第1章

随机事件及其概率

概率论与数理统计是从数量化的角度来研究现实世界中的一类不确定性现象(随机现象)及其规律性的一门应用数学学科,也是理论联系实际最活跃的学科之一.生活中的很多问题都可归结为概率问题,而概率方法在各领域的广泛应用和卓有成效的贡献,几乎使它成为一切专家、学者、实业家手中强有力的工具,它在科学上的地位也越来越被人们重视.本章所介绍的随机事件及其概率是概率论中最基本、最重要的概念之一.

1.1 随机事件

1.1.1 随机现象

在自然界以及生产实践和科学实验中普遍存在着两类现象.一类是在一定条件下重复进行试验,某一结果必然发生或必然不发生,即是可以事先预言的,称为**确定性现象**.例如:

(1) 一口袋中有十只完全相同的红球,从中任取一只必然为红球.

(2) 异性电荷相互吸引,同性电荷相互排斥.

除去确定性现象,人们发现还存在另一类现象,它是事先不可预言的,即在相同条件下重复进行试验,每次的结果不一定相同,称这一类现象为**偶然性现象**或**随机现象**.例如:

(1) 在相同条件下,抛掷一枚质地均匀的硬币,结果可能是正面向上,也可能是

反面向上，每次抛掷之前无法确定结果.

（2）从一批产品中，随机抽取检验，结果可能是合格品，也可能是不合格品. 在每次抽检之前无法预知结果.

是否这些随机现象都没有什么规律可循呢？事实上并非如此. 人们在长期的观察和实践中逐渐发现，所谓的不可预言只是对一次或是少数几次观察或实践而言，当在相同条件下进行大量观察时，这类现象都呈现出某种规律. 例如，均匀的硬币抛掷足够多次，正面向上和反面向上的比例总是近似 1∶1，而且大致上抛掷次数越多，就越接近这个比例. 表 1-1 给出了历史上抛掷硬币实验的记录. 再比如人的高度虽然各不相同，但通过大量的统计，如果在一定范围内按同一高度人的数量占总人数的比例画出“直方图”（x 轴表示人的高度，y 轴表示同一高度的人的数量占总人数的比例），就可以连成一条曲线（见图 1-1）.

表 1-1　历史上抛均匀硬币试验的记录

实 验 者	试验次数（n）	正面向上次数（n_A）	正面向上频率（$\frac{n_A}{n}$）
德摩根（De Morgan）	2 048	1 061	0.518 1
布丰（Buffon）	4 040	2 048	0.506 9
皮尔逊（Pearson）	12 000	6 019	0.501 6
皮尔逊	24 000	12 012	0.500 5

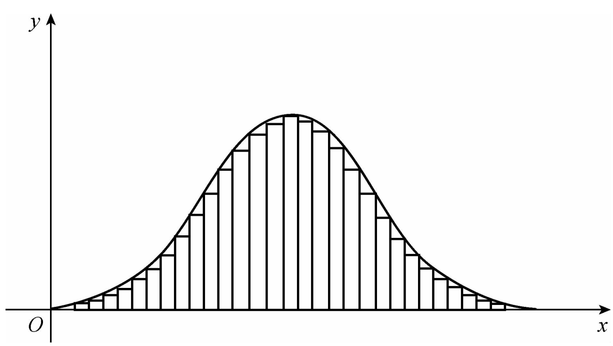


图 1-1

在一定条件下，随机现象有多种可能的结果发生，事先不能预知将出现哪种结果，但通过大量的重复观察，出现的结果会呈现出某种规律，称为随机现象的**统计规律性**.

概率论与数理统计就是研究和揭示随机现象统计规律性的一门数学学科，由于

随机现象的普遍性,概率论与数理统计的理论与方法得到了越来越广泛的应用,其应用几乎遍及所有科学技术领域、工农业生产和国民经济各个部门. 例如,气象预报、水文、地震预报、寻求最佳生产条件的试验设计等. 与其他学科相结合发展了许多边缘学科:生物统计、数学地质、环境数学等. 概率论与数理统计也是一些重要学科的理论基础,如可靠性理论、控制论、人工智能等.

1.1.2　随机试验与样本空间

要对随机现象的统计规律性进行研究,就需要对随机现象进行重复观察. 对随机现象的一次观察称为一个**试验**. 如果这个试验具有如下特点:

(1) **可重复性**　试验可以在相同条件下重复进行.

(2) **可观察性**　每次试验的可能结果不止一个,但能事先明确试验的所有可能结果.

(3) **不确定性**　每次试验前不能确定哪个结果将出现.

称这样的试验为**随机试验**,通常用字母 E 表示. 在实际问题中,也有很多随机现象是不能重复的. 例如,某场球赛的输赢、某个学生参加一次大型考试的成绩等都是不能重复的,对这些现象的观察就不再是一个随机试验了. 本书中以后提到的试验都是指随机试验,就是通过研究随机试验来研究随机现象的.

例 1.1　下列试验都是随机试验.

(1) 考试结束后,某个学生做了这样几个阄:A 表示"90 分及 90 分以上",B 表示"$80\sim89$ 分",C 表示"$70\sim79$ 分",D 表示"$60\sim69$ 分",E 表示"不及格",从中抓一个,观察出现的结果.

(2) 掷两颗骰子, 观察骰子朝上的点数.

(3) 3 件产品中 2 件正品,1 件次品,① 依次取出 2 件;② 同时取出 2 件,观察结果.

(4) 从一批电脑中,任取一台,① 观察无故障运行的时间 t;② 若电脑无故障运行 1 000 小时以上为合格品,否则为不合格品,观察取出的电脑是合格品还是不合格品.

(5) 向坐标平面区域 D:$x^2+y^2\leqslant 100$ 内随机投掷一点 M(设点必落在 D 上),观察点 M 的坐标.

随机试验 E 的每一种可能的结果称为一个**样本点**,它们的全体称为 E 的**样本空间**,记为 S(或 Ω).

例 1.2　写出例 1.1 中各随机试验的样本空间.

(1) $S=\{A,B,C,D,E\}$.

(2) $S=\{(i,j)\mid i,j=1,2,3,4,5,6\}$,其中$(i,j)$表示第一颗骰子朝上的点数为 i,第二颗骰子朝上的点数为 j.

(3) ① $S_1=\{$(正品,次品),(正品,正品),(次品,正品)$\}$;② $S_2=\{$(正品,次

品),(正品,正品)}.

若用"1"表示"正品","0"表示"次品",这里的两个样本空间又可表示为

① $S_1 = \{(1,0),(1,1),(0,1)\}$;② $S_2 = \{(1,0),(1,1)\}$.

(4) ① $S_1 = \{t \mid t \geqslant 0\}$;② $S_2 = \{$合格品,不合格品$\}$.

若用"1"表示"合格品","0"表示"不合格品",S_2 又可表示为 $S_2 = \{1,0\}$.

(5) $S_5 = \{(x,y) \mid x^2 + y^2 \leqslant 100\}$.

由上例可知,样本空间可以是有限集,也可以是无限集.

注:① 对于同一个随机试验,试验的样本点和样本空间可能不一样,这要根据需要观察的内容来确定,如上例中的(3)和(4).

② 有时,同一个样本空间可概括各种实际内容完全不同的问题.例如,只包含两个样本点的样本空间 $S = \{1,0\}$,既可以作为抛硬币出现正面或反面的模型,也可表示学生成绩及格或不及格的模型,还可表示产品验收中合格与不合格的模型等.

1.1.3 随机事件

在进行随机试验时,人们除了关心试验的结果,常常还关心试验的结果是否具备某种指定的可观察的特性.例如,在例 1.1(1) 中,抓阄者关心的是抓到的阄能否表达"考试及格"这一信息,即抓到的阄是否是 $\{A,B,C,D\}$ 中的一个.随机试验中,类似这样的样本空间的子集,称之为**随机事件**,简称**事件**.通常用大写字母 $A,B,C,\cdots$ 表示.在一次试验中,当且仅当这一子集中的某个样本点出现时,称这一**事件发生**.随机事件是概率论研究的主要对象.特别地,

(1) **基本事件** 由样本空间 S 中单个样本点组成的单点集称为**基本事件**,常用 e 或 ω 表示.

(2) **复合事件** 由样本空间 S 中两个或两个以上样本点组成的集合称为**复合事件**.

(3) **必然事件** 样本空间 S 作为自身的子集,它在每次试验中必然发生,称为**必然事件**,用 S 表示.

(4) **不可能事件** 样本空间 S 的最小子集(即空集),它在每次试验中必然不发生,称为**不可能事件**,用 $\varnothing$ 表示.

必然事件和不可能事件都是确定性事件,为了讨论问题方便,将其作为两个特殊的随机事件,它们在概率问题讨论中起着重要作用.

例 1.3 掷一颗骰子的样本空间为 $S = \{1,2,3,4,5,6\}$.

记 $A = \{$出现 1 点$\}$,A 为基本事件;

$B = \{$出现奇数点$\}$,B 为复合事件;

$C = \{$出现的点数不超过 6$\}$,C 为必然事件;

$D=\{$出现 8 点$\}$，D 为不可能事件.

1.1.4　事件的关系与运算

在实际问题中，往往要在同一个试验中同时研究几个事件，这些事件是相互联系的，因而就不能只是孤立地研究单个事件，还要考虑它们之间的关系和运算性质. 分析事件之间的关系和运算，不仅能帮助人们更深入地认识事件的本质，还可以大大简化一些复杂事件的计算.

下面的讨论总假设在同一样本空间中进行. 事件作为样本空间的一个子集，它们之间的关系和运算与集合之间的关系和运算是类似的. 20 世纪 30 年代初，冯・米泽斯(Von Mises) 最先开始用集合论的观点来研究随机事件. 下面给出这些关系和运算在概率论中的含义.

(1) **包含关系**　若 $A\subset B$，称事件 B 包含事件 A，或事件 A 包含于事件 B，或 A 是 B 的子事件. 其含义为：如果事件 A 发生必然导致事件 B 发生.

注：不可能事件 $\varnothing$ 因不含有任何样本点，因此对于任意事件 A，有 $\varnothing\subset A$.

(2) **相等关系**　若 $A=B$，称事件 A 与事件 B **相等**. 其含义为：若事件 A 发生必然导致事件 B 发生，同时若事件 B 发生也必然导致事件 A 发生，即 $A\subset B$ 且 $B\subset A$. 此时事件 A 与事件 B 所包含的样本点相同.

(3) **事件的和**　事件 $A\cup B=\{x\mid x\in A$ 或 $x\in B\}$ 称为事件 A 与事件 B 的**和事件**. 其含义为：事件 A 与 B 中至少有一个发生，即由事件 A 与 B 中所有的样本点(相同的只计入一次) 组成的新事件. 称 $\bigcup\limits_{i=1}^{n}A_i$ 为 n 个事件 $A_1,A_2,\cdots,A_n$ 的和事件，称 $\bigcup\limits_{i=1}^{+\infty}A_i$ 为可列个事件 $A_1,A_2,\cdots,A_n,\cdots$ 的和事件.

(4) **事件的积**　事件 $A\cap B=\{x\mid x\in A$ 且 $x\in B\}$ 称为事件 A 与事件 B 的**积事件**. 其含义为：事件 A 与事件 B 同时发生. 事件 $A\cap B$ 也记作 AB. 称 $\bigcap\limits_{i=1}^{n}A_i$ 为 n 个事件 $A_1,A_2,\cdots,A_n$ 的积事件，称 $\bigcap\limits_{i=1}^{+\infty}A_i$ 为可列个事件 $A_1,A_2,\cdots,A_n,\cdots$ 的积事件.

(5) **事件的差**　事件 $A-B=\{x\mid x\in A$ 且 $x\notin B\}$ 称为事件 A 与事件 B 的**差事件**. 其含义为：事件 A 发生且事件 B 不发生.

(6) **互不相容(或互斥)**　若 $A\cap B=\varnothing$，则称事件 A 与事件 B 是**互不相容的**或**互斥的**. 其含义为：事件 A 和事件 B 不同时发生.

(7) **对立事件(或逆事件)**　若 $A\cup B=S$ 且 $A\cap B=\varnothing$，则称事件 A 与事件 B 互为**对立事件**，或称事件 A 与事件 B 互为**逆事件**. 其含义是：在每次试验中，事件 A 与事件 B 中有且仅有一个发生. 事件 A 的对立事件记为 $\overline{A}$，因而 $\overline{A}=S-A$.

注:互不相容的事件不一定是对立事件,但对立事件一定是互不相容的.

英国逻辑学家韦恩(Venn)给出了一种用几何图形表示事件之间关系和运算的方法,这些几何图形称为**韦恩图**(或**文氏图**)(见图 1-2).

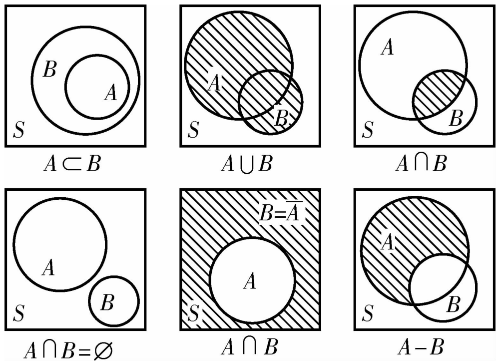


图 1-2

注:由事件的关系和运算的定义,不难看出,事件的运算满足下列关系:

① 若 $A \subset B$,则 $A \cup B = B, A \cap B = A$.

② $A\overline{B} = A - B = A - AB, A \cup B = A \cup (B - A)$.

事件的运算具有如下基本性质:

(1) **交换律** $A \cup B = B \cup A, A \cap B = B \cap A$.

(2) **结合律** $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$.

(3) **分配律** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

(4) **对偶律(德摩根律)** $\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$.

注:上述运算律可推广到有限个或可列个事件的情形. 例如,对偶律有 $\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}$, $\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}$, $\overline{\bigcup_{i=1}^{+\infty} A_i} = \bigcap_{i=1}^{+\infty} \overline{A_i}$, $\overline{\bigcap_{i=1}^{+\infty} A_i} = \bigcup_{i=1}^{+\infty} \overline{A_i}$.

例 1.4 向指定目标连续射击三枪,设 $A_i = \{$第 i 枪击中目标$\}(i = 1,2,3)$,则

(1)"至少有一枪击中目标":$A_1 \cup A_2 \cup A_3$.

(2)"三枪都击中目标":$A_1A_2A_3$.

(3)"只有第一枪击中目标":$A_1\overline{A}_2\overline{A}_3$.

(4)"恰有一枪击中目标":$A_1\overline{A}_2\overline{A}_3 \cup \overline{A}_1A_2\overline{A}_3 \cup \overline{A}_1\overline{A}_2A_3$.

(5)"三枪都没有击中目标":$\overline{A_1 \cup A_2 \cup A_3} = \overline{A}_1\overline{A}_2\overline{A}_3$.

(6)"至少有一枪没有击中目标":$\overline{A_1A_2A_3} = \overline{A}_1 \cup \overline{A}_2 \cup \overline{A}_3$.

1.2　随机事件的概率

在一次随机试验中，一个事件的发生是具有偶然性的. 人们常常希望知道在一次试验中某事件发生的可能性有多大. 事件的概率就是用来描述事件在试验中出现的可能性大小的数量指标，那么如何定义概率?很自然地想到做试验，如果某事件出现得越频繁，就认为其概率越大. 为了引入概率这个概念，下面先讨论事件频率的概念及相关性质.

1.2.1　频率及其性质

定义 1.1　在相同条件下，将随机试验重复进行了 n 次，随机事件 A 发生的次数 n_A 称为事件 A 发生的**频数**，$f_n(A)=\dfrac{n_A}{n}$ 称为事件 A 发生的**频率**.

易见，频率具有下述基本性质：

(1) $0 \leqslant f_n(A) \leqslant 1$.

(2) $f_n(S)=1$.

(3) 设 $A_1, A_2, \cdots, A_n$ 是两两互不相容的事件，则

$$f_n(A_1 \cup A_2 \cup \cdots \cup A_n)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_n).$$

以抛硬币试验为例，抛掷一枚均匀硬币，可能有两种结果：出现正面或出现反面. 就一次试验而言，看不出这些结果发生的规律，但进行大量重复试验时，发现出现正面和反面的次数大致相等(见表 1-1)，即出现正面的频率总是在 0.5 左右摆动，而且随着试验次数的增加，这个频率更加稳定地趋于 0.5.

例 1.5　考察英语中给定字母出现的频率时人们发现，当观察字母的数量 n 较小时，频率有较大幅度的波动，但随着 n 值的增大，频率呈现出稳定性(见表 1-2).

表 1-2　英文字母的使用频率

字母	使用频率	字母	使用频率	字母	使用频率
E	0.126 8	L	0.039 4	P	0.018 6
T	0.097 8	D	0.038 9	B	0.015 6
A	0.078 8	U	0.028 0	V	0.010 2
O	0.077 6	C	0.026 8	K	0.006 0

续表

字母	使用频率	字母	使用频率	字母	使用频率
I	0.070 7	F	0.025 6	X	0.001 6
N	0.070 6	M	0.024 4	J	0.001 0
S	0.063 4	W	0.021 4	Q	0.000 9
R	0.059 4	Y	0.020 2	Z	0.000 6
H	0.057 3	G	0.018 7		

字母频率分析有着重要的实用价值，如在密码学解码理论中可以利用它有效地破解单字母替换密码；在编码理论中可以用较短的码编排使用频率较高的字母；计算机键盘设计时使用频率较高的字母安排在较方便的位置等.

在实际观察中，通过大量重复试验得到随机事件频率稳定于某个数值的实例还有很多. 它们都表明，当重复试验的次数 n 逐渐增大时，频率 $f_n(A)$ 呈现出稳定性，逐渐稳定于某个常数，这个性质称为“频率的稳定性”. 由此可以看出，随机事件发生的可能性的大小可以用一个数来表示，下面给出随机事件概率的统计定义.

1.2.2 概率的定义

定义 1.2 在相同条件下重复进行 n 次试验，若随机事件 A 发生的频率 $f_n(A)=\frac{n_A}{n}$ 随着试验次数的增大而稳定地在某个常数 $p(0\leqslant p\leqslant 1)$ 附近摆动，则称 p 为事件 A 的**概率**，记为 $P(A)$.

此定义给出了随机事件概率的近似计算方法，在许多实际问题中，当事件的概率不容易计算时，往往可用频率近似估计概率的大小，而且随着试验次数的增加，估计的精度会越来越高，这正是 1946 年冯 · 诺依曼（Von Neumann）和乌拉姆（Ulam）创立的蒙特卡洛方法的基本思想.

但是概率的统计定义有一定的局限性，在实际问题中，不可能对每一个事件都做大量试验，再求得事件的概率. 1933 年前苏联著名数学家柯尔莫哥洛夫（Kolmogorov）在他的《概率论的基本概念》一书中提出了概率的公理化结构，使概率论成为严谨的数学分支. 下面给出概率的公理化定义.

定义 1.3 设随机试验 E 的样本空间为 S，对于 E 的每个事件 A 赋予一个实数 $P(A)$，若 $P(A)$ 满足如下三个条件：

（1）**非负性** $P(A)\geqslant 0$.

（2）**规范性** $P(S)=1$.

(3) **可列可加性**　对可列个两两互不相容的事件 $A_1, A_2, \cdots, A_n, \cdots$，有

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i),$$

则称 $P(A)$ 为事件 A 发生的**概率**.

