, 4, 5, 在其中同时取 3 只, 以  $X$  表示取出的 3 只球中的最大号码, 写出随机变量  $X$  的分布律.

2. 设在 15 只同类型零件中有 2 只为次品, 在其中取 3 次, 每次任取 1 只, 作不放回抽样, 以  $X$  表示取出的次品个数, 求:

(1)  $X$  的分布律;

(2)  $X$  的分布函数并作图;

(3)  $P\{X \leq \frac{1}{2}\}$ ,  $P\{1 < X \leq \frac{3}{2}\}$ ,  $P\{1 \leq X \leq \frac{3}{2}\}$ ,  $P\{1 < X < 2\}$ .

3. 射手向目标独立地进行了 3 次射击, 每次击中率为 0.8, 求 3 次射击中击中目标的次数的分布律及分布函数, 并求 3 次射击中至少击中 2 次的概率.

4. (1) 设随机变量  $X$  的分布律为

$$P\{X=k\} = a \frac{\lambda^k}{k!},$$

其中  $k=0, 1, 2, \dots$ ,  $\lambda > 0$  为常数, 试确定常数  $a$ .

(2) 设随机变量  $X$  的分布律为

$$P\{X=k\} = a/N, \quad k=1, 2, \dots, N,$$

试确定常数  $a$ .

5. 甲、乙两人投篮, 投中的概率分别为 0.6, 0.7, 今各投 3 次, 求:

(1) 两人投中次数相等的概率;

(2) 甲比乙投中次数多的概率.

6. 设某机场每天有 200 架飞机在此降落, 任一飞机在某一时刻降落的概率设为 0.02, 且设各飞机降落是相互独立的. 试问该机场需配备多少条跑道, 才能保证某一时刻飞机需立即降落而没有空闲跑道的概率小于 0.01 (每条跑道只能允许一架飞机降落)?

7. 有一繁忙的汽车站, 每天有大量汽车通过, 设每辆车在一天的某时段出事故的概率为 0.0001, 在某天的该时段内有 1000 辆汽车通过, 问出事故的次数不小于 2 的概率是多少 (利用泊松定理)?

8. 已知 5 重伯努利试验中成功的次数  $X$  满足:  $P\{X=1\} = P\{X=2\}$ , 求概率  $P\{X=4\}$ .

9. 设事件  $A$  在每一次试验中发生的概率为 0.3, 当  $A$  发生不少于 3 次时, 指示灯发出信号.

(1) 进行了 5 次独立试验, 试求指示灯发出信号的概率;

(2) 进行了 7 次独立试验, 试求指示灯发出信号的概率.

10. 某公安局在长度为  $t$  的时间间隔内收到的紧急呼救的次数  $X$  服从参数为  $\frac{1}{2}t$  的泊松分布, 而与时间间隔起点无关 (时间以 h 计). 求

- (1) 求某一天 12:00-15:00 没收到呼救的概率;  
 (2) 求某一天 12:00-17:00 至少收到 1 次呼救的概率.

11. 设  $P\{X=k\} = C_2^k p^k (1-p)^{2-k}, \quad k=0,1,2,$

$$P\{Y=m\} = C_4^m p^m (1-p)^{4-m}, \quad m=0,1,2,3,4$$

分别为随机变量  $X, Y$  的概率分布, 如果已知  $P\{X \geq 1\} = \frac{5}{9}$ , 试求  $P\{Y \geq 1\}$ .

12. 某教科书印刷了 2000 册, 因装订等原因造成错误的概率为 0.001, 试求在这 2000 册书中恰有 5 册错误的概率.

13. 进行某种试验, 成功的概率为  $\frac{3}{4}$ , 失败的概率为  $\frac{1}{4}$ . 以  $X$  表示试验首次成功所需试验的次数, 试写出  $X$  的分布律, 并计算  $X$  取偶数的概率.