1.2.3　概率的性质

由概率的公理化定义，可以得到概率的一些重要性质.

性质 1　$P(\varnothing) = 0$.

证明　令 $A_i = \varnothing, i = 1,2,3,\cdots$，则 $\bigcup\limits_{i=1}^{+\infty} A_i = \varnothing$，且有 $A_iA_j = \varnothing, i \neq j, i,j = 1,2,3,\cdots$，由概率的可列可加性得

$$P(\varnothing) = P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i) = \sum_{i=1}^{+\infty} P(\varnothing),$$

由概率的非负性知 $P(\varnothing) \geqslant 0$，结合上式可得 $P(\varnothing) = 0$.

注：由 $P(A) = 0$，不能得到 $A = \varnothing$；由 $P(A) = 1$，也不能得到 $A = S$.

性质 2(有限可加性)　若事件 $A_1, A_2, \cdots, A_n$ 两两互不相容，则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

证明　令 $A_{n+1} = A_{n+2} = \cdots = \varnothing$，则有 $A_iA_j = \varnothing (i \neq j, i,j = 1,2,3,\cdots)$，由概率的可列可加性知

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{+\infty} P(A_i) \\ &= \sum_{i=1}^{n} P(A_i) + 0 = P(A_1) + P(A_2) + \cdots + P(A_n). \end{aligned}$$

性质 3(减法公式)　若 A, B 为任意两个事件，则

$$P(A - B) = P(A) - P(AB).$$

特别地，若 $B \subset A$，则有

$$P(A - B) = P(A) - P(B)，且\ P(A) \geqslant P(B).$$

证明　由于 $A = AB \cup (A - B)$，且 AB 与 $A - B$ 互不相容，由性质 2 得

$$P(A) = P(AB) + P(A - B),$$

因此

$$P(A - B) = P(A) - P(AB).$$

特别地，若 $B \subset A$，则

$$P(A) = P(AB) + P(A - B) = P(B) + P(A - B),$$

即

$$P(A - B) = P(A) - P(B),$$

且由概率的非负性知 $P(A-B)\geqslant 0$,因而 $P(A)\geqslant P(B)$.

性质 4 对任一事件 A,有 $P(\overline{A})=1-P(A)$.

证明 由于 $A\cup\overline{A}=S$,且 $A\overline{A}=\varnothing$,由性质 2 得

$$1=P(S)=P(A\cup\overline{A})=P(A)+P(\overline{A}),$$

即 $P(\overline{A})=1-P(A)$.

性质 5 对于任意一个事件 A,有 $P(A)\leqslant 1$.

证明 因 $A\subset S$, 由性质 3 得

$$P(A)\leqslant P(S)=1.$$

性质 6(加法公式) 若 A,B 为任意两个事件,则

$$P(A\cup B)=P(A)+P(B)-P(AB).$$

证明 因 $A\cup B=A\cup(B-AB)$, $A\cap(B-AB)=\varnothing$ 且 $AB\subset B$,因而

$$P(A\cup B)=P(A)+P(B-AB)=P(A)+P(B)-P(AB).$$

注:性质 6 可以推广到 n 个事件的情形,如当 $n=3$ 时,

$$P(A_1\cup A_2\cup A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3).$$

一般地,对于任意 n 个事件 $A_1,A_2,\cdots,A_n$,有

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n).$$

例 1.6 设事件 $A,B,A\cup B$ 的概率分别为 p,q,r.求:

(1) $P(AB)$;(2) $P(A\overline{B})$;(3) $P(\overline{A}\,\overline{B})$.

解 (1) 由 $P(A\cup B)=P(A)+P(B)-P(AB)$ 知

$$P(AB)=P(A)+P(B)-P(A\cup B)=p+q-r.$$

(2) 由 $A=AS=A(B\cup\overline{B})=AB\cup A\overline{B}$, 且 AB 与 $A\overline{B}$ 互不相容知

$$P(A)=P(AB)+P(A\overline{B}),$$

即

$$P(A\overline{B})=P(A)-P(AB)=r-q.$$

(3) $P(\overline{A}\,\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-r$.

例 1.7 观察某地区未来5天的天气情况,记 $A_i=\{$有 i 天不下雨$\}$,$i=0,1,2,3,4,5$.已知 $P(A_j)=jP(A_0)$,$j=1,2,3,4,5$.求下列各事件的概率:

(1) 5 天均下雨;(2) 至少 1 天不下雨;(3) 至多 3 天不下雨.

解 显然 $A_0,A_1,\cdots,A_5$ 是两两不相容的事件且 $\bigcup\limits_{i=0}^{5}A_i=S$,故

$$1=P(S)=P\left(\bigcup_{i=0}^{5}A_i\right)=\sum_{i=0}^{5}P(A_i)=P(A_0)+\sum_{i=1}^{5}iP(A_0)=16P(A_0),$$

于是

$$P(A_0)=\frac{1}{16},\ P(A_i)=\frac{i}{16},i=1,2,3,4,5.$$

记(1),(2),(3) 中三个事件分别为 A,B,C,则

(1) $P(A)=P(A_0)=\dfrac{1}{16}$.

(2) $P(B)=P\left(\bigcup\limits_{i=1}^{5}A_i\right)=1-P(A_0)=\dfrac{15}{16}$.

(3) $P(C)=P\left(\bigcup\limits_{i=0}^{3}A_i\right)=\sum\limits_{i=0}^{3}P(A_i)=\dfrac{7}{16}$.

1.3　古典概型和几何概型

概率的公理化定义并没有告诉我们怎么去确定一个事件的概率. 本节将讨论两类比较简单的随机试验,这两类随机试验中每个样本点的出现是等可能的. 下面先介绍在有限集计数中常用的排列与组合公式.

1.3.1　排列与组合公式

排列与组合都是用来计算“从 n 个元素中任取 r 个” 的取法总数的公式,它们的区别在于, 如果不考虑取出元素间的次序,则用组合公式,否则用排列公式. 有限集合的计数都基于如下两条基本计数原理:

(1) **加法原理**　若完成一件事有 m 类不同途径,其中第一类途径有 n_1 种方法,第二类途径有 n_2 种方法,…,第 m 类途径有 n_m 种方法,则完成这件事的方法总数为 $n_1+n_2+\cdots+n_m$.

(2) **乘法原理**　若一件事需经 m 个步骤才能完成,其中第一个步骤有 n_1 种方法,第二个步骤有 n_2 种方法,…,第 m 个步骤有 n_m 种方法,则完成这件事的方法总数为 $n_1\cdot n_2\cdot\cdots\cdot n_m$.

排列与组合的定义及其计算公式如下:

(1) **排列**　从 n 个不同元素中任取 $r(r\leqslant n)$ 个排成一列(考虑元素间的先后次序), 称此为一个排列,此种排列的总数记为 P_n^r. 按乘法原理,取出的第一个元素有 n 种取法,第二个元素有 $n-1$ 种取法,…,第 r 个元素有 $n-r+1$ 种取法,因而有

$$\mathrm{P}_n^r=n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot(n-r+1)=\frac{n!}{(n-r)!}.$$

特别地，当 $r=n$ 时，$\mathrm{P}_n^n=n!$.

(2) **组合**　从 n 个不同元素中任取 $r(r\leqslant n)$ 个并成一组(不考虑元素间的先后次序)，称此为一个组合，此种组合的总数记为 C_n^r 或 $\binom{n}{r}$，按乘法原理，此种组合的总数为

$$\mathrm{C}_n^r=\frac{\mathrm{P}_n^r}{r!}=\frac{n!}{r!(n-r)!}.$$

这里规定 $0!=1$ 和 $\mathrm{C}_n^0=1$.

注：组合数满足下列性质：

① $\mathrm{C}_n^r=\mathrm{C}_n^{n-r}$.

② $\mathrm{C}_n^r=\mathrm{C}_{n-1}^r+\mathrm{C}_{n-1}^{r-1}$.

③ $\mathrm{C}_n^0+\mathrm{C}_n^1+\cdots+\mathrm{C}_n^n=2^n$.

④ $\mathrm{C}_n^1+2\mathrm{C}_n^2+\cdots+n\mathrm{C}_n^n=\sum\limits_{i=1}^{n}i\mathrm{C}_n^i=n2^{n-1}$.

⑤ $\sum\limits_{i=0}^{m}\mathrm{C}_n^i\mathrm{C}_m^{m-i}=\mathrm{C}_{m+n}^m(n\geqslant m)$.

1.3.2　古典概型

在概率论发展的初期，人们研究的是一类特殊的随机试验，这类试验具有以下两个特点：

(1) **有限性**　随机试验只有有限个样本点.

(2) **等可能性**　每一个样本点发生的可能性相同.

具备以上两个特征的随机试验称为**等可能概型**，由于这种等可能的数学模型曾经是概率论发展初期的主要研究对象，所以也称为**古典概型**. 古典概型在概率论中有很重要的地位，它是一类简单、直观的随机试验，如抛硬币、掷骰子、抽签等. 下面利用概率的性质来讨论古典概型中事件概率的计算公式.

设试验的样本空间为 $S=\{e_1,e_2,\cdots,e_n\}$. 在古典概型的假设下，试验中每个基本事件发生的可能性相同，即有

$$P(\{e_1\})=P(\{e_2\})=\cdots=P(\{e_n\}).$$

又由于基本事件是两两互不相容的，因而

$$\begin{aligned}1=P(S)&=P(\{e_1\}\cup\{e_2\}\cup\cdots\cup\{e_n\})\\&=P(\{e_1\})+P(\{e_2\})+\cdots+P(\{e_n\})=nP(\{e_i\}),\end{aligned}$$

即

$$P(\{e_i\})=\frac{1}{n},\ i=1,2,\cdots,n.$$

若事件 A 包含其样本空间 S 中 k 个基本事件，即

$$A=\{e_{i_1}\}\cup\{e_{i_2}\}\cup\cdots\cup\{e_{i_k}\},$$

则事件 A 发生的概率

$$P(A)=P\Big(\bigcup_{j=1}^{k}\{e_{i_j}\}\Big)=\sum_{j=1}^{k}P(\{e_{i_j}\})=\frac{k}{n}=\frac{A\text{ 包含的基本事件数}}{S\text{ 中基本事件的总数}}.\tag{1.1}$$

式(1.1)就是古典概型中事件概率的计算公式，它把计算事件概率的问题转化为对基本事件的计数问题.

例 1.8　将 1,2,3,4 四个数随意地排成一行，求下列各事件的概率：

(1) 自左至右或自右至左恰好排成 1,2,3,4 的顺序.

(2) 数字 1 排在最右边或最左边.

(3) 数字 1 与数字 2 相邻.

(4) 数字 1 排在数字 2 的右边(不一定相邻).

解　将 4 个数随意地排成一行有 $4!=24$ 种排法，即基本事件总数为 24. 记(1),(2),(3),(4) 的事件分别为 A,B,C,D.

(1) 事件 A 中共有 2 种排法，因而

$$P(A)=\frac{2}{24}=\frac{1}{12}.$$

(2) 事件 B 中有 $2\times(3!)=12$ 种排法，故有

$$P(B)=\frac{12}{24}=\frac{1}{2}.$$

(3) 先将数字 1 和 2 排在任意相邻两个位置，共有 2×3 种排法，其余两个数可在其余两个位置任意排放，共有 $2!$种排法，因而事件 C 有 $2\times3\times2=12$ 种排法，即

$$P(C)=\frac{12}{24}=\frac{1}{2}.$$

(4) 数字 1 排在数字 2 的右边的每一种排法，交换数字 1 和数字 2 的位置便对应于数字 1 排在数字 2 的左边的一种排法，反之亦然. 因而，数字 1 排在数字 2 的右边与数字 1 排在数字 2 的左边的排法种数相同，各占总排法数的$\frac{1}{2}$，故有

$$P(D)=\frac{1}{2}.$$

例 1.9　一盒中共有 N 只球，其中 M 只红球，$N-M$ 只白球，从中抽取 n 个，考虑两种取球方式：

(1) 每次取一只球，观察其颜色后放回盒中，搅匀后再取，直至取 n 次，这种取球方式称为**放回抽样**；

(2) 第一次取一球不放回，第二次从剩余的球中再取一球，如此继续下去，直至

取出 n 只球，这种取球方式称为**不放回抽样**.

试分别就上面的两种情况，设 $A_m=\{$取出的 n 只球中至少有 m 只红球$\}$，$B_m=\{$取出的 n 只球中恰有 m 只红球$\}$，求 $P(A_m)$ 及 $P(B_m)$（$m<\min(n,M)$）.

解 （1）放回抽样.

在放回抽样的情况下，从 N 只球中取 n 只，共有 N^n 种取法.

事件 A_m 相当于从 n 次取球中先选取 m 次，使得这 m 次都取红球，剩下的 $n-m$ 次可以任意取，因而 A_m 中总的取法有 $\mathrm{C}_n^m M^m N^{n-m}$ 种.

事件 B_m 相当于从 n 次取球中先选取 m 次，使得这 m 次都取红球，剩下的 $n-m$ 次都取白球，因而 B_m 中总的取法有 $\mathrm{C}_n^m M^m (N-M)^{n-m}$ 种.

从而，根据古典概率计算

$$P(A_m)=\frac{\mathrm{C}_n^m M^m N^{n-m}}{N^n},$$

$$P(B_m)=\frac{\mathrm{C}_n^m M^m (N-M)^{n-m}}{N^n}.$$

（2）不放回抽样.

在不放回抽样的情况下，从 N 只球中取 n 只，从排列的角度看共有 $\mathrm{C}_N^1\mathrm{C}_{N-1}^1\cdots\mathrm{C}_{N-n+1}^1=\mathrm{P}_N^n$ 种取法.

事件 A_m 相当于从 n 次取球中先选取 m 次，这 m 次是不放回地顺次取 m 只红球，剩下的 $n-m$ 次可以不放回地任意取，因而 A_m 中总的取法有 $\mathrm{C}_n^m\mathrm{P}_M^m\mathrm{P}_{N-m}^{n-m}$ 种.

事件 B_m 相当于从 n 次取球中先选取 m 次，使得这 m 次都不放回地顺次取红球，剩下的 $n-m$ 次都不放回地顺次取白球，因而 B_m 中总的取法有 $\mathrm{C}_n^m\mathrm{P}_M^m\mathrm{P}_{N-M}^{n-m}$ 种.

从而，根据古典概率计算

$$P(A_m)=\frac{\mathrm{C}_n^m\mathrm{P}_M^m\mathrm{P}_{N-m}^{n-m}}{\mathrm{P}_N^n},P(B_m)=\frac{\mathrm{C}_n^m\mathrm{P}_M^m\mathrm{P}_{N-M}^{n-m}}{\mathrm{P}_N^n}.$$

注：对于(2) 不放回抽样的情形，也可以从组合的角度来计数，这时“从 N 只球中取 n 只”共有 C_N^n 种取法，事件“取出的 n 只球中至少有 m 只红球”共有 $\mathrm{C}_M^m\mathrm{C}_{N-m}^{n-m}$ 种取法；事件“取出的 n 只球中恰有 m 只红球”共有 $\mathrm{C}_M^m\mathrm{C}_{N-M}^{n-m}$ 种取法.

在用排列组合公式计算古典概率时，先要看清所考虑的是排列问题还是组合问题，在计算样本空间 S 和事件 A 所包含的基本事件数时，两者一定要保持一致，是排列就都是排列，是组合就都是组合. 不要重复计数，也不要遗漏.

例 1.10 袋中有 a 支红签，b 支白签，它们除颜色外无差别，现有 $a+b$ 个人无放回地去抽签，求第 k 个人抽到红签的概率（$1\leqslant k\leqslant a+b$）.

解法一 记 $A=\{$第 k 个人抽到红签$\}$.

假设这 $a+b$ 支签都不相同，让这 $a+b$ 个人抽完签以后排成一排，相当于把这

$a+b$ 支签排成一排，共有 $(a+b)!$ 种排法. 第 k 个人抽到红签相当于第 k 个位置放的是红签，先在第 k 个位置放一支红签，有 C_a^1 种放法，其余各签随意放，有 $(a+b-1)!$ 种放法，故事件 A 包含 $C_a^1(a+b-1)!$ 个基本事件，因而

$$P(A)=\frac{C_a^1(a+b-1)!}{(a+b)!}=\frac{a}{a+b}.$$

解法二　记 $A=\{$第 k 个人抽到红签$\}$.

假设这 $a+b$ 支签都不相同，只看前 k 个人的抽签情况，共有 P_{a+b}^k 种可能排法，事件 A 包含了 $C_a^1 P_{a+b-1}^{k-1}$ 个基本事件，所以

$$P(A)=\frac{C_a^1 P_{a+b-1}^{k-1}}{P_{a+b}^k}=\frac{a(a+b-1)!(a+b-k)!}{(a+b-k)!(a+b)!}=\frac{a}{a+b}.$$

注：由上例可以看到，对于古典概型可用不同的等可能基本事件组来描述，对同一事件的概率也常有多种不同的解法. 本题结果还说明，每个人抽到红签的概率与抽签的先后次序无关. 所以，进行分组的时候采用抽签或抓阄的方法是公平的.

例 1.11　在 $1\sim 2\,000$ 的整数中随机地取一个数，问取到的整数既不能被 6 整除，又不能被 8 整除的概率.

解　记 $A=\{$取到的数能被 6 整除$\}$，$B=\{$取到的数能被 8 整除$\}$，则所求概率为 $P(\overline{A}\,\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-[P(A)+P(B)-P(AB)]$.

由于 $333<\frac{2\,000}{6}<334$，因而

$$P(A)=\frac{333}{2\,000}.$$

同理，由于 $\frac{2\,000}{8}=250$，得

$$P(B)=\frac{250}{2\,000}.$$

又由于一个数同时能被 6 与 8 整除，就相当于能被 24 整除，且由 $83<\frac{2\,000}{24}<84$，得

$$P(AB)=\frac{83}{2\,000}.$$

于是所求概率为

$$P(\overline{A}\,\overline{B})=1-\left(\frac{333}{2\,000}+\frac{250}{2\,000}-\frac{83}{2\,000}\right)=\frac{3}{4}.$$

1.3.3　几何概型

古典概型考虑的是有有限个样本点且每个样本点是等可能发生的随机试验的

概率模型,当试验的样本点有无穷多个的时候,古典概型就不适用了.下面考虑如何将古典概型作必要的推广,使给出的模型能适用于有无穷多个样本点且每个样本点发生的可能性相等的试验.

设随机试验 E 的样本空间有无限多个样本点,其可用欧氏空间的某个区域 S 来表示(假设区域 S 以及 S 中的任意小区域 A 都是可以度量的,其度量大小用 $\mu(A)$ 表示.例如,一维区间的长度、二维平面的面积、三维空间的体积等),各个样本点出现的可能性相同,即向 S 中任投一点,则此点落在 S 中某区域 A 中的可能性大小只与区域的度量成正比,而与区域 A 的位置及形状无关,设事件 A 的概率为 $P(A)=k\mu(A)$,其中 k 为常数,而 $P(S)=k\mu(S)$,于是 $k=\dfrac{1}{\mu(S)}$,从而事件 A 的概率为

$$P(A)=\frac{\mu(A)}{\mu(S)},$$

称具有这种特性的试验为**几何概型**.

例 1.12(会面问题) 甲、乙两人相约在 8 点到 9 点之间在某处会面,并约定先到者应等候另一人 20 分钟,过时就离开.如果每个人可在指定的一小时内任意时刻到达,试计算二人能够会面的概率.

解 设 $A=\{$甲、乙两人能会面$\}$.记 8 点为计算时刻的 0 时,以分钟为单位,设甲、乙两人到达的时刻分别是 x,y,则样本空间为

$$S=\{(x,y)\mid 0\leqslant x\leqslant 60,0\leqslant y\leqslant 60\}.$$

事件 A(见图 1-3)可表示为

$$A=\{(x,y)\mid (x,y)\in S,\ |x-y|\leqslant 20\}.$$

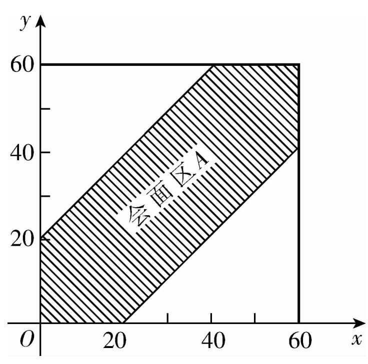


图 1-3

由题意知,这是一个几何概型问题,于是

$$P(A)=\frac{\mu(A)}{\mu(S)}=\frac{60^2-40^2}{60^2}=\frac{5}{9}.$$

注：求几何概型的概率关键是用图形(一般用线段、平面或空间图形)清楚地描述样本空间 S 和所求事件 A，然后计算出相关图形的度量(一般为长度、面积或体积).

例 1.13(三角形构成问题)　在长度为 a 的线段内任取两点将其分为三段，求它们可以构成一个三角形的概率.

解　令被截成的三段线段长度分别为 $x,y,a-x-y$，设 $A=\{$截成的三线段能构成三角形$\}$，则样本空间 S 和事件 A(见图 1-4) 可表示为

$$\begin{aligned} S &= \{(x,y) \mid 0<x<a, 0<y<a, 0<a-x-y<a\} \\ &= \{(x,y) \mid 0<x<a, 0<y<a, 0<x+y<a\}, \\ A &= \{(x,y) \mid 0<x<a, 0<y<a, 0<x+y<a, \\ &\qquad x+y>a-x-y, a-x>x, a-y>y\} \\ &= \{(x,y) \mid x<\frac{a}{2}, y<\frac{a}{2}, x+y>\frac{a}{2}\}. \end{aligned}$$

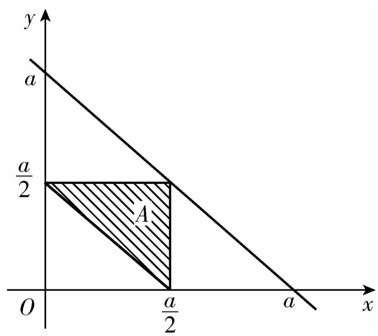


图 1-4

所以

$$P(A)=\frac{\mu(A)}{\mu(S)}=\frac{1}{4}.$$

1.4　条件概率与事件的独立性

1.4.1　条件概率

在实际问题中，除了考虑某个事件 A 的概率 $P(A)$，经常还需要考虑在另一个事件 B 已经发生的条件下，事件 A 发生的概率. 例如，在遗传学上，考虑在已知父亲的

身高不低于1.75米的情况下，子辈的身高不低于1.70米的概率；在统计某城市居民人均年收入时，通常也是考虑在已知居民的学历水平为本科的条件下，他们的人均年收入不低于某个值的概率．类似于这样的，在某个事件 B 已经发生的条件下，事件 A 发生的概率，称为条件概率，记为 $P(A \mid B)$．下面通过一个例子来给出条件概率的数学定义．

例 1.14 掷一颗均匀的骰子，若已知掷出的是奇数点，问掷出的点数小于 3 的概率．

解 这个问题的解决并不困难，由题意知掷出的是奇数点，只是不能肯定这个奇数点是 1，3，5 中的哪一个．显然，“掷出的点数小于 3”这一事件就相当于“掷出的是 1 点”，因而，在已知掷出的是奇数点的条件下，掷出的点数小于 3 的概率是$\frac{1}{3}$．记“掷出的点数小于 3”这一事件为 A，“掷出的是奇数点”这一事件为 B，即$P(A \mid B)=\frac{1}{3}$．

注意到 $P(A \mid B)=\frac{1}{3}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{P(AB)}{P(B)}$．

事实上，不难验证，对于一般的古典概型问题，只要 $P(B)>0$，都有

$$P(A \mid B)=\frac{P(AB)}{P(B)}.$$

同样，在几何概型中，在已知事件 A 发生的条件下，事件 B 发生的概率也可以表示为

$$P(A \mid B)=\frac{\mu(AB)}{\mu(B)}=\frac{\mu(AB)/\mu(S)}{\mu(B)/\mu(S)}=\frac{P(AB)}{P(B)}.$$

因而，在古典概型和几何概型中都有 $P(A \mid B)=\frac{P(AB)}{P(B)}$．在一般场合，将上述关系式作为条件概率的数学定义．

定义 1.4 设 A，B 为两个事件，且 $P(B)>0$，称

$$P(A \mid B)=\frac{P(AB)}{P(B)}$$

为事件 B 发生的条件下事件 A 发生的**条件概率**．

注：相对于条件概率 $P(A \mid B)$，$P(A)$ 称为**无条件概率**．$P(A)$ 表示在样本空间 S 中事件 A 发生的概率．而 $P(A \mid B)$ 则是局限在事件 B 发生的范围内，事件 A 发生的概率，相当于把事件 A 所在的样本空间缩小了．

设 B 是一事件，且 $P(B)>0$．容易验证，条件概率 $P(\cdot \mid B)$ 符合概率定义中的三个条件，即

(1) 对于任意事件 A，$0 \leqslant P(A \mid B) \leqslant 1$.

(2) $P(S \mid B) = 1$.

(3) 设 $A_1, A_2, \cdots, A_n, \cdots$ 是两两互不相容的事件，则有

$$P\left(\bigcup_{i=1}^{+\infty} A_i \mid B\right) = \sum_{i=1}^{+\infty} P(A_i \mid B).$$

此外，前面证明的概率的性质都适用于条件概率.

计算条件概率有两种方法：

(1) 在缩减的样本空间 B 中求事件 A 的概率，就得到 $P(A \mid B)$.

(2) 在样本空间 S 中，先求 $P(AB)$ 和 $P(B)$，再按定义计算 $P(A \mid B)$.

例 1.15　袋中有 3 个红球和 2 个白球. 现从袋中不放回地连取 2 个. 已知第一次取得红球，求第二次取得白球的概率.

解　记 $A=\{$第二次取得白球$\}$，$B=\{$第一次取得红球$\}$，依题意要求 $P(A \mid B)$.

解法一　从袋中不放回地连续取两个球，则第一次取得红球的取法有 $\mathrm{P}_3^1\mathrm{P}_4^1$ 种，在缩减的样本空间 B 中考虑第二次取得白球的取法有 $\mathrm{P}_3^1\mathrm{P}_2^1$ 种，所以

$$P(A \mid B) = \frac{\mathrm{P}_3^1\mathrm{P}_2^1}{\mathrm{P}_3^1\mathrm{P}_4^1} = \frac{1}{2}.$$

也可以直接计算，因为第一次取走了 1 个红球，袋中只剩下 4 个球，其中有 2 个白球和 2 个红球，再从中任取 1 个，取得白球的概率为 $\frac{2}{4}$，所以

$$P(A \mid B) = \frac{2}{4} = \frac{1}{2}.$$

解法二　在 5 个球中不放回连取 2 个球的取法有 P_5^2 种，其中第一次取得红球的取法有 $\mathrm{P}_3^1\mathrm{P}_4^1$ 种，第一次取得红球第二次取得白球的取法有 $\mathrm{P}_3^1\mathrm{P}_2^1$ 种，所以

$$P(B) = \frac{\mathrm{P}_3^1\mathrm{P}_4^1}{\mathrm{P}_5^2} = \frac{3}{5}, P(AB) = \frac{\mathrm{P}_3^1\mathrm{P}_2^1}{\mathrm{P}_5^2} = \frac{3}{10}.$$

由定义得

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{3/10}{3/5} = \frac{1}{2}.$$

1.4.2　乘法公式

由条件概率的定义，容易得到

$$P(AB) = P(A)P(B \mid A)(P(A) > 0) \tag{1.2}$$

或者

$$P(AB) = P(B)P(A \mid B)(P(B) > 0). \tag{1.3}$$

式(1.2) 和式(1.3) 称为**乘法公式**，利用乘法公式可以计算积事件的概率.

乘法公式可以推广到多个事件的积事件的情形. 例如,对于 3 个事件 A_1,A_2,A_3,且 $P(A_1A_2)>0$,有

$$P(A_1A_2A_3)=P(A_3\mid A_1A_2)P(A_2\mid A_1)P(A_1).$$

一般地, 对于 n 个事件 $A_1,A_2,\cdots,A_n,n\geqslant 2$,且 $P(A_1A_2\cdots A_{n-1})>0$(此假设可保证下式中所有条件概率都有意义),有

$$P(A_1A_2\cdots A_n)=P(A_n\mid A_1\cdots A_{n-1})P(A_{n-1}\mid A_1\cdots A_{n-2})\cdots P(A_2\mid A_1)P(A_1).$$

例 1.16 一批零件共有 100 个,其中有 10 个不合格品. 从中一个一个取出,求第三次才取得不合格品的概率.

解 令 $A_i=\{$第 i 次取出的是不合格品$\}$,$i=1,2,3$,则所求概率为

$$P(\overline{A}_1\overline{A}_2A_3)=P(A_3\mid \overline{A}_1\overline{A}_2)P(\overline{A}_2\mid \overline{A}_1)P(\overline{A}_1)=\frac{10}{98}\times\frac{89}{99}\times\frac{90}{100}\approx 0.082\,6.$$

例 1.17(**罐子模型**,又叫**波利亚模型**) 设罐中有 b 个黑球、r 个红球,每次随机取出一个球,观察其颜色后将球放回,同时加进 c 个同色球和 d 个异色球. 若在袋中连续取球四次,试求第一、三次取红球且第二、四次取黑球的概率.

解 令 $B_i=\{$第 i 次取出的是黑球$\}$,$i=1,2,3,4$;$R_j=\{$第 j 次取出的是红球$\}$,$j=1,2,3,4$. 由题意, 所求概率为

$$\begin{aligned}P(R_1B_2R_3B_4)&=P(B_4\mid R_1B_2R_3)P(R_3\mid R_1B_2)P(B_2\mid R_1)P(R_1)\\&=\frac{b+c+2d}{b+r+3c+3d}\cdot\frac{r+c+d}{b+r+2c+2d}\cdot\frac{b+d}{b+r+c+d}\cdot\frac{r}{b+r}.\end{aligned}$$

1.4.3 事件的独立性

条件概率 $P(A\mid B)$ 反映了事件 B 对事件 A 的影响,一般来说,$P(A)$ 与$P(A\mid B)$ 是不相等的. 但在许多实际问题中,常会遇到这样一种情况:两个事件中的一个事件的发生不影响另一个事件的发生,即有 $P(A)=P(A\mid B)$. 由此引出了事件独立的概念.

定义 1.5 若两事件 A 与 B 满足

$$P(AB)=P(A)P(B),$$

则称事件 A 与 B **相互独立**,简称 A 与 B **独立**.

注:① 不可能事件及必然事件与任何事件都独立.

② 两事件相互独立与互不相容是两个完全不同的概念,它们分别从不同角度刻画了两事件之间的某种关系. 互不相容是事件的集合属性,而相互独立是事件的概率属性. 可以知道,若 $P(A)>0$,$P(B)>0$,则事件 A 与 B 互不相容和独立不能同时成立,或者说,若 A 与 B 既独立又互不相容,则一定有 $P(A)=0$ 或 $P(B)=0$.

由条件概率和独立性的定义，有下面的结论.

定理 1.1　设 A 与 B 是两事件，且 $P(B)>0$，若 A 与 B 相互独立，则 $P(A)=P(A\mid B)$. 反之亦然.

定理 1.2　若事件 A,B 相互独立，则 (1) A 与 $\overline{B}$ 独立；(2) $\overline{A}$ 与 B 独立；(3) $\overline{A}$ 与 $\overline{B}$ 独立.

证明　(1) 因为

$$\begin{aligned}P(A\overline{B})&=P(A)-P(AB)=P(A)-P(A)P(B)\\&=P(A)[1-P(B)]=P(A)P(\overline{B}),\end{aligned}$$

所以 A 与 $\overline{B}$ 独立.

同理可证(2)，(3).

因此，从概率的角度看，两事件相互独立是指一个事件的发生不影响另一个事件的发生. 在实际应用中，判断两事件是否独立也常采用这个标准.

独立的概念可以推广到多个事件的情形.

定义 1.6　设 A,B,C 是三个事件，如果有

$$P(AB)=P(A)P(B),$$
$$P(AC)=P(A)P(C),$$
$$P(BC)=P(B)P(C),$$

则称事件 A,B,C **两两独立**.

若同时还有

$$P(ABC)=P(A)P(B)P(C),$$

则称事件 A,B,C **相互独立**.

注：相互独立的事件一定是两两独立的，反之不成立.

例 1.18　设有 4 张卡片，其中 3 张上分别标有数字 1，2，3，剩下的 1 张上同时标有这 3 个数字. 现从 4 张卡片中任取 1 张，记 $A_i=\{$取出的卡片上标有数字 $i\}$，$i=1,2,3$. 试分析 A_1,A_2,A_3 三个事件间的独立情况.

解　由题意知

$$P(A_1)=P(A_2)=P(A_3)=\frac{1}{2},$$

$$P(A_1A_2)=P(A_1A_3)=P(A_2A_3)=\frac{1}{4}.$$

由于

$$P(A_1A_2)=\frac{1}{4}=P(A_1)P(A_2),$$

$$P(A_1A_3)=\frac{1}{4}=P(A_1)P(A_3),$$

$$P(A_2A_3)=\frac{1}{4}=P(A_2)P(A_3),$$

因而 A_1,A_2,A_3 三个事件两两独立.

又知 $P(A_1A_2A_3)=\frac{1}{4}$，但 $P(A_1)P(A_2)P(A_3)=\frac{1}{8}$，即 $P(A_1A_2A_3)\neq P(A_1)P(A_2)P(A_3)$，因而 A_1,A_2,A_3 不相互独立.

一般地，任意 n 个事件 $A_1,A_2,\cdots,A_n$，若对任意的 $k(1<k\leqslant n)$ 个事件 $A_{i_1},A_{i_2},\cdots,A_{i_k}$，$1\leqslant i_1<i_2<\cdots<i_k\leqslant n$，都有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}),$$

则称此 n 个事件 $A_1,A_2,\cdots,A_n$ **相互独立.**

显然，若 n 个事件相互独立，则它们中的任意 $m(2\leqslant m\leqslant n)$ 个事件也相互独立.

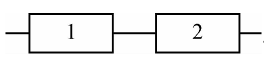
此外，对于多个相互独立的事件也有类似于两个事件独立的定理 1.2 的结论，即若 $A_1,A_2,\cdots,A_n$ 相互独立，则将它们中的任意 $k(1\leqslant k\leqslant n)$ 个事件换成它们的对立事件后所得到的 n 个事件仍相互独立.

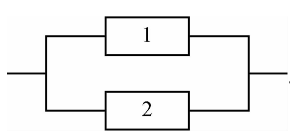
注：在计算 n 个相互独立的事件的和的概率时，常常用到如下计算公式：

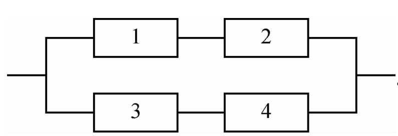
$$\begin{aligned}P(A_1\cup A_2\cup\cdots\cup A_n)&=1-P(\overline{A_1\cup A_2\cup\cdots\cup A_n})\\&=1-P(\overline{A}_1\overline{A}_2\cdots\overline{A}_n)\\&=1-P(\overline{A}_1)P(\overline{A}_2)\cdots P(\overline{A}_n).\end{aligned}$$

近年来，可靠性理论得到迅速发展. 元件的可靠度指的是一个原件在规定的时间内和规定的条件下能完成规定功能的概率. 系统的可靠度是指由元件组成的一个系统在规定时间内和规定的条件下能正常工作的概率. 下面通过例子来说明事件的独立性在可靠性理论中的应用.

例 1.19 系统由多个元件组成，且所有元件都独立地工作. 设每个元件正常工作的概率都是 $p=0.9$，试求以下系统正常工作的概率：

(1) 串联系统 S_1：.

(2) 并联系统 S_2：.

(3) 混联系统 S_3：.

解 设 $A_i=\{$第 S_i 个系统正常工作$\}$，$i=1,2,3$；$B_j=\{$第 j 个元件正常工作$\}$，$j=1,2,3,4$.

(1) $P(A_1)=P(B_1B_2)=P(B_1)P(B_2)=p^2=0.81$.

(2) $P(A_2)=P(B_1\cup B_2)=P(B_1)+P(B_2)-P(B_1)P(B_2)=p+p-p^2=0.99$.

(3) 易知 $A_3=B_1B_2\cup B_3B_4$，由此

$$\begin{aligned}P(A_3)&=P(B_1B_2\cup B_3B_4)=P(B_1B_2)+P(B_3B_4)-P(B_1B_2B_3B_4)\\&=P(B_1)P(B_2)+P(B_3)P(B_4)-P(B_1)P(B_2)P(B_3)P(B_4)\\&=2p^2-p^4=0.9639.\end{aligned}$$

1.4.4　伯努利试验

利用事件的独立性可以定义两个或更多个试验的独立性.

定义 1.7　设有两个试验 E_1 和 E_2，假如试验 E_1 的任一结果（事件）与试验 E_2 的任一结果（事件）都是相互独立的事件，则称这**两个试验相互独立**.

类似地，可定义 n 个试验 $E_1,E_2,\cdots,E_n$ 的相互独立性：如果 E_1 的任一结果、E_2 的任一结果、…、E_n 的任一结果都是相互独立的事件，则称**试验 $\boldsymbol{E_1,E_2,\cdots,E_n}$ 相互独立**. 如果这 n 个试验还是相同的，则称其为 **$\boldsymbol{n}$ 重独立重复试验**；如果在 n 重独立重复试验中，每次试验只有两个可能结果：事件 A 发生及事件 A 不发生，则称这类试验为 **$\boldsymbol{n}$ 重伯努利（Bernoulli）试验**，也称为**伯努利概型**.

例如，掷 n 枚硬币、掷 n 颗骰子、检查 n 个产品等等都是 n 重独立重复试验，它们也可以看做 n 重伯努利试验.

定理 1.3　在 n 重伯努利试验中，设每次试验中事件 A 发生的概率为 $p(0<p<1)$，则在 n 次试验中，事件 A 恰好发生 k 次的概率为

$$P_n(k)=\mathrm{C}_n^kp^k(1-p)^{n-k},k=0,1,\cdots,n.$$

证明　记 $A_i=\{$第 i 次试验中事件 A 发生$\}$，$i=1,2,\cdots,n$. 在 n 次试验中，事件 A 在指定的 k 次试验中发生（比如前 k 次），其余 $n-k$ 次试验中不发生的概率为

$$P(A_1A_2\cdots A_k\bar{A}_{k+1}\cdots\bar{A}_n)=P(A_1)\cdot\cdots\cdot P(A_k)\cdot P(\bar{A}_{k+1})\cdot\cdots\cdot P(\bar{A}_n)=p^k(1-p)^{n-k}.$$

由于 n 次试验中 A 发生 k 次的方式共有 C_n^k 种，而 C_n^k 种方式对应 C_n^k 个事件. C_n^k 个事件中的任一个发生都可导致事件 A 的发生，且这 C_n^k 个事件两两互不相容，于是有

$$P_n(k)=\mathrm{C}_n^kp^k(1-p)^{n-k},k=0,1,\cdots,n.$$

由于 $\mathrm{C}_n^kp^k(1-p)^{n-k}$ 正好是二项式 $(px+(1-p))^n$ 的展开式中 x^k 的系数，因此称上述公式为**二项概率公式**.