14. 有 2500 名同一年龄和同社会阶层的人参加了保险公司的人寿保险. 在一年中每个人死亡的概率为 0.002, 每个参加保险的人在 1 月 1 日需交 12 元保险费, 而在死亡时家属可从保险公司领取 2000 元赔偿金. 求:

- (1) 保险公司亏本的概率;  
 (2) 保险公司获利分别不少于 10000 元、20000 元的概率.

15. 已知随机变量  $X$  的密度函数为

$$f(x) = Ae^{-|x|}, \quad -\infty < x < +\infty,$$

求: (1)  $A$  值; (2)  $P\{0 < X < 1\}$ ; (3)  $F(x)$ .

16. 设某种仪器内装有 3 只同样的电子管, 电子管使用寿命  $X$  的密度函数为

$$f(x) = \begin{cases} \frac{100}{x^2}, & x \geq 100, \\ 0, & x < 100. \end{cases}$$

求:

- (1) 在开始 150h 内没有电子管损坏的概率;  
 (2) 在 (1) 中这段时间内有一只电子管损坏的概率;  
 (3) 分布函数  $F(x)$ .

17. 在区间  $[0, a]$  上任意投掷一个质点, 以  $X$  表示这质点的坐标, 设这质点落在  $[0, a]$  中任意小区间内的概率与这小区间长度成正比例, 试求  $X$  的分布函数.

18. 设随机变量  $X$  在  $[2, 5]$  上服从均匀分布. 现对  $X$  进行 3 次独立观测, 求至少有两次的观测值大于 3 的概率.

19. 设顾客在某银行的窗口等待服务的时间  $X$  (以 min 计) 服从指数分布  $E(1/5)$ . 某顾客在窗口等待服务, 若超过 10min 他就离开. 他一个月要到银行 5 次, 以  $Y$  表示一个月内他未等到服务而离开窗口的次数, 试写出  $Y$  的分布律, 并求  $P\{Y \geq 1\}$ .

20. 某人乘汽车去火车站乘火车, 有两条路可走. 第一条路程较短但交通拥挤, 所需时间  $X$  服从  $N(40, 10^2)$ ; 第二条路程较长, 但阻塞少, 所需时间  $X$  服从  $N(50, 4^2)$ .

- (1) 若动身时离火车开车只有 1h, 问应走哪条路能乘上火车的把握大些?  
 (2) 又若离火车开车时间只有 45min, 问应走哪条路赶上火车把握大些?

21. 设  $X \sim N(3, 2^2)$ ,

- (1) 求  $P\{2 < X \leq 5\}$ ,  $P\{-4 < X \leq 10\}$ ,  $P\{|X| > 2\}$ ,  $P\{X > 3\}$ ;  
 (2) 确定  $c$ , 使  $P\{X > c\} = P\{X \leq c\}$ .

22. 由某机器生产的螺栓长度 (cm)  $X \sim N(10.05, 0.06^2)$ , 规定长度在  $10.05 \pm 0.12$  内为合格品, 求一螺栓为不合格品的概率.

23. 一工厂生产的电子管寿命  $X$  (h) 服从正态分布  $N(160, \sigma^2)$ , 若要求  $P\{120 < X \leq 200\} \geq 0.8$ , 允许  $\sigma$  最大不超过多少?

24. 设随机变量  $X$  的分布函数为

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

(1) 求常数  $A, B$ ;

(2) 求  $P\{X \leq 2\}, P\{X > 3\}$ ;

(3) 求概率密度  $f(x)$ .

25. 设随机变量  $X$  的概率密度为

$$f(x) = \begin{cases} x, & 0 \leq x < 1, \\ 2-x, & 1 \leq x < 2, \\ 0, & \text{其他.} \end{cases}$$

求  $X$  的分布函数  $F(x)$ , 并画出  $f(x)$  及  $F(x)$ .

26. 设随机变量  $X$  的密度函数为

$$(1) f(x) = \begin{cases} bx, & 0 < x < 1, \\ \frac{1}{x^2}, & 1 \leq x < 2, \\ 0, & \text{其他.} \end{cases}$$

试确定常数  $a, b$ , 并求其分布函数  $F(x)$ .