注：在 n 重伯努利试验中，事件 A 至少发生一次的概率为 $1-P_n(0)=1-(1-p)^n$.

例 1.20　一袋中有 4 个红球和 6 个白球，甲、乙两人采用取后放回的方式各取 n 次，每次取 1 个球，求甲取到的红球数与乙取到的白球数相等的概率.

解　设 $A_k=\{$甲取到 k 个红球$\}$，$B_k=\{$乙取到 k 个白球$\}$，$C=\{$甲取到的红球

数与乙取到的白球数相等}.

采用取后放回的方式取 n 次球是一个 n 重伯努利试验，所以

$$P(A_k)=\mathrm{C}_n^k\left(\frac{2}{5}\right)^k\left(\frac{3}{5}\right)^{n-k},P(B_k)=\mathrm{C}_n^k\left(\frac{3}{5}\right)^k\left(\frac{2}{5}\right)^{n-k},$$

又 $C=\bigcup\limits_{k=0}^{n}A_kB_k$，由于 A_k 与 B_k 相互独立，且 A_iB_i 与 $A_jB_j(i\neq j)$ 互不相容，因此事件 C 发生的概率为

$$P(C)=P\left(\bigcup_{k=0}^{n}A_kB_k\right)=\sum_{k=0}^{n}P(A_k)P(B_k)=\left(\frac{6}{25}\right)^n\sum_{k=0}^{n}\mathrm{C}_n^k\mathrm{C}_n^{n-k}=\left(\frac{6}{25}\right)^n\mathrm{C}_{2n}^n.$$

1.5 全概率公式与贝叶斯公式

1.5.1 全概率公式

从已知简单事件的概率推算出未知复杂事件的概率是概率论的重要研究课题之一. 为了计算复杂事件的概率，经常把一个复杂事件分解为若干个互不相容的简单事件的和，通过分别计算简单事件的概率，并利用概率的加法公式和乘法公式等得到最终的结果. 在这类计算中，全概率公式起着重要作用. 下面先介绍样本空间划分的定义.

定义 1.8 设 S 为试验 E 的样本空间，$B_1,B_2,\cdots,B_n$ 为 E 的一组事件. 若有

(1) $B_1\cup B_2\cup\cdots\cup B_n=S$.

(2) $B_iB_j=\varnothing(i\neq j,1\leqslant i,j\leqslant n)$，

则称 $B_1,B_2,\cdots,B_n$ 构成了样本空间 S 的一个**划分**.

若 $B_1,B_2,\cdots,B_n$ 为样本空间 S 的一个划分，那么对于每次试验，事件 $B_1,B_2,\cdots,B_n$ 中有一个且仅有一个发生.

注：划分可看作是对立事件概念上的推广，其特点是“不重不漏”. 对于给定试验的样本空间，其划分是不唯一的.

定理 1.4(全概率公式) 若事件 $B_1,B_2,\cdots,B_n$ 构成了试验 E 的样本空间 S 的一个划分，且 $P(B_i)>0,i=1,2,\cdots,n$，则对 E 的任一事件 A 都有

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A\mid B_i). \tag{1.4}$$

式(1.4) 称为**全概率公式**.

证明 由划分的定义知

$$P(A) = P(A \cap S) = P\Big(A \cap \big(\bigcup_{i=1}^{n} B_i\big)\Big) = P\Big(\bigcup_{i=1}^{n}(A \cap B_i)\Big)$$
$$= \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

特别地，当 $n=2$ 时，若 $0 < P(B) < 1$，记 B_1 为 B，则 B_2 为$\overline{B}$，于是

$$P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B}).$$

全概率公式的直观意义：对于某一复杂事件 A，若 A 的发生有各种可能的原因 B_i，$i=1,2,\cdots,n$，则事件 A 发生的概率即是在各事件 B_i 发生的条件下引起 A 发生的概率的总和.

例 1.21　设在 n 张彩票中有一张奖券，求第二个人摸到奖券的概率.

解　记 $A_i=\{$第 i 个人摸到奖券$\}$，$i=1,2,\cdots,n$，则

$$P(A_1) = \frac{1}{n}, P(\overline{A}_1) = \frac{n-1}{n}, P(A_2 \mid A_1) = 0, P(A_2 \mid \overline{A}_1) = \frac{1}{n-1}.$$

由全概率公式可得

$$P(A_2) = P(A_1)P(A_2 \mid A_1) + P(\overline{A}_1)P(A_2 \mid \overline{A}_1)$$
$$= \frac{1}{n} \cdot 0 + \frac{n-1}{n} \cdot \frac{1}{n-1} = \frac{1}{n}.$$

类似可得

$$P(A_3) = P(A_4) = \cdots = P(A_n) = \frac{1}{n}.$$

如果设 n 张彩票中有 $k(k \leqslant n)$ 张奖券，则

$$P(A_1) = P(A_2) = \cdots = P(A_n) = \frac{k}{n}.$$

这说明，购买彩票时，不论先买后买，中奖的机会是均等的.

例 1.22　设某仓库有一批产品，已知其中甲、乙、丙三个工厂生产的产品依次占 50%，30% 和 20%，且甲、乙、丙厂的次品率分别为$\frac{1}{10}$，$\frac{1}{15}$ 和$\frac{1}{20}$，现从这批产品中任取一件，求取得正品的概率.

解　记 $A_1=\{$取得的这件产品是甲厂生产的$\}$，$A_2=\{$取得的这件产品是乙厂生产的$\}$，$A_3=\{$取得的这件产品是丙厂生产的$\}$，$B=\{$取得的产品是正品$\}$. 于是

$$P(A_1) = \frac{5}{10},\ P(A_2) = \frac{3}{10},\ P(A_3) = \frac{2}{10},$$
$$P(B \mid A_1) = \frac{9}{10},\ P(B \mid A_2) = \frac{14}{15},\ P(B \mid A_3) = \frac{19}{20},$$

由全概率公式知

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + P(B \mid A_3)P(A_3)$$

$$= \frac{9}{10} \times \frac{5}{10} + \frac{14}{15} \times \frac{3}{10} + \frac{19}{20} \times \frac{2}{10} = 0.92.$$

注:利用全概率公式计算复杂事件的概率时,最关键的一步是找样本空间的划分.

例 1.23 要验收一批产品(共 100 件),验收方案如下:从这批产品中随机取 3 件检测(设对 3 件产品的检测是独立的),若 3 件中至少有 1 件在检测中被认为不合格,则这批产品被拒收.设一件不合格品被检测为不合格的概率为 0.95,而一件合格品被误测为不合格品的概率为 0.01,如果已知 100 件产品中恰有 4 件是不合格品,试求这批产品被接收的概率.

解 记 $A=\{$这批产品被接收$\}$,$B_i=\{$随机地取出 3 件产品,其中恰有 i 件不合格品$\}$,$i=0,1,2,3$,则有

$$P(A \mid B_0)=(0.99)^3, P(A \mid B_1)=(0.99)^2 \times 0.05,$$
$$P(A \mid B_2)=0.99 \times (0.05)^2, P(A \mid B_3)=(0.05)^3,$$

而

$$P(B_0)=\frac{C_{96}^3}{C_{100}^3}, P(B_1)=\frac{C_4^1 C_{96}^2}{C_{100}^3}, P(B_2)=\frac{C_4^2 C_{96}^1}{C_{100}^3}, P(B_3)=\frac{C_4^3}{C_{100}^3},$$

由全概率公式知

$$P(A)=\sum_{i=0}^{3} P(A \mid B_i) P(B_i)=0.8629.$$

1.5.2 贝叶斯公式

利用全概率公式,可通过综合分析一事件发生的不同原因及其可能性来求得该事件发生的概率.但在实践中,往往会遇到这样的问题:一事件已经发生了,要考虑该事件发生的各种原因的可能性的大小.例如,在疾病诊断问题中,已知出现某种症状有多种原因,要研究引起这种症状的各种病因的概率是多少,哪种病因的概率最大,以此来作为疾病诊断的依据.对于类似于这种"由果溯因"的推断问题,可用下面的贝叶斯公式来解决.

定理 1.5 若事件 $B_1, B_2, \cdots, B_n$ 构成了试验 E 的样本空间 S 的一个划分,且 $P(B_i)>0, i=1,2,\cdots,n$, A 为 E 的任一事件,$P(A)>0$,则

$$P(B_i \mid A)=\frac{P(B_i) P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j) P(A \mid B_j)}, i=1,2,\cdots,n. \tag{1.5}$$

公式(1.5)称为**贝叶斯公式**,也称为**后验公式**.

证明 由条件概率的定义和全概率公式即得

$$P(B_i \mid A)=\frac{P(AB_i)}{P(A)}=\frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}, i=1,2,\cdots,n.$$

特别地，若取 $n=2$，记 B_1 为 B，则 B_2 为 $\overline{B}$，于是

$$P(B \mid A)=\frac{P(AB)}{P(A)}=\frac{P(B)P(A \mid B)}{P(B)P(A \mid B)+P(\overline{B})P(A \mid \overline{B})}.$$

例 1.24　发报台分别以概率 0.6 和 0.4 发出信号“•”和“—”，由于通讯系统受到干扰，当发出信号“•”时，收报台未必收到信号“•”，而是分别以 0.8 和 0.2 的概率收到“•”和“—”；同样，发出“—”时分别以 0.9 和 0.1 的概率收到“—”和“•”. 如果收报台收到“•”，求它没收错的概率.

解　记 $A=$｛发出信号“•”｝，$B=$｛收到信号“•”｝，则 $\overline{A}=$｛发出信号“—”｝，$\overline{B}=$｛收到信号“—”｝. 于是

$$P(A)=0.6, P(\overline{A})=0.4, P(B \mid A)=0.8,$$
$$P(\overline{B} \mid A)=0.2, P(B \mid \overline{A})=0.1, P(\overline{B} \mid \overline{A})=0.9,$$

由贝叶斯公式知

$$\begin{aligned} P(A \mid B) &= \frac{P(AB)}{P(B)}=\frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\overline{A})P(B \mid \overline{A})} \\ &= \frac{0.6\times 0.8}{0.6\times 0.8+0.4\times 0.1}=\frac{12}{13}, \end{aligned}$$

所以没收错的概率为 $\frac{12}{13}$.

例 1.25　对于例 1.22，若取出的是正品，问该产品是由甲厂生产的概率.

解　沿用例 1.22 的记号，要求的是 $P(A_1 \mid B)$，由贝叶斯公式知

$$\begin{aligned} P(A_1 \mid B) &= \frac{P(A_1B)}{P(B)}=\frac{P(B \mid A_1)P(A_1)}{P(B \mid A_1)P(A_1)+P(B \mid A_2)P(A_2)+P(B \mid A_3)P(A_3)} \\ &= 0.4891. \end{aligned}$$

在此例中，$P(A_1)$ 和 $P(A_1 \mid B)$ 分别称为 A_1 的**先验概率与后验概率**，$P(A_1)$ 是在没有进一步信息（不知道事件 B 是否发生）的情况下 A_1 发生的概率，在获得新的信息（知道 B 发生）后，人们对 A_1 发生的概率可以重新加以修正，得到 $P(A_1 \mid B)$，因而称为后验概率.

本章小结

概率论的研究对象是随机现象，是通过研究随机试验来研究随机现象的. 随机试验的全部可能结果组成的集合称为样本空间. 样本空间作为一个集合，它的子集

称为事件，在一次随机试验中，当且仅当这一子集中的一个样本点出现时，称这一事件发生. 事件之间的关系和运算可按照集合之间的关系和运算来处理.

在一次随机试验中，一个事件可能发生也可能不发生，怎样来描述某一事件在一次试验中发生的可能性的大小?频率是处理这个问题的一个古老、直观的概念. 从频率的稳定性以及频率的性质得到启发，给出了概率的公理化定义. 本章定义了一个集合(事件)的函数 $P(\cdot)$，它满足三条基本性质：非负性、规范性、可列可加性. 这样定义的函数的函数值 $P(A)$ 称为事件 A 的概率.

对于一个事件的概率值具体是多少，概率的定义并没有给出一个通用的计算方法. 对于两种经典的等可能情形的概率模型 —— 古典概型和几何概型，给出了事件概率的求法.

从古典概型和几何概型中得到启发和抽象，给出了条件概率的公式

$$P(A \mid B) = \frac{P(AB)}{P(B)}, P(B) > 0. \tag{1}$$

计算条件概率的时候，通常采用两种方法：① 在样本空间 S 中计算 $P(AB)$ 和 $P(B)$，按式(1) 计算；② 在事件 B 已发生的范围内，求事件 A 的概率，即在缩减的样本空间 $S' = B$ 中计算事件 A 的概率.

式(1) 也可以表示为

$$P(AB) = P(A \mid B)P(B), P(B) > 0, \tag{2}$$

式(2) 称为乘法公式.

一般情况下，条件概率 $P(A \mid B)$ 和无条件概率 $P(A)$ 是不相等的，当两者相等时，称两事件 A 和 B 相互独立. 事件的独立性是概率论中的一个非常重要的概念，概率论和数理统计中很多内容都是在独立的前提下讨论的. 在实际应用中，往往是根据实际背景判断事件的独立性.

在求一个复杂事件 A 的概率时，若 A 的发生有各种可能的原因 B_i，$i = 1, 2, \cdots, n$，则事件 A 发生的概率是各事件 B_i 发生的条件下引起 A 发生的概率的总和，即将复杂事件 A 分解成若干个互不相容的简单事件之和，通过利用全概率公式求解. 若一事件已经发生了，要考虑引发该事件发生的各种原因的可能性的大小，可利用贝叶斯公式求解.

本章知识要点

随机试验　样本空间　随机事件　事件的关系和运算　频率　概率的公理化定义　古典概型　几何概型　条件概率　乘法公式　事件的独立性　伯努利试验　全概率公式　贝叶斯公式

本章常用结论

1）事件的运算律

交换律　$A \cup B = B \cup A, A \cap B = B \cap A$.

结合律　$A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$.

分配律　$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

对偶律(德摩根律)　$\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$;

减法运算　$A - B = A\overline{B} = A - AB$.

2）概率的性质

(1) $P(\varnothing) = 0, P(S) = 1$.

(2) 对于任意一个事件 A,有 $0 \leqslant P(A) \leqslant 1$.

(3) 有限可加性　若事件 $A_1, A_2, \cdots, A_n$ 两两互不相容,则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

(4) 可列可加性　对可列个两两互不相容的事件 $A_1, A_2, \cdots, A_n, \cdots$,有

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i).$$

(5) 减法公式　若 A, B 为任意两个事件,则

$$P(A - B) = P(A) - P(AB).$$

特别地,若 $B \subset A$,则有

$$P(A - B) = P(A) - P(B).$$

(6) 加法公式　若 A, B 为任意两个事件,则

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

(7) $P(\overline{A}) = 1 - P(A)$.

3）古典概型概率的计算

古典概型中事件 A 的概率为

$$P(A) = \frac{A\text{ 包含的基本事件数}}{S\text{ 中基本事件的总数}}.$$

4）条件概率与乘法公式

设 A, B 为两个事件,且 $P(B) > 0$,则

$$P(A \mid B) = \frac{P(AB)}{P(B)},$$

此时

$$P(AB)=P(B)P(A\mid B)(P(B)>0).$$

5）事件独立性的结论

（1）事件 A 与 B 独立 $\Leftrightarrow P(AB)=P(A)P(B)$.

（2）若事件 A 与 B 独立，则 A 与 $\overline{B}$ 独立，$\overline{A}$ 与 B 独立，$\overline{A}$ 与 $\overline{B}$ 独立.

（3）若事件 $A_1,A_2,\cdots,A_n$ 相互独立，则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n),$$

$$P(\bigcup_{i=1}^{n}A_i)=1-\prod_{i=1}^{n}P(\overline{A_i}).$$

6）伯努利试验的概率公式

在 n 重伯努利试验中，设每次试验中事件 A 发生的概率为 $p(0<p<1)$，则在 n 次试验中，事件 A 恰好发生 k 次的概率为

$$P_n(k)=\mathrm{C}_n^k p^k(1-p)^{n-k},k=0,1,\cdots,n.$$

7）全概率公式与贝叶斯公式

若事件 $B_1,B_2,\cdots,B_n$ 构成了试验 E 的样本空间 S 的一个划分，且 $P(B_i)>0$，$i=1,2,\cdots,n$，则对 E 的任一事件 A 都有

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i).$$

若有 $P(A)>0$，则

$$P(B_i\mid A)=\frac{P(AB_i)}{P(A)}=\frac{P(B_i)P(A\mid B_i)}{\sum_{j=1}^{n}P(B_j)P(A\mid B_j)},i=1,2,\cdots,n.$$

习题 1

（1）填空题.

① 已知 A,B 两个事件满足条件 $P(AB)=P(\overline{A}\,\overline{B})$，且 $P(A)=p$，则 $P(B)=$ ________.

② 设 A,B 是相互独立的随机事件，A,B 都不发生的概率为 $\frac{1}{16}$，A 发生 B 不发生的概率等于 A 不发生而 B 发生的概率，则 $P(A)=$ ________.

③ 一个小组有 8 名学生，每个学生在一周 7 天的任何一天出生是等可能的，则 8 个人的生日都不在星期一的概率为________.

④ 设工厂甲和工厂乙的产品的次品率分别为 1% 和 2%，现从由甲和乙的产品

分别占 60% 和 40% 的一批产品中随机抽取一件，发现是次品，则该次品属甲生产的概率是________.

⑤ 从数 1，2，3，4 中任取一数，记为 X，再从 $1,2,\cdots,X$ 中任取一个数，记为 Y，则 Y 是 2 的概率为________.

(2) 选择题.

① 设随机事件 A,B 同时发生时，事件 C 一定发生，则(　　).

A. $P(C)=P(AB)$　　B. $P(C)=P(A\cup B)$

C. $P(C)\leqslant P(A)+P(B)-1$　　D. $P(C)\geqslant P(A)+P(B)-1$

② 随机事件 A,B 满足 $P(A)=P(B)=\frac{1}{2}$，$P(A\cup B)=1$，则下列结论正确的是(　　).

A. $A\cup B=S$　　B. $AB=\varnothing$

C. $P(A-B)=P(A)$　　D. $P(\overline{A}\cup\overline{B})=0$

③ 袋中有 5 只球，其中 3 只红球，2 只黑球，先后从袋中取 3 次球，每次任取 1 只不放回，则第二次取到红球的概率为(　　).

A. $\frac{1}{5}$　　B. $\frac{2}{5}$

C. $\frac{3}{5}$　　D. $\frac{4}{5}$

④ 设随机事件 A,B 互不相容，且 $P(A)>0$，$P(B)>0$，则下列结论正确的是(　　).