27. 求标准正态分布的上  $\alpha$  分位点:

(1)  $\alpha=0.01$ , 求  $z_\alpha$ ;

(2)  $\alpha=0.003$ , 求  $z_\alpha, z_{\alpha/2}$ .

28. 设随机变量  $X$  的分布律如表 2-9 所示, 求  $Y=X^2$  的分布律.

表 2-9

$X$	-2	-1	0	1	3
$P_k$	1/5	1/6	1/5	1/15	11/30

29. 设  $P\{X=k\} = \frac{1}{2^k}, k=1, 2, \dots$ , 令

$$Y = \begin{cases} 1, & \text{当 } X \text{ 取偶数时,} \\ -1, & \text{当 } X \text{ 取奇数时.} \end{cases}$$

求随机变量  $X$  的函数  $Y$  的分布律.

30. 设  $X \sim N(0, 1)$ . 求:

(1) 求  $Y=e^X$  的概率密度;

(2) 求  $Y=2X^2+1$  的概率密度;

(3) 求  $Y=|X|$  的概率密度.

31. 设随机变量  $X \sim U(0, 1)$ , 试求:

(1)  $Y=e^X$  的分布函数及密度函数;

(2)  $Z=-2\ln X$  的分布函数及密度函数.

32. 设随机变量  $X$  的密度函数为

$$f(x)=\begin{cases} \frac{2x}{\pi^2}, & 0 < x < \pi, \\ 0, & \text{其他.} \end{cases}$$

试求  $Y=\sin X$  的密度函数.

33. 设随机变量  $X$  的分布函数如下:

$$F(x)=\begin{cases} \frac{1}{1+x^2}, & x < \underline{(1)}, \\ \underline{(2)}, & x \geq \underline{(3)}. \end{cases}$$

试填上(1),(2),(3)项.

34. 同时掷两枚骰子, 直到一枚骰子出现 6 点为止, 求抛掷次数  $X$  的分布律.

35. 随机数字序列要多长才能使数字 0 至少出现一次的概率不小于 0.9?

36. 已知

$$F(x)=\begin{cases} 0, & x < 0, \\ x+\frac{1}{2}, & 0 \leq x < \frac{1}{2}, \\ 1, & x \geq \frac{1}{2}. \end{cases}$$

则  $F(x)$  是 ( ) 随机变量的分布函数.

(A) 连续型

(B) 离散型

(C) 非连续亦非离散型.

$$37. \text{ 设随机变量 } X \text{ 的分布函数 } F(x)=\begin{cases} 0 & x < 0, \\ \frac{1}{2} & 0 \leq x < 1, \text{ 则 } P(X=1) \text{ 等于 } ( ). \\ 1-e^{-x} & x \geq 1, \end{cases}$$

(A) 0

(B)  $\frac{1}{2}$

(C)  $\frac{1}{2}-e^{-1}$

(D)  $1-e^{-1}$

(2010 研考)

38. 设随机变量  $X \sim N(0, \sigma^2)$ , 问: 当  $\sigma$  取何值时,  $X$  落入区间  $(1, 3)$  的概率最大?

39. 设在一段时间内进入某一商店的顾客人数  $X$  服从泊松分布  $P(\lambda)$ , 每个顾客购买某种物品的概率为  $p$ , 并且各个顾客是否购买该种物品相互独立, 求进入商店的顾客购买这种物品的人数  $Y$  的分布律.

40. 设随机变量  $X$  服从参数为 2 的指数分布. 证明:  $Y=1-e^{-2X}$  在区间  $(0, 1)$  上服从均匀分布.

41. 设随机变量  $X$  的密度函数为



$$f(x)=\begin{cases} \frac{1}{3}, & 0 \leq x \leq 1, \\ \frac{2}{9}, & 3 \leq x \leq 6, \\ 0, & \text{其他.} \end{cases}$$

若  $k$  使得  $P\{X \geq k\} = 2/3$ , 求  $k$  的取值范围.