A. $\overline{A}$ 与 $\overline{B}$ 互不相容　　B. $\overline{A}$ 与 $\overline{B}$ 不是互不相容的

C. $P(AB)=P(A)P(B)$　　D. $P(A-B)=P(A)$

⑤ 某人向同一目标独立重复射击，每次射击命中目标的概率为 $p(0<p<1)$，则此人第 4 次射击恰好第 2 次命中目标的概率为(　　).

A. $3p(1-p)^2$　　B. $6p(1-p)^2$

C. $3p^2(1-p)^2$　　D. $6p^2(1-p)^2$

(3) 袋中有 10 个球，分别标有号码 1～10，从中任取一球，设 $A=\{$取到的球的号码是偶数$\}$，$B=\{$取到的球的号码是奇数$\}$，$C=\{$取到的球的号码小于 5$\}$，试问下列运算分别表示什么事件：

① $A\cup B$；② $\overline{B\cup C}$；③ AB；④ AC；⑤ $\overline{A}\,\overline{C}$.

(4) 若要击落飞机必须同时击毁两个发动机或击毁驾驶舱，记 $A_1=\{$击毁第一个发动机$\}$，$A_2=\{$击毁第二个发动机$\}$，$B=\{$击毁驾驶舱$\}$，试用 A_1,A_2,B 表示事件“飞机被击落”.

(5) 若$P(A)=0.5$,$P(B)=0.4$,$P(A-B)=0.3$,求$P(A\cup B)$和$P(\overline{A}\cup\overline{B})$.

(6) 设A,B,C是三个事件,且$P(A)=P(B)=P(C)=\frac{1}{4}$,$P(AB)=P(BC)=0$,$P(AC)=\frac{1}{8}$,求$A,B,C$至少有一个发生的概率.

(7) 将3个球随机放入4个杯子中,问杯子中球的个数最多为1,2,3的概率各是多少?

(8) 某个家庭有两个小孩,求至少有一个女孩的概率(设男女出生率相同).

(9) 一盒子中有9张卡片分别编号1,2,…,9,采用取后放回的方式,连续取3张卡片,求前两次取到卡片编号为偶数最后取到卡片编号为奇数的概率.

(10) 掷两颗均匀的骰子,求出现点数之和为8的概率.

(11) 口袋中有4个白球,6个红球.从中任取2个,求取到1个白球1个红球的概率.

(12) 一赌徒认为掷一颗骰子4次至少出现一次6点与掷两颗骰子24次至少出现一次双6点的机会是相等的,你认为如何?

(13) 将15名新生(其中有3名优秀生)随机地分配到三个班级中,其中一班4名,二班5名,三班6名,求:

① 每一个班级各分配到1名优秀生的概率;

② 3名优秀生被分配到一个班级的概率.

(14) 一个袋子中装有$a+b$只球,其中a只黑球,b只白球,随意的每次从中取出一只球(不放回),求下列各事件的概率:

① 第i次取到的是黑球;

② 第i次才取到黑球.

(15) 有n个球,每个球都等可能地被放到N个不同盒子中的任一个,每个盒子所放球数不限,试求:

① 指定的$n(n\leqslant N)$个盒子中各有一球的概率$P(A)$;

② 恰好有$n(n\leqslant N)$个盒子中各有一球的概率$P(B)$.

(16) 从1,2,…,9中任取1个数,取后放回,先后取出5个数,求下列事件的概率:

①“最后取出的数字是奇数”;

②“5个数字全不同”;

③“1恰好出现2次”;

④“1至少出现2次”;

⑤“恰好出现不同的2对数字”;

⑥“总和为 10”.

(17) 在一个有 n 个人参加的晚会上，每个人带了一件礼物，且假定每个人带的礼物都不相同. 晚会期间各人从放在一起的 n 个礼物中随机地抽取一件，求至少有一个人自己抽到自己的礼物的概率.

(18) 从 5 双不同的鞋子中任取 4 只，求这 4 只鞋子中至少有 2 只能配成 1 双的概率.

(19) 500 km^2 的海域里有面积达 40 km^2 的大陆架藏着石油. 在此海域里任选一点钻探，求钻到石油的概率.

(20) 一个家庭有两个小孩，问该家在有一个女孩的条件下，另一个也是女孩的概率是多少？

(21) 设某光学仪器厂制造的透镜，第一次落下时被打破的概率为 $\frac{1}{2}$；若第一次落下未被打破，第二次落下时被打破的概率为 $\frac{7}{10}$；若前两次落下都未被打破，第三次落下时被打破的概率为 $\frac{9}{10}$. 试求透镜落下三次而未被打破的概率.

(22) 播种用的一等小麦种子中混有 5% 的二等种子和 10% 的三等种子，这三个等级的种子长出的穗含有 50 颗以上麦粒的概率分别为 0.90，0.80 和 0.65. 求这批种子任选一粒，长出的穗含有 50 颗以上麦粒的概率.

(23) 甲袋中有 6 个红球和 4 个白球，乙袋中有 12 个红球和 8 个白球，现从两袋子中任选一个袋子，然后采用不放回的方式连续取 2 个球，求第一次取到白球的概率.

(24) 某彩票每周开奖一次，每次提供十万分之一的中奖机会，且各周开奖是相互独立的. 某彩民每周买一次彩票，坚持十年(每年 52 周)，那么他从未中奖的可能性是多少？

(25) A 系与 B 系举行篮球、排球、足球比赛，篮球赛 A 胜 B 的概率为 0.8，排球赛 A 胜 B 的概率为 0.4，足球赛 A 胜 B 的概率为 0.4，若在三项比赛中至少胜两项才算获胜，试计算哪个系获胜的概率较大.

(26) 对以往数据分析结果表明，当机器调整得良好时，产品的合格率为 98%，而当机器发生某种故障时，其合格率为 55%. 每天早上机器开动时，机器调整良好的概率为 95%. 试求已知某日早上第一件产品是合格时，机器调整得良好的概率.

(27) 有两箱零件，第一箱装 50 件，其中 10 件一等品；第二箱装 30 件其中 18 件一等品，现从两箱中随意挑出一箱，然后从该箱中任取 2 个零件，试求：

① 第一次取出的零件是一等品的概率；

② 在第一次取出的是一等品的条件下，第二次取出的仍是一等品的概率.

(28) 假设每个人的血清中含有肝炎病毒的概率为 0.4%，100 人的血清混合在一起，求此中含有肝炎病毒的概率.

(29) 一猎人用猎枪射击一只兔子，第一枪距离兔子 200 m 远，如果未击中，他追到离兔子 150 m 远处进行第二次射击，如果仍未击中，他追到离兔子 100 m 远处再进行第三次射击，此时击中的概率为 $\frac{1}{2}$，如果这个猎人射击的命中率与他离兔子的距离的平方成反比，求猎人击中兔子的概率.

(30) 根据临床记录，某种诊断癌症的试验有如下的效果：若记 $A=\{$试验反应为阳性$\}$，$C=\{$被诊断者患有癌症$\}$，则有 $P(A\mid C)=0.95$，$P(\overline{A}\mid\overline{C})=0.95$. 现在对自然人群进行普查，设被试验的人患有癌症的概率为 0.005，即 $P(C)=0.005$，试求 $P(C\mid A)$.

(31) 一个人的血型为 A，B，AB，O 型的概率分别为 0.37，0.21，0.08，0.34. 现任意挑选四个人，试求：

① 此四人的血型全不相同的概率；

② 此四人的血型全部相同的概率.

(32) 甲、乙两选手进行乒乓球单打比赛，已知在每局中甲胜的概率为 0.6，乙胜的概率为 0.4，比赛可采用三局二胜制或五局三胜制，问哪一种比赛制度对甲更有利？

第 2 章

随机变量及其分布

在第 1 章中用样本空间的子集表示随机事件，这种表示方法对分析随机现象的统计规律有较大的局限性. 为了更好地利用数学工具来研究随机试验的结果，将随机试验的结果数量化，由此就产生了随机变量的概念. 本章将用随机变量来表示随机事件，它对概率论的研究有着十分重要的作用.

2.1　随 机 变 量

随机试验的结果有些本身就是数量. 例如，掷骰子出现的点数、产品抽样中的次品数、电话总机在单位时间内接到的呼叫次数、测量中出现的误差等. 而有些随机试验表面上看其试验结果与数量没有直接关系. 例如，掷硬币的结果“正面”、“反面”. 我们可以将其数量化，当出现正面时对应数 1，出现反面时对应数 0. 这样随机试验的结果就是随机变化的变量，把随机试验的结果数量化，便于应用数学知识研究随机现象，使对随机现象的研究更深入和简单.

例 2.1　抛掷一枚硬币两次，观察出现正面(记为 H) 和反面(记为 T) 的情况.

样本空间是 $S = \{HH, HT, TH, TT\}$. 以 X 表示抛掷两次硬币中出现正面 H 的次数，那么对于样本空间中的每一个样本点 e，X 都有一个数与之对应. X 是定义在样本空间 S 上的一个实值单值函数，它的定义域是样本空间 S，值域是实数集合 $\{0,1,2\}$，使用函数记号将 X 写成

$$X = X(e) = \begin{cases} 0, & e = TT, \\ 1, & e = HT \text{ 或 } TH, \\ 2, & e = HH. \end{cases}$$

例 2.2 测试灯泡的寿命.

样本空间是 $S=\{t \mid t\geqslant 0\}$. 每一个灯泡的实际使用寿命可能是$[0,+\infty)$中的任何一个实数. 以 X 表示灯泡的寿命(小时),它的数值由试验的结果来确定,即 X 是定义在样本空间 S 上的函数,定义域是样本空间 S,值域为$[0,+\infty)$.

上述例子说明,随机试验的结果总能够用一个实数来表示,这个数随试验结果的不同而变化,因而它是样本点的函数,这个函数就是要引入的随机变量.

定义 2.1 设随机试验的样本空间为 S,$X=X(e)$ 是定义在样本空间 S 上的实值单值函数,称 $X=X(e)$ 为**随机变量**,通常用大写字母 $X,Y,Z,\cdots$ 或希腊字母 $\xi,\eta,\zeta,\cdots$ 表示,而随机变量所取的值一般用小写字母 $x,y,z,\cdots$ 表示.

图 2-1 所示为样本点 e 与实数 $X=X(e)$ 对应的示意图.

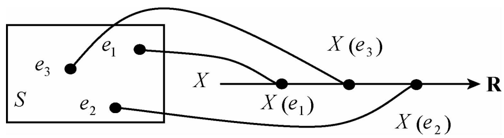


图 2-1

随机变量的取值随试验的结果而定. 在试验之前不能预知它取什么值,且它的取值有一定的概率,这些性质显示出随机变量与普通函数有着本质的区别.

引入随机变量后,便可以用随机变量来描述事件.

例如,掷一枚骰子朝上一面的点数 X 是一个随机变量. 事件"朝上一面的点数为 2"可表示为$\{X=2\}$;事件"朝上一面的点数至多为 3"可表示为$\{X\leqslant 3\}$. $P\{X=2\}=\dfrac{1}{6}$,$P\{X\leqslant 3\}=\dfrac{1}{2}$ 分别表示两事件发生的概率.

一般地,对任意实数集 I,随机变量 X 在 I 上取值常写成$\{X\in I\}$,它表示事件$\{e \mid X(e)\in I\}$,此时有 $P\{X\in I\}=P\{e \mid X(e)\in I\}$.

随机变量的产生使人们可利用数学分析的方法对随机试验的结果进行广泛而深入的研究,概率论的研究由古典概率时期进入分析概率时期,概率论的研究取得了飞速的进展.

随机变量因取值方法不同,通常分为离散型和非离散型两类. 而非离散型随机变量中最重要的是连续型随机变量. 本书主要讨论离散型随机变量和连续型随机变量.

2.2　离散型随机变量及其分布

2.2.1　离散型随机变量及其分布律

定义 2.2　若随机变量 X 的所有可能取值是有限个或可列无限多个，则称此随机变量为**离散型随机变量**.

例如，掷骰子朝上一面的点数、一昼夜 120 接到的呼叫次数等均为离散型随机变量，而某元件寿命的所有可能取值充满一个区间，无法按一定次序一一列举出来，因而它是一个非离散型随机变量.

定义 2.3　设离散型随机变量 X 所有可能取值为 $x_i(i=1,2,3,\cdots)$，则称 X 取 x_i 的概率

$$P\{X=x_i\}=p_i,i=1,2,3,\cdots$$

为 X 的**分布律(列)**或**概率分布**.

分布律也可以表示为表格形式

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

或矩阵形式

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}.$$

由概率的定义，分布律具有以下性质：

(1) **非负性**　$p_i \geqslant 0(i=1,2,3,\cdots)$.

(2) **归一性**　$\sum\limits_{i=1}^{+\infty} p_i = 1$.

证明　显然，(1) 成立.

(2) 由于 $\{X=x_i\} \cap \{X=x_j\} = \varnothing(i \neq j)$，$\bigcup\limits_{i=1}^{+\infty}\{X=x_i\}$ 是必然事件，所以

$$P\left(\bigcup_{i=1}^{+\infty}\{X=x_i\}\right)=\sum_{i=1}^{+\infty}P\{X=x_i\}=\sum_{i=1}^{+\infty}p_i=1.$$

注：① 上述两条性质是分布律必须具有的性质. 如果一个数列 $\{p_i\}$ 具有以上两条性质，则它可以作为某离散型随机变量的分布律.

② 利用上述性质，可以验证随机变量的分布律的正确性.

分布律能够完全刻画离散型随机变量的统计规律性.只要给出随机变量的分布律,利用概率的可列可加性,任一事件发生的概率都可以求出来.例如:

$$P\{a \leqslant X \leqslant b\} = P\Big(\bigcup_{a \leqslant x_i \leqslant b}\{X = x_i\}\Big) = \sum_{a \leqslant x_i \leqslant b} p_i;$$

$$P\{X \in I\} = \sum_{x_i \in I} p_i.$$

例 2.3 一个贸易总公司分别为下设的两个分公司下达销售任务,两个分公司完成任务的概率分别为 0.8 和 0.4.若完成任务,那么两个分公司将分别获得奖金 4 万 元和 6 万元.设 X 为总公司应付出的奖金,求 X 的分布律并计算 $P\{4 \leqslant X < 10\}$ 和 $P\{X \leqslant 6\}$.

解 X 的所有可能取值为 0,4,6,10(单位:万元).设 $A_i = \{$第 i 个分公司获得奖金$\}(i = 1,2)$,则 $P(A_1) = 0.8$,$P(A_2) = 0.4$,且 A_1,A_2 相互独立.因此

$$P\{X = 0\} = P(\overline{A_1}\,\overline{A_2}) = P(\overline{A_1})P(\overline{A_2}) = 0.12,$$

$$P\{X = 4\} = P(A_1\overline{A_2}) = P(A_1)P(\overline{A_2}) = 0.48,$$

$$P\{X = 6\} = P(\overline{A_1}A_2) = P(\overline{A_1})P(A_2) = 0.08,$$

$$P\{X = 10\} = P(A_1A_2) = P(A_1)P(A_2) = 0.32,$$

即 X 的分布律为

X	0	4	6	10
P	0.12	0.48	0.08	0.32

于是

$$P\{4 \leqslant X < 10\} = P\{X = 4\} + P\{X = 6\} = 0.56,$$

$$P\{X \leqslant 6\} = 1 - P\{X > 6\} = 1 - P\{X = 10\} = 0.68.$$

2.2.2 常用离散型随机变量的分布

1)(0-1) 分布

定义 2.4 若随机变量 X 只可能取 0 与 1 两个值,它的分布律是

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

即

X	0	1
P	$1-p$	p

则称 X 服从以 p 为参数的**(0-1) 分布**或**两点分布**.

对于一个随机试验，若它的样本空间只包含两个样本点，即 $S=\{e_1,e_2\}$，则总能在 S 上定义一个服从(0-1)分布的随机变量 $X=X(e)=\begin{cases}0, & e=e_1,\\ 1, & e=e_2,\end{cases}$ 用其来描述这个随机试验的结果. 例如，抛掷一枚硬币出现正面或反面；检查产品的质量是否合格等都可以用(0-1)分布来描述.

若对于一个随机试验，它的可能结果不止两个，但人们只关心其中的事件 A 发生或不发生，且 $P(A)=p(0<p<1)$，若设 $\{X=1\}=A$，$\{X=0\}=\overline{A}$，则 X 为服从(0-1)分布的随机变量. 例如，打靶是否中靶；掷骰子是否出现 1 点等. (0-1)分布是经常用到的一种分布.

例 2.4　50 件产品中有 45 件正品，5 件次品，从中任取 1 件，若规定 $X=\begin{cases}0, & \text{取到次品},\\ 1, & \text{取到正品},\end{cases}$ 试求 X 的分布律.

解　取得次品的概率，即

$$P\{X=0\}=\frac{5}{50}=\frac{1}{10},$$

取得正品的概率，即

$$P\{X=1\}=\frac{45}{50}=\frac{9}{10},$$

于是，X 是服从 $p=\frac{9}{10}=0.9$ 的(0-1)分布.

2）二项分布

由第 1 章介绍的 n 重伯努利实验知，若事件 A 在每次试验中发生的概率为 $P(A)=p(0<p<1)$，设 $A_k=\{n$ 次独立重复试验 A 恰好发生 k 次$\}$，$k=0,1,2,\cdots,n$，则

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

设 X 表示 n 重伯努利试验中事件 A 发生的次数，则 X 的可能取值为 $0,1,2,\cdots,n$，$\{X=k\}=A_k$.

定义 2.5　设随机变量 X 的可能取值为 $0,1,2,\cdots,n$，其分布律为

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

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

显然，

(1) $P(X=k)\geqslant 0(k=0,1,2,\cdots,n)$.

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

这里 $C_n^k p^k (1-p)^{n-k}$ 恰好是二项式 $(p+q)^n (q=1-p)$ 展开式中含 p^k 的那项，故名二项分布.

特别地，当 $n=1$ 时，二项分布即为(0-1)分布，故(0-1)分布可记为 $X \sim b(1,p)$. 实际上，二项分布是 n 重伯努利试验的概率模型，是一种常用的离散分布.

例 2.5 一份考卷中有10道选择题，每题有4个可能答案，其中只有1个答案是正确的.

(1) 某学生随机猜测，求他答对题数的分布律及至少能答对2道题的概率.

(2) 若一人答对6道题，则推测他是猜对的还是有答题能力.

解 设 X 表示该学生靠猜测答对的题数，则 $X \sim b\left(10,\frac{1}{4}\right)$.

(1) X 的分布律为

$$P\{X=k\}=C_{10}^k\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{10-k}, k=0,1,2,\cdots,10,$$

于是

$$\begin{aligned}P\{X\geqslant 2\}&=1-P\{X=0\}-P\{X=1\}\\&=1-C_{10}^0\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10}-C_{10}^1\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^9=0.756.\end{aligned}$$

(2) 因为

$$P\{X=6\}=C_{10}^6\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^4=0.016,$$

即单靠猜测答对6道题的可能性是0.016，概率很小，所以由实际推断原理(即概率很小的事件在一次试验中实际上几乎是不发生的)可推测，此学生是有答题能力的.

例 2.6 某种技术竞赛有5个项目，参赛者若有3项以上取得优秀就可以取得技术过硬荣誉证书. 某单位选派了4人参加竞赛. 已知每个人每项取优的概率为0.8，求下列事件的概率：

(1) 1个参赛者获得荣誉证书.

(2) 该单位至少有2个人获得荣誉证书.