(2000 研考)

42. 设随机变量  $X$  的分布函数为

$$F(x)=\begin{cases} 0, & x < -1, \\ 0.4, & -1 \leq x < 1, \\ 0.8, & 1 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

求  $X$  的概率分布.

43. 设 3 次独立试验中, 事件  $A$  出现的概率相等. 若已知  $A$  至少出现一次的概率为  $19/27$ , 求  $A$  在一次试验中出现的概率.

44. 若随机变量  $X$  在  $(1, 6)$  上服从均匀分布, 则方程  $y^2 + Xy + 1 = 0$  有实根的概率是多少?

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

46. 假设一厂家生产的每台仪器, 以概率 0.7 可以直接出厂; 以概率 0.3 需进一步调试, 经调试后以概率 0.8 可以出厂, 以概率 0.2 定为不合格品不能出厂. 现该厂新生产了  $n(n \geq 2)$  台仪器 (假设各台仪器的生产过程相互独立). 求

- (1) 全部能出厂的概率  $\alpha$ ;
- (2) 其中恰好有两台不能出厂的概率  $\beta$ ;
- (3) 其中至少有两台不能出厂的概率  $\theta$ .

47. 某地抽样调查结果表明, 考生的外语成绩 (百分制) 近似服从正态分布, 平均成绩为 72 分, 96 分以上的占考生总数的 2.3%, 试求考生的外语成绩在 60 分至 84 分之间的概率.

48. 在电源电压不超过 200V、200V~240V 和超过 240V 3 种情形下, 某种电子元件损坏的概率分别为 0.1, 0.001 和 0.2 (假设电源电压  $X$  服从正态分布  $N(220, 25^2)$ ). 试求:

- (1) 该电子元件损坏的概率  $\alpha$ ;
- (2) 该电子元件损坏时, 电源电压在 200~240V 的概率  $\beta$ .

49. 设随机变量  $X$  在区间  $(1, 2)$  上服从均匀分布, 试求随机变量  $Y = e^{2X}$  的概率密度  $f_Y(y)$ .

50. 设随机变量  $X$  的密度函数为

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

求随机变量  $Y = e^X$  的密度函数  $f_Y(y)$ .

51. 设随机变量  $X$  的密度函数为

$$f_X(x) = \frac{1}{\pi(1+x^2)},$$

求  $Y = 1 - \sqrt[3]{X}$  的密度函数  $f_Y(y)$ .

52. 假设一大型设备在任何长为  $t$  的时间内发生故障的次数  $N(t)$  服从参数为  $\lambda t$  的泊松

分布.求:

(1) 求相继两次故障之间时间间隔  $T$  的概率分布;

(2) 求在设备已经无故障工作 8h 的情形下, 再无故障运行 8h 的概率  $Q$ .

53. 设随机变量  $X$  的绝对值不大于 1,  $P\{X=-1\}=1/8$ ,  $P\{X=1\}=1/4$ . 在事件  $\{-1 < X < 1\}$  出现的条件下,  $X$  在  $(-1, 1)$  内任一子区间上取值的条件概率与该子区间长度成正比, 试求  $X$  的分布函数  $F(x) = P\{X \leq x\}$ .

54. 设随机变量  $X$  服从正态分布  $N(\mu_1, \sigma_1^2)$ ,  $Y$  服从正态分布  $N(\mu_2, \sigma_2^2)$ , 且  $P\{|X - \mu_1| < 1\} > P\{|Y - \mu_2| < 1\}$ , 试比较  $\sigma_1$  与  $\sigma_2$  的大小. (2006 研考)

55. 设  $F_1(x)$  与  $F_2(x)$  为两个随机变量的分布函数, 其相应的概率密度  $f_1(x)$  和  $f_2(x)$  是连续函数, 则必为概率密度的是 ( ).

(A)  $f_1(x)f_2(x)$

(B)  $2f_1(x)F_1(x)$

(C)  $f_1(x)F_2(x)$

(D)  $f_1(x)F_2(x) + f_2(x)F_1(x)$  (2011 研考)