解 (1) 1个参赛者参加5项技术竞赛相当于进行5次独立试验，每项竞赛只有两个可能结果，即取得优秀和没取得优秀. 设 $A=\{$取得优秀$\}$，则 $P(A)=p=0.8$；X 为参赛者在5个项目中取得优秀的项目数，则 $X \sim b(5,0.8)$.

于是，1个参赛者获得荣誉证书的概率

$P\{X\geqslant 3\}=C_5^3(0.8)^3(0.2)^2+C_5^4(0.8)^4(0.2)+C_5^5(0.8)^5=0.942\,1.$

(2) 4个参赛者的竞赛成绩是4次独立试验的结果，每个参赛者只有两种可能结果，即获得荣誉证书和未获得荣誉证书. 设 $B=\{$获得荣誉证书$\}$，则 $P(B)=0.942\,1$；设 Y 为4人中获得荣誉证书的人数，则 $Y \sim b(4,0.942\,1)$.

故该单位至少有 2 人获得荣誉证书的概率

$$
\begin{aligned}
P\{Y \geqslant 2\} &= 1 - P\{Y = 0\} - P\{Y = 1\} \\
&= 1 - C_4^0 (0.057\,9)^4 - C_4^1 (0.942\,1)(0.057\,9)^3 = 0.999\,3.
\end{aligned}
$$

二项分布 $b(n,p)$ 和(0-1) 分布 $b(1,p)$ 还有一层密切关系. 设一个随机试验只有两个结果 A 和 $\overline{A}$，且 $P(A) = p$，现将试验独立进行 n 次，记 X 为 n 次试验中 A 出现的次数，则 $X \sim b(n,p)$，记 X_i 为第 i 次试验中 A 出现的次数，即 $X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中 } A \text{ 出现}, \\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不出现}, \end{cases}$ $i = 1,2,\cdots,n$，则 $X_i \sim b(1,p)$，并且 $X_1, X_2, \cdots, X_n$ 相互独立，由 X 和 $X_1, X_2, \cdots, X_n$ 的定义，显然有 $X = X_1 + X_2 + \cdots + X_n$.

3）泊松分布

定义 2.6　设随机变量 X 所有可能取值为 $0,1,2,\cdots$，其分布律为

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

则称 X 服从参数为 λ 的**泊松分布**，记为 $X \sim \pi(\lambda)$ 或 $X \sim P(\lambda)$.

显然，

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

(2) $\displaystyle\sum_{k=0}^{+\infty} P\{X = k\} = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$

定理 2.1（泊松定理）　对二项分布 $b(n,p)$，设 $np = \lambda, \lambda > 0$，则对于任一固定的非负整数 k，有

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

证明　由 $p = \dfrac{\lambda}{n}$，有

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

对任意固定的非负整数 k，当 $n \to +\infty$ 时，

$$
1 \cdot \left(1 - \frac{1}{n}\right) \cdot \cdots \cdot \left(1 - \frac{k-1}{n}\right) \to 1, \left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1,
$$

故有 $\displaystyle\lim_{n \to +\infty} C_n^k p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$.

由泊松定理可知，泊松分布是二项分布的极限分布. 对于 $X \sim b(n,p)$，当 n 充分大而 p 又很小($n \geqslant 20, p \leqslant 0.05$) 时，有以下近似公式

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} (\lambda = np), \tag{2.1}$$

即以 n,p 为参数的二项分布的概率值可以由参数为 $\lambda = np$ 的泊松分布的概率值近似，而泊松分布的计算可借助于泊松分布数值表直接查得(见附表 2).

历史上，泊松分布是作为二项分布的近似，于 1837 年由法国数学家泊松(Poisson)引入的. 泊松分布是概率中最重要的分布之一，现已发现许多随机现象都服从或近似服从泊松分布. 例如，服务系统中对服务的呼叫数、公共汽车站候车的人数、布匹上的疵点数、放射性物质分裂落在某区域的质点数、显微镜下落在某区域中血球或微生物数目等，都是泊松分布的概率模型.

例 2.7 某种铸件的砂眼数服从参数为 $\lambda = 0.5$ 的泊松分布，试求该铸件至多有 1 个砂眼的概率和至少有 2 个砂眼的概率.

解 以 X 表示铸件的砂眼数，由题意知 $X \sim \pi(0.5)$，则该种铸件上至多有 1 个砂眼的概率

$$\begin{aligned} P\{X \leqslant 1\} &= P\{X = 0\} + P\{X = 1\} \\ &= \frac{0.5^0}{0!} e^{-0.5} + \frac{0.5}{1!} e^{-0.5} = 0.91, \end{aligned}$$

至少有 2 个砂眼的概率

$$P\{X \geqslant 2\} = 1 - P\{X < 2\} = 1 - P\{X \leqslant 1\} = 0.09.$$

例 2.8 某一交通路口每天有大量汽车通过，设每辆汽车在每天某段时间内发生事故的概率是 0.000 1，设在该路口的某个时间段内有若干辆汽车通过，求：

(1) 若有 1 500 辆汽车通过，不发生事故及发生事故次数小于 2 的概率.

(2) 若有 20 000 辆汽车通过，则至少发生 1 次事故的概率.

解法一 (1) 用 X 表示发生事故的次数，则 $X \sim b(1\,500, 0.000\,1)$，

$$P\{X = 0\} = C_{1\,500}^0 (0.000\,1)^0 (1 - 0.000\,1)^{1\,500} = 0.860\,7,$$

$$\begin{aligned} P\{X < 2\} &= P\{X = 0\} + P\{X = 1\} \\ &= C_{1\,500}^0 (0.000\,1)^0 (1-0.000\,1)^{1\,500} + C_{1\,500}^1 (0.000\,1)^1 (1-0.000\,1)^{1\,500-1} \\ &= 0.989\,8. \end{aligned}$$

(2) 用 X 表示发生事故的次数，则 $X \sim b(20\,000, 0.000\,1)$，

$$\begin{aligned} P\{X \geqslant 1\} &= 1 - P\{X = 0\} \\ &= 1 - C_{20\,000}^0 (0.000\,1)^0 (1 - 0.000\,1)^{20\,000} = 0.864\,7. \end{aligned}$$

解法二 利用式(2.1)计算.

(1) $\lambda = 1\,500 \times 0.000\,1 = 0.15, P\{X = 0\} = e^{-0.15} = 0.860\,7,$

$P\{X < 2\} = P\{X = 0\} + P\{X = 1\} \approx e^{-0.15} + 0.15e^{-0.15} = 0.989\,8.$

(2) $\lambda = 20\,000 \times 0.000\,1 = 2, P\{X \geqslant 1\} = 1 - P\{X = 0\} \approx 1 - e^{-2} = 0.864\,7.$

显然，用式(2.1)计算较为简便.

例 2.8 中每辆车发生事故的概率非常小，当某段时间内通过汽车为 1 500 辆时，至少发生一次事故的可能性为 0.139 3，发生事故的概率还是很小. 但是当汽车量达到 20 000 辆时，至少发生一次事故的可能性达到 0.864 7，发生事故的概率就比较大了，还可以计算如果汽车量达到 30 000 辆时，至少发生一次事故的可能性达到 0.95，事故几乎总有发生了. 此例说明，小概率事件也是不可忽视的.

2.3　随机变量的分布函数

2.3.1　分布函数

对于离散型的随机变量，分布律便能完全刻画其统计规律性，但对于非离散型的随机变量，由于其可能取值不能一一列举出来，因而就不能采用分布律来描述. 另外非离散型随机变量通常取任一指定值的概率都等于 0(下一节将给出证明)，在实际中对于非离散型的随机变量，如测量的误差、元件的寿命等，并不会对它们取某值的概率感兴趣，而是考察误差落在某个区间内的概率或寿命大于某数的概率. 因而，需要研究随机变量取值落在一个区间内的概率问题 $P\{x_1 < X \leqslant x_2\}$，而$P\{x_1 < X \leqslant x_2\} = P\{X \leqslant x_2\} - P\{X \leqslant x_1\}$，所以只需要知道 $P\{X \leqslant x_1\}$ 和$P\{X \leqslant x_2\}$ 即可. 为了更普遍地研究随机变量的分布，把讨论事件$\{X = x_i\}$的概率转为研究事件$\{X \leqslant x\}$的概率，其中 x 为任意实数，显然这个事件的概率依赖于 x 的变化而变化，是 x 的一个函数，这个函数称为随机变量的分布函数，一般定义如下.

定义 2.7　设 X 是一个随机变量，对于任意实数 x，称函数

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

为 X 的**概率分布函数**，简称**分布函数**，记为 $X \sim F(x)$ 或 $F_X(x)$.

对任意实数 $x_1, x_2\,(x_1 < x_2)$，随机点落在区间$(x_1, x_2]$内的概率为

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

因此，如果已知 $F(x)$，便可求出随机变量 X 的值落在任一区间$(x_1, x_2]$上的概率，从这个意义上讲，分布函数描述了随机变量的统计规律性.

随机变量的分布函数是一个普通的函数，其定义域为$(-\infty, +\infty)$，值域为$[0, 1]$，由于分布函数的引入使得人们能用数学分析的方法研究随机变量.

若将 X 看作数轴上随机点的坐标，那么分布函数 $F(x)$ 在 x 处的值就表示 X 落在区间$(-\infty, x]$上的概率.

分布函数 $F(x)$ 的基本性质：

(1) **有界性**　对于任意实数 x，$0\leqslant F(x)\leqslant 1$.

事实上，对于任意实数 x，函数 $F(x)$ 的值是事件 $\{X\leqslant x\}$ 发生的概率，而概率总在 0 与 1 之间.

(2) $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$，$F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

事实上，当 $x\to-\infty$ 时，事件 $\{X\leqslant x\}$ 趋于不可能事件，故 $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$；当 $x\to+\infty$ 时，事件 $\{X\leqslant x\}$ 趋于必然事件，故 $F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

(3) **单调性**　$F(x)$ 是单调非减函数，即对于任意 $x_1<x_2$，有 $F(x_1)\leqslant F(x_2)$.

事实上，对于任意 $x_1<x_2$，事件 $\{X\leqslant x_1\}$ 包含于事件 $\{X\leqslant x_2\}$，故

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

(4) **右连续**　$F(x)=F(x+0)$（证明略）.

注：① 如果一个函数同时满足上述四条性质，则该函数一定是某个随机变量的分布函数.

② 利用分布函数 $F(x)$，可以很方便地求出下列事件的概率：

$$P\{X>a\}=1-P\{X\leqslant a\}=1-F(a),$$
$$P\{X<a\}=\lim_{x\to a^-}F(x)=F(a-0),$$
$$P\{X=a\}=F(a)-F(a-0),$$

其中 a 是任意给定的实数.

例 2.9　向半径为 r 的圆内随机抛一点，求此点到圆心的距离 X 的分布函数，并求 $P\left\{X>\dfrac{r}{2}\right\}$.

解　若 $x<0$，$\{X\leqslant x\}$ 是不可能事件，则 $F(x)=P\{X\leqslant x\}=0$；若 $x\geqslant r$，$\{X\leqslant x\}$ 是必然事件，则 $F(x)=P\{X\leqslant x\}=1$；若 $0\leqslant x<r$，由几何概型知

$$F(x)=P\{X\leqslant x\}=\frac{\pi x^2}{\pi r^2}=\left(\frac{x}{r}\right)^2.$$

从而，X 的分布函数

$$F(x)=\begin{cases}0, & x<0,\\ \left(\dfrac{x}{r}\right)^2, & 0\leqslant x<r,\\ 1, & x\geqslant r,\end{cases}$$

且

$$P\left\{X>\frac{r}{2}\right\}=1-P\left\{X\leqslant\frac{r}{2}\right\}=1-F\left(\frac{r}{2}\right)=1-\left(\frac{1}{2}\right)^2=\frac{3}{4}.$$

例 2.10　验证 $F(x)=\frac{1}{\pi}\left(\arctan x+\frac{\pi}{2}\right),-\infty<x<+\infty$ 是一个分布函数.

证明　对 $\forall x\in(-\infty,+\infty),0\leqslant F(x)\leqslant 1,F(x)$ 连续且单调递增，

$$F(-\infty)=0,F(+\infty)=1,$$

因此，它满足分布函数的四条性质，故 $F(x)$ 是一个分布函数，该函数称为**柯西分布函数**.

2.3.2　离散型随机变量的分布函数

由分布函数 $F(x)$ 的定义，若已知离散型随机变量 X 的分布律

$$P\{X=x_i\}=p_i(i=1,2,3,\cdots),$$

则它的分布函数 $F(x)=P\{X\leqslant x\}=\sum\limits_{x_i\leqslant x}P\{X=x_i\}=\sum\limits_{x_i\leqslant x}p_i$，即

当 $x<x_1$ 时，$F(x)=0$；

当 $x_1\leqslant x<x_2$ 时，$F(x)=p_1$；

当 $x_2\leqslant x<x_3$ 时，$F(x)=p_1+p_2$；

…

当 $x_{n-1}\leqslant x<x_n$ 时，$F(x)=p_1+p_2+\cdots+p_{n-1}$；

…

如图 2-2 所示，$F(x)$ 是一个右连续的阶梯函数，它在 $X=x_i(i=1,2,3,\cdots)$ 处有跳跃，其跳跃度恰为随机变量 X 在 $X=x_i$ 点的概率 p_i.

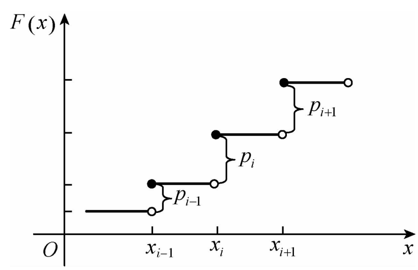


图 2-2

反之，若一个随机变量 X 的分布函数为阶梯函数，则 X 一定是一个离散型随机变量，其分布函数 $F(x)$ 唯一确定.

例 2.11　设随机变量 X 的分布律为

X	2	3	4
P	0.2	0.3	0.5

求 X 的分布函数，并求 $P\{X\leqslant 2\}$，$P\{2.4<X\leqslant 3.8\}$，$P\{3\leqslant X\leqslant 4\}$.

解 当 $x<2$ 时，$F(x)=P\{X\leqslant x\}=0$；

当 $2\leqslant x<3$ 时，$F(x)=\sum\limits_{x_i\leqslant x}P\{X=x_i\}=P\{X=2\}=0.2$；

当 $3\leqslant x<4$ 时，$F(x)=\sum\limits_{x_i\leqslant x}P\{X=x_i\}=P\{X=2\}+P\{X=3\}=0.2+0.3=0.5$；

当 $x\geqslant 4$ 时，$F(x)=P\{X=2\}+P\{X=3\}+P\{X=4\}=1$.

综上所述，X 的分布函数

$$F(x)=\begin{cases}0, & x<2,\\ 0.2, & 2\leqslant x<3,\\ 0.5, & 3\leqslant x<4,\\ 1, & x\geqslant 4.\end{cases}$$

函数 $F(x)$ 是阶梯形右连续函数，其图形如图 2-3 所示，在 $x=2,3,4$ 处有跳跃，其跃度分别等于 $P\{X=2\}$，$P\{X=3\}$，$P\{x=4\}$.

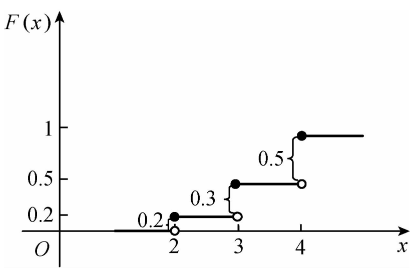


图 2-3

因此有

$$P\{X\leqslant 2\}=F(2)=0.2,$$
$$P\{2.4<X\leqslant 3.8\}=F(3.8)-F(2.4)=0.5-0.2=0.3,$$
$$P\{3\leqslant X\leqslant 4\}=F(4)-F(3)+P\{X=3\}=1-0.5+0.3=0.8.$$

例 2.12 设随机变量 X 的分布函数

$$F(x)=\begin{cases}0, & x<-1,\\ \dfrac{1}{4}, & -1\leqslant x<2,\\ \dfrac{3}{4}, & 2\leqslant x<3,\\ 1, & x\geqslant 3,\end{cases}$$

求 X 的分布律.

解　由于 $F(x)$ 是一个阶梯函数,故知 X 是一个离散型随机变量,$F(x)$ 的跳跃点分别为 $-1,2,3$,由 $P\{X=a\}=F(a)-F(a-0)$,得

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

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

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

故 X 的分布律为

X	-1	2	3
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

2.4　连续型随机变量及其分布

2.4.1　连续型随机变量及其概率密度

容易看到,在上一节例 2.8 中,对于任意 x,分布函数 $F(x)$ 均可表示为 $F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$,其中 $f(t)=\begin{cases}\dfrac{2t}{r^2}, & 0<t<r,\\ 0, & 其他,\end{cases}$ 这就是说 $F(x)$ 是非负函数 $f(t)$ 在区间 $(-\infty,x]$ 上的积分,称 X 为连续型随机变量,因而引入连续型随机变量的定义.

定义 2.8　如果对于随机变量 X 的分布函数 $F(x)$,存在非负函数 $f(x)$,使得对于任意实数 x,有

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

则称 X 为**连续型随机变量**,函数 $f(x)$ 称为 X 的**概率密度函数**,简称**概率密度**或**密度函数**,记为 $X\sim f(x)$ 或 $f_X(x)$.

注:对一个连续型随机变量 X,若已知其概率密度 $f(x)$,则根据定义可求得分布函数 $F(x)$. 分布函数 $F(x)$ 与概率密度 $f(x)$ 的关系的几何解释如图 2-4 所示.

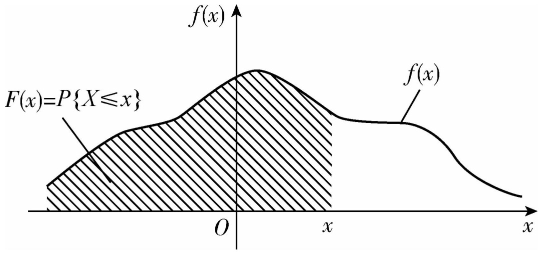


图 2-4

根据以上定义,易得到概率密度 $f(x)$ 的性质:

(1) **非负性** $f(x) \geqslant 0(-\infty < x < +\infty)$.

(2) **归一性** $\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1$.

注:以上两条性质是确定或判断某个函数是否为概率密度的充要条件,并有明显的几何意义,即介于概率密度曲线 $y = f(x)$ 及 x 轴之间的区域面积为 1(见图2-5).

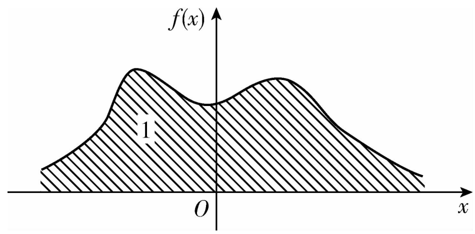


图 2-5

(3) $P\{a < X \leqslant b\} = F(b) - F(a) = \int_{a}^{b} f(x)\mathrm{d}x$.

注:性质(3) 将连续型随机变量 X 在区间$(a,b]$ 上取值的概率转化成了概率密度在区间$(a,b]$ 上的定积分(见图 2-6),从而可以利用高等数学中微积分知识求解概率计算问题.

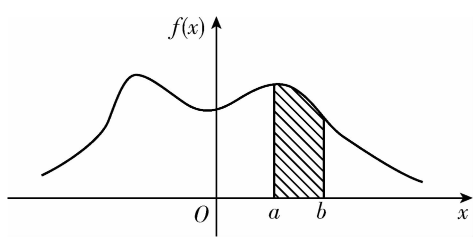


图 2-6

(4) 在 $f(x)$ 的连续点处，有 $F'(x)=f(x)$.

对于连续型随机变量 X，还要指出两点：① $F(x)$ 是连续函数；② $P\{X=a\}=0$（a 为任意实数）.

证明　(2) 因为 $P\{X=a\}=F(a)-F(a-0)$，而 $F(x)$ 是连续函数，有 $F(a-0)=F(a)$，所以 $P\{X=a\}=0$.

注：① $P\{X=a\}=0$ 说明连续型随机变量 X 取每一特定值的概率为 0，它并不表示不可能事件. 从而，概率为 0 的事件不一定是不可能事件，概率为 1 的事件也不一定为必然事件.

② 在计算连续型随机变量落在某一区间的概率时，可以不必区分该区间是开区间或闭区间或半开半闭区间，即

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

由性质(4) 知

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

上式可理解为：X 在点 x 处的概率密度 $f(x)$ 恰好是 X 落在区间 $(x,x+\Delta x]$ 上的概率与区间长度 Δx 之比的极限（这与物理学中线密度的定义类似，这也是称 $f(x)$ 为概率密度的原因）.

当 Δx 充分小时，有 $P\{x<X\leqslant x+\Delta x\}\approx f(x)\Delta x$，即 X 落在小区间 $(x,x+\Delta x]$ 上的概率近似等于 $f(x)\Delta x$.

例 2.13　设随机变量 X 的概率密度

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

(1) 确定常数 A；(2) 求 X 的分布函数；(3) 求 $P\{0<X<1\}$.

解　(1) 由概率密度性质(2) 得

$$1=\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\int_{-1}^{1}\frac{A}{\sqrt{1-x^2}}\mathrm{d}x=2A\arcsin x\Big|_0^1=2A\cdot\frac{\pi}{2}=\pi A,$$

因此 $A=\dfrac{1}{\pi}$，于是

$$f(x)=\begin{cases}\dfrac{1}{\pi\sqrt{1-x^2}}, & |x|<1,\\ 0, & |x|\geqslant 1.\end{cases}$$

(2) 由式(2.2) 得

当 $x \leqslant -1$ 时，$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t=\int_{-\infty}^{x} 0\mathrm{d}t=0$；

当 $-1<x<1$ 时，$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t=\int_{-\infty}^{-1} 0\mathrm{d}t+\int_{-1}^{x} \frac{1}{\pi\sqrt{1-t^2}}\mathrm{d}t=\frac{1}{\pi}\arcsin t\Big|_{-1}^{x}=\frac{1}{\pi}\arcsin x+\frac{1}{2}$；

当 $x \geqslant 1$ 时，$F(x)=\int_{-\infty}^{x} f(t)\mathrm{d}t=\int_{-\infty}^{-1} 0\mathrm{d}t+\int_{-1}^{1} \frac{1}{\pi\sqrt{1-t^2}}\mathrm{d}t+\int_{1}^{x} 0\mathrm{d}t=\frac{2}{\pi}\arcsin t\Big|_{0}^{1}=1$.

所以，X 的分布函数

$$F(x)=\begin{cases}0, & x\leqslant -1,\\ \frac{1}{\pi}\arcsin x+\frac{1}{2}, & -1<x<1,\\ 1, & x\geqslant 1.\end{cases}$$

(3) $P\{0<X<1\}=F(1)-F(0)=1-\left(\frac{1}{\pi}\arcsin 0+\frac{1}{2}\right)=\frac{1}{2}$.

例 2.14 设随机变量 X 的分布函数

$$F(x)=\begin{cases}A-B\mathrm{e}^{-2x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

(1) 确定常数 A,B；(2) 求概率密度 $f(x)$.

解 (1) 由分布函数的性质(2)知

$$1=F(+\infty)=\lim_{x\to+\infty}F(x)=\lim_{x\to+\infty}(A-B\mathrm{e}^{-2x})=A,$$

又由 $F(x)$ 的连续性得 $F(0+0)=F(0-0)=F(0)$，即 $A-B=0$，故 $B=1$.

所以，X 的分布函数

$$F(x)=\begin{cases}1-\mathrm{e}^{-2x}, & x\geqslant 0,\\ 0, & x<0.\end{cases}$$

(2) 由概率密度的性质(4)知，当 $x>0$ 时，$f(x)=F'(x)=2\mathrm{e}^{-2x}$；当 $x<0$ 时，$f(x)=F'(x)=0$；当 $x=0$ 时，$F'_-(0)=0$，$F'_+(0)=2$，即 $F(x)$ 在 $x=0$ 处的导数是不存在的，但可以补充定义，设该点的概率密度为 $f(0)=0$，因为改变概率密度个别点处的值不影响其在区间上的积分值. 因此

$$f(x)=\begin{cases}2\mathrm{e}^{-2x}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

注：若已知 $F(x)$ 为某连续型随机变量 X 的分布函数，则当 $F(x)$ 在点 x 处导数存在时，由 $F'(x)=f(x)$ 可直接求得概率密度；在点 x 处导数不存在时，可定义

$f(x)$ 为任一非负值，一般地，可取 $f(x)=0$. 以后本书中遇到这种情况不再赘述.

2.4.2　常用连续型随机变量的分布

1）均匀分布

定义 2.9　设连续型随机变量 X 的概率密度

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其他},\end{cases} \tag{2.3}$$

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

易见 $f(x)\geqslant 0$，$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$.

由式(2.2) 得 X 的分布函数为

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

$f(x)$ 和 $F(x)$ 的图形分别如图 2-7 和图 2-8 所示.

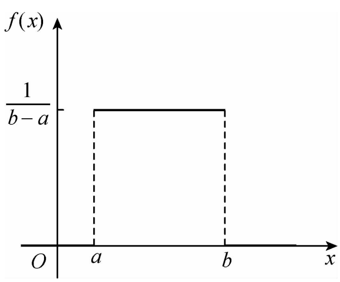


图 2-7

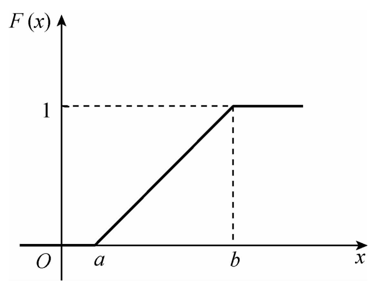


图 2-8

注：式(2.3) 中的“$a<x<b$”也可定义为“$a<x\leqslant b$”，“$a\leqslant x<b$”或“$a\leqslant x\leqslant b$”.

若 $X\sim U(a,b)$，则对任一长度 l 的子区间 $(c,c+l)$，$a\leqslant c<c+l\leqslant b$，有

$$P\{c<X<c+l\}=\int_{c}^{c+l}\frac{1}{b-a}\mathrm{d}x=\frac{l}{b-a},$$

说明服从均匀分布的随机变量 X 落在 (a,b) 内任何一个子区间的概率与该区间的长度成正比，而与该区间的位置无关，即 X 的取值落在任意相等的子区间的可能性是相同的，因此均匀分布常用作随机投点的数学模型.

在实际生活中有许多随机变量是服从均匀分布的. 例如，每隔一定时间有一辆公共汽车通过汽车站，乘客候车时间是服从均匀分布的. 再如，司机刹车时，轮胎接

触地面的点与地面摩擦会有一定的磨损，刹车时与地面接触点的位置服从均匀分布，即在任一等长的小区间上发生磨损的可能性是相同的.

例 2.15 某公共汽车站从上午 7 时起，每 15 分钟来一班车，即 7:00，7:15，7:30，7:45 等时刻有汽车到达此站，如果乘客到达此站的时间 X 是 7:00 到 7:30 之间的均匀随机变量，试求他候车时间少于 5 分钟的概率.

解 以 7:00 为起点 0，以分为单位. 依题意，$X \sim U(0,30)$，于是

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

为使候车时间少于 5 分钟，乘客必须在 7:10 到 7:15 之间，或在 7:25 到 7:30 之间到达车站，故所求概率

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

即乘客候车时间少于 5 分钟的概率为$\dfrac{1}{3}$.

例 2.16 设随机变量 X 在(2,5)上服从均匀分布，现对 X 进行三次独立观测，试求至少有两次观测值大于 3 的概率.

解 据题意，X 在(2,5)上服从均匀分布，所以 X 的概率密度

$$f(x)=\begin{cases}\dfrac{1}{3}, & 2<x<5,\\ 0, & \text{其他},\end{cases}$$

于是，事件"对 X 的观测值大于 3"的概率

$$P\{X>3\}=\int_{3}^{5}\frac{1}{3}\mathrm{d}x=\frac{2}{3}.$$

设 Y 表示三次独立观测值大于 3 的次数，则 $Y\sim b(3,\dfrac{2}{3})$，因此

$$P\{Y\geqslant 2\}=C_3^2\left(\frac{2}{3}\right)^2\times\frac{1}{3}+C_3^3\left(\frac{2}{3}\right)^3=\frac{20}{27}.$$

2）指数分布

定义 2.10 设连续型随机变量 X 的概率密度

$$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}\tag{2.4}$$

其中 $\lambda(\lambda>0)$ 为常数，则称 X 服从参数为 λ 的**指数分布**，记为 $X\sim Exp(\lambda)$.

易见 $f(x)\geqslant 0$，$\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$.

由式(2.2)得 X 的分布函数

$$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, & x>0,\\ 0, & x\leqslant 0.\end{cases} \tag{2.5}$$

$f(x)$ 的图形如图 2-9 所示.

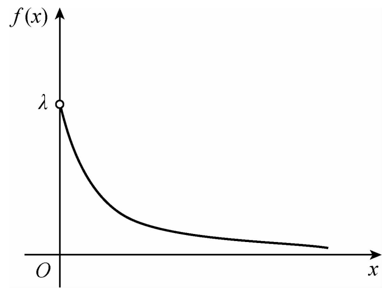


图 2-9

注:在式(2.4)中,若令 $\lambda=\dfrac{1}{\theta}$,则指数分布的概率密度 $f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{1}{\theta}x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$ 其中 $\theta>0$,则称 X 服从参数为 θ 的指数分布,记为 $X\sim Exp(\theta)$. 有时为讨论问题的方便,也可用 θ 为参数的指数分布.

若 $X\sim Exp(\lambda)$,则对任意 $s>0,t>0$,有

$$P\{X>s+t\mid X>s\}=P\{X>t\}, \tag{2.6}$$

事实上,

$$P\{X>s+t\mid X>s\}=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>s\}}=\frac{P\{X>s+t\}}{P\{X>s\}}$$

$$=\frac{1-F(s+t)}{1-F(s)}=\frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}}=\mathrm{e}^{-\lambda t}=P\{X>t\}.$$

指数分布的这一特性称为**无记忆性**. 若 X 表示某一元件的寿命,则式(2.6)表明,已知元件已使用了 s 小时,则再使用 t 小时的概率与 s 无关. 这就是说,元件对它已使用过 s 小时没有记忆,这一性质使指数分布具有广泛的应用性.

在实际中,指数分布常用来作为各种“寿命”分布的近似分布. 例如,电子元件熔丝等的寿命就是服从指数分布. 再如,随机服务系统中的服务时间、每次打电话的通话时间、事故发生间隔的时间、汽车行驶的里程数等都服从指数分布. 指数分布在可靠性理论与排队论中也有着广泛的应用.

例 2.17　某电子元件无故障地工作的总时间 X(单位:h)服从参数为 $\lambda=\dfrac{1}{100}$ 的指数分布,试求这个电子元件无故障地工作 $50\sim150$ h 的概率,以及工作时间少

于 100 h 的概率.

解 由题意,随机变量 X 的概率密度

$$f(x)=\begin{cases}\dfrac{1}{100}e^{-\frac{x}{100}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$$

则所求概率分别

$$P\{50\leqslant X\leqslant 150\}=\frac{1}{100}\int_{50}^{150}e^{-\frac{x}{100}}dx=0.382,$$

$$P\{X<100\}=\frac{1}{100}\int_{0}^{100}e^{-\frac{x}{100}}dx=0.632.$$

注:在讨论服从泊松分布的一类问题中,等待某一事件发生所需要的时间是服从指数分布的随机变量,且可证明泊松分布的参数 λ 即为指数分布中的参数 λ.

例 2.18 假定自动取款机对每位顾客的服务时间(单位:min)服从 $\lambda=\dfrac{1}{3}$ 的指数分布,如果有一顾客恰好在你前头走到空闲的取款机,求:

(1) 你至少等候 3 min 的概率.

(2) 你等候时间在 $3\sim 6$ min 的概率.

(3) 如果你到达取款机时,正有一名顾客使用着取款机,则(1) 和(2) 中的概率又分别是多少?

解 以 X 表示你前面这位顾客所用服务时间,$F(x)$ 为 X 的分布函数,由式(2.5)所求概率为

(1) $P\{X\geqslant 3\}=1-F(3)=1-(1-e^{-\frac{3}{3}})=e^{-1}=0.368.$

(2) $P\{3\leqslant X\leqslant 6\}=F(6)-F(3)=(1-e^{-\frac{6}{3}})-(1-e^{-\frac{3}{3}})=e^{-1}-e^{-2}=0.233.$

(3) 如果你到达时,一名顾客已使用取款机 s 分钟,同时没有其他人在排队等候,那么由指数分布的无记忆性,你等候时间的概率与 s 分钟无关,从而问题(1) 和(2) 的答案均不变.

3) 正态分布

定义 2.11 设连续型随机变量 X 的概率密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in\mathbf{R},$$

其中 $\mu,\sigma(\sigma>0)$ 为常数,则称 X 服从参数为 μ,σ^2 的**正态分布**或**高斯分布**,记为 $X\sim N(\mu,\sigma^2)$.

易见 $f(x)\geqslant 0$,下面来证明 $\int_{-\infty}^{+\infty}f(x)dx=1$.

若令$\dfrac{x-\mu}{\sigma}=t$,得到

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

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

$$I^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{x^2+y^2}{2}}\mathrm{d}x\mathrm{d}y,$$

利用极坐标计算二重积分

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

而 $I>0$,故有 $I=\sqrt{2\pi}$,即$\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\sqrt{2\pi}$,于是

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

习惯上常把服从正态分布的随机变量称为**正态变量**.

$f(x)$ 的图形如图 2-10 所示.

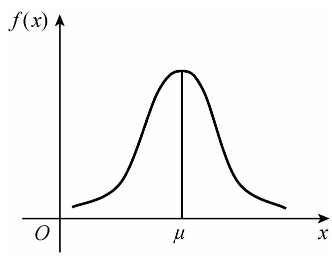


图 2-10

$f(x)$ 的图形有以下几个特征:

(1) 曲线关于 $x=\mu$ 对称.

(2) 曲线在 $x=\mu$ 时达到最大值 $f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$.

(3) 曲线在 $x=\mu\pm\sigma$ 处有拐点,且以 x 轴为渐近线.

(4) μ 能够确定曲线的位置,若固定 σ 而改变 μ 的值,则图形沿 x 轴平移而不改变其形状,称 μ 为**位置参数**(见图 2-11).

(5) σ 能够确定曲线中峰的陡峭程度,若固定 μ 而改变 σ 的值,则 σ 越大,曲线越平缓,σ 越小,曲线越陡峭,称 σ 为**形状参数**(见图 2-12),σ 的大小反映了 X 取值的集中或分散程度.

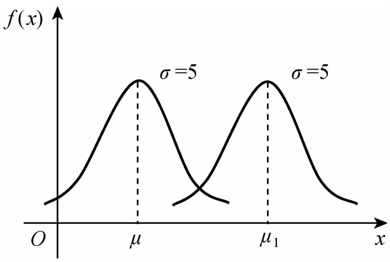


图 2-11

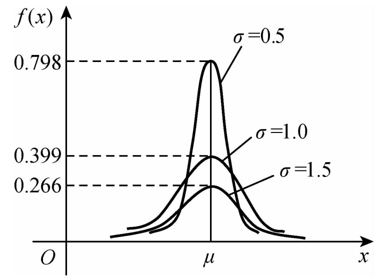


图 2-12

正态分布的分布函数

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

特别地，当$\mu=0,\sigma=1$时，称X服从**标准正态分布**，记为$X\sim N(0,1)$. 此时，概率密度和分布函数分别用$\varphi(x)$和$\Phi(x)$表示，即有

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

$\varphi(x)$和$\Phi(x)$的图形如图 2-13 和图 2-14 所示.

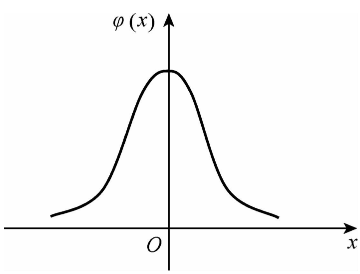


图 2-13

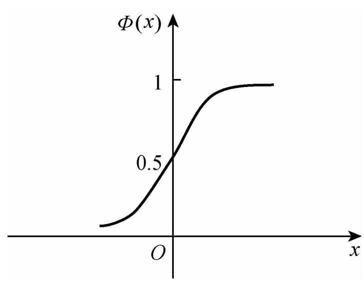


图 2-14

若$X\sim N(0,1)$，则由对称性知：

(1) $\Phi(0)=P\{X\leqslant 0\}=P\{X\geqslant 0\}=\dfrac{1}{2}$.

(2) $\Phi(-x)=1-\Phi(x)$.

标准正态分布的重要性在于任何一个正态分布都可以通过线性变换转化为标准正态分布.

定理 2.2 设$X\sim N(\mu,\sigma^2)$，则$Y=\dfrac{X-\mu}{\sigma}\sim N(0,1)$.

证明 $Y=\dfrac{X-\mu}{\sigma}$的分布函数为

$$P\{Y\leqslant x\}=P\{\frac{X-\mu}{\sigma}\leqslant x\}=P\{X\leqslant \mu+\sigma x\}$$

$$=\int_{-\infty}^{\mu+\sigma x}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t\xlongequal{u=\frac{t-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u=\Phi(x),$$

所以

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

因为正态分布函数是一个超越函数，直接计算有困难，人们已编制了标准正态分布表(见附表 3)，可供查阅.

标准正态分布表的使用方法如下：

(1) 表中给出了 $x>0$ 时的数值，当 $x<0$ 时，利用 $\Phi(-x)=1-\Phi(x)$ 求得.

(2) 若 $X\sim N(0,1)$，则

$$P\{a<X\leqslant b\}=P\{a\leqslant X<b\}=P\{a<X<b\}=P\{a\leqslant X\leqslant b\}$$
$$=\Phi(b)-\Phi(a).$$

(3) 若 $X\sim N(\mu,\sigma^2)$，则 X 的分布函数为

$$F(x)=P\{X\leqslant x\}=P\left\{\frac{X-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\right\}=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

对任意区间 $(a,b]$ 有

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

例 2.19　设 $X\sim N(90,0.5^2)$，求 $P\{X<89\}$，$P\{88<X\leqslant 91.2\}$.

解　这里 $\mu=90$，$\sigma=0.5$. 于是

$$P\{X<89\}=P\left\{\frac{X-90}{0.5}<\frac{89-90}{0.5}\right\}=\Phi\left(\frac{89-90}{0.5}\right)$$
$$=\Phi(-2)=1-\Phi(2)=1-0.977\,2=0.022\,8,$$

$$P\{88<X\leqslant 91.2\}=P\left\{\frac{88-90}{0.5}<\frac{X-90}{0.5}\leqslant\frac{91.2-90}{0.5}\right\}$$
$$=\Phi\left(\frac{91.2-90}{0.5}\right)-\Phi\left(\frac{88-90}{0.5}\right)$$
$$=\Phi(2.4)-\Phi(-4)$$
$$=\Phi(2.4)-[1-\Phi(4)]$$
$$=\Phi(2.4)=0.991\,8.$$

例 2.20　设 $X\sim N(\mu,\sigma^2)$，求 $P\{|X-\mu|<\sigma\}$，$P\{|X-\mu|<2\sigma\}$，$P\{|X-\mu|<3\sigma\}$.

解　$P\{|X-\mu|<\sigma\}=P\left\{-1<\frac{X-\mu}{\sigma}<1\right\}=\Phi(1)-\Phi(-1)$

$$=2\Phi(1)-1=2\times 0.841\,3-1=0.682\,6,$$

$$P\{|X-\mu|<2\sigma\}=2\Phi(2)-1=0.9544,$$
$$P\{|X-\mu|<3\sigma\}=2\Phi(3)-1=0.9974.$$

由例2.20可以看出，虽然正态变量X的取值范围是$(-\infty,+\infty)$，但它的取值几乎都落在区间$(\mu-3\sigma,\mu+3\sigma)$内，被称为正态分布的“$3\sigma$”法则，这也是一些随机变量即使不在$(-\infty,+\infty)$内取值，也可看成正态分布的原因.

例 2.21 在车床上加工金属圆杆，已知圆杆直径(单位:cm)$X\sim N(12.4,\sigma^2)$，规定直径在$12.0\sim12.8$ cm之间为合格品，要求产品合格的概率至少为0.95，试确定σ最多为多少?

解 依题意，需求σ，使得

$$0.95\leqslant P\{12.0\leqslant X\leqslant 12.8\}=\Phi\left(\frac{12.8-12.4}{\sigma}\right)-\Phi\left(\frac{12.0-12.4}{\sigma}\right)=2\Phi\left(\frac{0.4}{\sigma}\right)-1,$$

即$\Phi\left(\frac{0.4}{\sigma}\right)\geqslant 0.975=\Phi(1.96)$(反查标准正态分布表).

又由$\Phi(x)$的单调性知，

$$\frac{0.4}{\sigma}\geqslant 1.96,\sigma\leqslant 0.204,$$

也就是说，σ最多为0.204 cm.

为了便于今后在数理统计中的应用，对标准正态随机变量，引入上α分位点的定义.

定义 2.12 设$X\sim N(0,1)$，若z_α满足$P\{X>z_\alpha\}=\alpha(0<\alpha<1)$，则称点$z_\alpha$为标准正态分布的**上 α 分位点**(见图2-15).

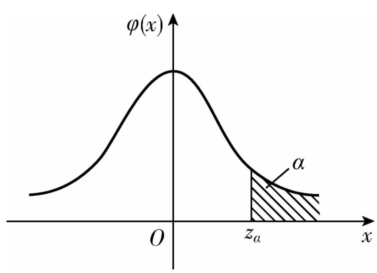


图 2-15

由$\varphi(x)$图形的对称性知，$z_{1-\alpha}=-z_\alpha$.

由关系式$\Phi(z_\alpha)=1-P\{X>z_\alpha\}=1-\alpha$，对于给定的$\alpha$的值，反查标准正态分布表可得$z_\alpha$的值.例如，$\alpha=0.05$，由$\Phi(z_{0.05})=1-0.05=0.95$知，$z_\alpha=1.645$.

数理统计中还用到双侧α分位点.

定义 2.13　设 $X \sim N(0,1)$，若 $z_{\alpha/2}$ 满足 $P\{|x| \geqslant z_{\alpha/2}\} = \alpha(0 < \alpha < 1)$，则称点 $z_{\alpha/2}$ 为标准正态分布的**双侧 α 分位点**（见图 2-16）.

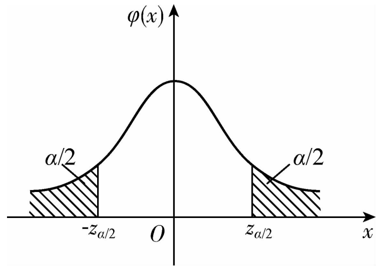


图 2-16

正态分布是概率论中最重要的一种分布，也是自然界中最常见的一种分布. 例如，人的身高、体重、学生的考试成绩、农作物的产量、加工产品的尺寸、测量误差等，它们都服从或近似服从正态分布. 在第 5 章，还会看到，许多非正态变量也和正态变量有着密切的联系，因此，正态变量在概率论与数理统计的理论研究和实际应用中起到非常重要的作用.

2.5　随机变量的函数的分布

在实际问题中，人们所关心的随机变量以及它们的分布往往难以直接得到，而与它们有关的另一些随机变量及其分布却是容易知道的. 例如，某商场销售一种商品，销售量 X 是一个随机变量且服从指数分布，而销售收入 $Y = mX$（m 是该商品的价格）是销售量的函数，因而销售收入 Y 也是一个随机变量，因此，需要讨论如何由已知的销售量 X 及其分布求销售收入 Y 及其分布.

下面将讨论如何由已知的随机变量 X 的分布去求它的函数 $Y = g(X)$（$g(x)$ 是已知的连续函数）的分布.

2.5.1　离散型随机变量的函数的分布

设离散型随机变量 X 的分布律为

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

给定函数关系 $y=g(x)$，则 $Y=g(X)$ 仍是离散型随机变量.

记 $y_i=g(x_i)(i=1,2,3,\cdots)$，若随机变量 X 与 Y 的取值一一对应，则 $Y=g(X)$ 的分布律为

Y	y_1	y_2	$\cdots$	y_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

这是因为事件$\{Y=y_i\}$发生当且仅当事件$\{X=x_i\}$发生，故

$$P\{Y=y_i\}=P\{X=x_i\}(i=1,2,3,\cdots).$$

通常把随机变量的可能取值按从小到大的次序排列起来. 若 $y_i=g(x_i)(i=1,2,3,\cdots)$ 中有相等的，则应使 $g(x_i)$ 相等的那些 x_i 所对应的概率相加作为 Y 取 $g(x_i)$ 的概率. 由此得到 Y 的分布律的方法通常称为**寻根法**.

例 2.22 设离散型随机变量 X 的分布律为

X	-1	0	1	2
P	0.3	0.4	0.1	0.2

求 $Y=2X+1, Z=(X-1)^2$ 的分布律.

解 据题意可得

X	-1	0	1	2
$Y=2X+1$	-1	1	3	5
$Z=(X-1)^2$	4	1	0	1
P	0.3	0.4	0.1	0.2

由此可得 Y,Z 的分布律分别为

Y	-1	1	3	5
P	0.3	0.4	0.1	0.2

Z	0	1	4
P	0.1	0.6	0.3

2.5.2 连续型随机变量的函数的分布

求连续型随机变量的函数的概率密度主要有两种方法，即分布函数法和公式法.

1）分布函数法

设连续型随机变量 X 的概率密度为 $f_X(x)$，求 X 的函数 $Y=g(X)$ 的概率密度 $f_Y(y)$ 的一般步骤为：

(1) 求 $Y=g(X)$ 的分布函数

$$F_Y(y)=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}=P\{X\in I_x\}=\int_{I_x} f_X(x)\mathrm{d}x,$$

其中 $I_x=\{x\mid g(x)\leqslant y\}$.

(2) 将上式两端对 y 求导，即得 Y 的概率密度

$$f_Y(y)=F'_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}\int_{I_x} f_X(x)\mathrm{d}x.$$

例 2.23　设随机变量 X 具有概率密度 $f_X(x)$，$-\infty<x<+\infty$，求 $Y=|x|$ 的概率密度.

解　分别记 X,Y 的分布函数为 $F_X(x)$，$F_Y(y)$，先求 Y 的分布函数 $F_Y(y)$.

$$F_Y(y)=P\{Y\leqslant y\}=P\{|x|\leqslant y\}.$$

当 $y<0$ 时，$\{Y\leqslant y\}$ 是不可能事件，知 $F_Y(y)=0$；

当 $y\geqslant 0$ 时，$F_Y(y)=P\{-y\leqslant X\leqslant y\}=F_X(y)-F_X(-y)$，

即

$$F_Y(y)=\begin{cases}F_X(y)-F_X(-y), & y\geqslant 0,\\ 0, & y<0.\end{cases}$$

将 $F_Y(y)$ 关于 y 求导数得 Y 的概率密度

$$f_Y(y)=\begin{cases}F'_X(y)-[F_X(-y)]', & y>0,\\ 0, & y\leqslant 0\end{cases}$$

$$=\begin{cases}f_X(y)+f_X(-y), & y>0,\\ 0, & y\leqslant 0.\end{cases}$$

特别地，若 $X\sim N(0,1)$，其概率密度 $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}(-\infty<x<+\infty)$，则

$$F_Y(y)=\begin{cases}\Phi(y)-\Phi(-y), & y\geqslant 0,\\ 0, & y<0\end{cases}=\begin{cases}2\Phi(y)-1, & y\geqslant 0,\\ 0, & y<0.\end{cases}$$

当 $y>0$ 时，$f_Y(y)=F'_Y(y)=2\Phi'(y)=2\varphi(y)=\sqrt{\dfrac{2}{\pi}}\mathrm{e}^{-\frac{y^2}{2}}$；

当 $y\leqslant 0$ 时，$f_Y(y)=0$，

即

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

例 2.24 设随机变量具有概率密度

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

求随机变量 $Y=X^2$ 的概率密度.

解 X,Y 的分布函数为 $F_X(x)$,$F_Y(y)$,先求 Y 的分布函数 $F_Y(y)$.

$$F_Y(y)=P\{Y\leqslant y\}=P\{X^2\leqslant y\}=\begin{cases}0, & y<0,\\ P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}, & y\geqslant 0.\end{cases}$$

当 $0\leqslant y<16$ 时,

$$P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}=\int_{-\sqrt{y}}^{\sqrt{y}}f_X(x)\mathrm{d}x=\int_{-\sqrt{y}}^{0}0\mathrm{d}x+\int_0^{\sqrt{y}}\frac{x}{8}\mathrm{d}x=\int_0^{\sqrt{y}}\frac{x}{8}\mathrm{d}x;$$

当 $y\geqslant 16$ 时,

$$\begin{aligned}P\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}&=\int_{-\sqrt{y}}^{\sqrt{y}}f_X(x)\mathrm{d}x=\int_{-\sqrt{y}}^{0}0\mathrm{d}x+\int_0^4\frac{x}{8}\mathrm{d}x+\int_4^{\sqrt{y}}0\mathrm{d}x\\&=\int_0^4\frac{x}{8}\mathrm{d}x.\end{aligned}$$

综上所述,

$$F_Y(y)=\begin{cases}0, & y<0,\\ \displaystyle\int_0^{\sqrt{y}}\frac{x}{8}\mathrm{d}x, & 0\leqslant y<16,\\ \displaystyle\int_0^4\frac{x}{8}\mathrm{d}x, & y\geqslant 16,\end{cases}$$

将 $F_Y(y)$ 关于 y 求导数,得 Y 的概率密度

$$f_Y(y)=\begin{cases}\dfrac{1}{16}, & 0<y<16,\\ 0, & \text{其他},\end{cases}$$

可见 Y 在区间 $(0,16)$ 上服从均匀分布,即 $Y\sim U(0,16)$.

用分布函数法可证明下述正态分布的重要性质.

定理 2.3 设 $X\sim N(\mu,\sigma^2)$,则 $Y=aX+b\sim N(a\mu+b,(a\sigma)^2)(a\neq 0)$.

证明 Y 的分布函数 $F_Y(y)=P\{Y\leqslant y\}=P\{aX+b\leqslant y\}$.

若 $a>0$,则有 $F_Y(y)=P\left\{X\leqslant\dfrac{y-b}{a}\right\}=F_X\left(\dfrac{y-b}{a}\right)$;

若 $a<0$,则有 $F_Y(y)=P\left\{X\geqslant\dfrac{y-b}{a}\right\}=1-F_X\left(\dfrac{y-b}{a}\right)$.

将上面两式分别对 y 求导得

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

故 $Y=aX+b\sim N(a\mu+b,(a\sigma)^2)$.

特别地,若取 $a=\frac{1}{\sigma}$,$b=-\frac{\mu}{\sigma}$,即得 $Y\sim N(0,1)$.

上面介绍的分布函数法是求随机变量函数 $Y=g(X)$ 的分布的主要方法,它适用范围广泛,对 $g(x)$ 为单调函数与非单调函数均适用,当 $g(x)$ 为单调函数时还可用公式法求 $Y=g(X)$ 的分布.

2)公式法

定理 2.4　设随机变量 X 具有概率密度 $f_X(x)$,$-\infty<x<+\infty$,又设函数 $g(x)$ 处处可导且恒有 $g'(x)>0$(或恒有 $g'(x)<0$),则 $Y=g(X)$ 是连续型随机变量,其概率密度

$$f_Y(y)=\begin{cases}f_X[h(y)]|h'(y)|, & \alpha<y<\beta,\\ 0, & \text{其他},\end{cases}\tag{2.7}$$

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

证明　先证 $g'(x)>0$ 的情形.

由 $g'(x)>0$,得 $y=g(x)$ 在$(-\infty,+\infty)$ 严格单调增加,它的反函数 $x=h(y)$ 存在,且在(α,β) 严格单调增加、可导,先求 Y 的分布函数 $F_Y(y)$.

因 $Y=g(X)$ 在(α,β) 取值,故

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

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

当 $\alpha<y<\beta$ 时,$F_Y(y)=P\{Y\leqslant y\}=P\{g(X)\leqslant y\}=P\{X\leqslant h(y)\}=F_X[h(y)]$.

将 $F_Y(y)$ 关于 y 求导数,得 Y 的概率密度

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

类似地,可证明 $g'(x)<0$ 的情形,有

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

综上所述,Y 的概率密度

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

若 $f(x)$ 在有限区间$[a,b]$以外等于零,当 $x\in[a,b]$ 时,$g(x)\in[\alpha,\beta]$($\alpha=\min\{g(a),g(b)\}$, $\beta=\max\{g(a),g(b)\}$)且在$[a,b]$上恒有 $g'(x)>0$(或恒有 $g'(x)<0$),则仍可按式(2.7)求得 $Y=g(X)$ 的概率密度.

下面用公式法证明定理 2.3.

证明 因 $y=g(x)=ax+b(a\neq 0)$ 严格单调,其反函数为 $x=h(y)=\dfrac{y-b}{a}$, $-\infty<y<+\infty$,且有 $h'(y)=\dfrac{1}{a}$. 又 $f_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}(-\infty<x<+\infty)$,由式(2.7)得 $Y=aX+b$ 的概率密度

$$
\begin{aligned}
f_Y(y)&=\frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)=\frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\\
&=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}},-\infty<y<+\infty,
\end{aligned}
$$

故 $Y=aX+b\sim N(a\mu+b,(a\sigma)^2)$.

例 2.25 设随机变量 X 的概率密度

$$
f_X(x)=\begin{cases}2x, & 0<x<1,\\ 0, & \text{其他},\end{cases}
$$

求 $Y=e^{-X}$ 的概率密度.

解 因 $y=e^{-x}$ 严格单调减少,其反函数为 $x=-\ln y$ 在$\left(\dfrac{1}{e},1\right)$上可导,又 $x'=(-\ln y)'=-\dfrac{1}{y}$,由式(2.7)得 $Y=e^{-X}$ 的概率密度

$$
f_Y(y)=\begin{cases}f_X(-\ln y)\left|-\dfrac{1}{y}\right|, & \dfrac{1}{e}<y<1,\\ 0, & \text{其他}\end{cases}=\begin{cases}-\dfrac{2}{y}\ln y, & \dfrac{1}{e}<y<1,\\ 0, & \text{其他}.\end{cases}
$$

可见当 $g(x)$ 为单调函数时,用公式法求 $Y=g(X)$ 的概率密度较为方便.

注:利用式(2.7)直接求出 $Y=g(X)$ 的概率密度时要注意两点:

① $y=g(x)$ 是严格单调函数;

② 在公式中,$h'(y)$ 要取绝对值,否则会出现 $f_Y(y)$ 的取值小于 0 的情形.

本章小结

本章首先引入了随机变量的概念,使随机试验中的各种事件可以用随机变量来描述,从而将对随机现象的研究转化成对随机变量的研究. 本书只讨论两类重要的随机变量:离散型和连续型随机变量. 主要研究这两类随机变量及其分布.

对于离散型随机变量,要找到它的所有可能取值,并一一计算出取到这些值的概率,这就是离散型随机变量取值的统计规律性.因而,对于离散型随机变量,用分布律

$$P\{X=x_i\}=p_i, i=1,2,3,\cdots$$

或

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

来描述它的取值的统计规律性较为直观和简洁.

由分布律可得分布函数

$$F(x)=P\{X\leqslant x\}=\sum_{x_i\leqslant x}P\{X=x_i\}, i=1,2,3,\cdots$$

由分布函数 $F(x)$ 可得分布律

$$P\{X=x_i\}=F(x_i)-F(x_i-0), i=1,2,3,\cdots$$

通常用概率密度 $f(x)$ 来刻画连续型随机变量,概率密度与离散情况下的分布律的作用完全相同.

由概率密度 $f(x)$ 可得分布函数

$$F(x)=P\{X\leqslant x\}=\int_{-\infty}^{x}f(t)\mathrm{d}t.$$

由上式可见,改变 $f(x)$ 在个别点上的函数值并不改变 $F(x)$ 的值,因此改变 $f(x)$ 在个别点上的值是无关紧要的.

由分布函数 $F(x)$ 可得概率密度 $f(x)$,在 $f(x)$ 的连续点处有 $f(x)=F'(x)$.

分布函数可以刻画一般的随机变量,它涵盖了离散型和连续型的随机变量,也可以刻画既非离散又非连续的随机变量,因此分布函数的概念更加一般.离散型随机变量 X 的分布函数是一个右连续的阶梯函数,在 $X=x_i(i=1,2,3,\cdots)$ 处有跳跃,跳跃值为 $p_i=P\{X=x_i\}(i=1,2,3,\cdots)$,而连续型随机变量 X 的分布函数是连续的,连续型随机变量取任一指定实数 a 的概率为 0,即 $P\{X=a\}=0$,离散型随机变量不具有这一特性.

随机变量 X 的函数 $Y=g(x)$ 也是一个随机变量.对于离散型随机变量的函数,已知 X 的分布律,求 Y 的分布律,用寻根法;对于连续型随机变量的函数,已知 X 的概率密度 $f_X(x)$,求 Y 的概率密度 $f_Y(y)$,用分布函数法,当该函数单调可导时可直接用公式.

本章应理解随机变量、分布律、概率密度以及分布函数的概念,掌握分布律、概率密度、分布函数的性质,熟知(0-1) 分布、二项分布、泊松分布、均匀分布、指数分布、正态分布这六大重要分布,会利用分布函数、分布律或概率密度计算随机事件的

概率，会由分布律或概率密度求分布函数，会由分布函数求分布律或概率密度，会求随机变量函数的分布.

本章知识要点

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概率密度　(0-1) 分布　二项分布　泊松分布　指数分布　均匀分布　正态分布　随机变量函数的分布

本章常用结论

1）关于分布的充要条件

(1) $P\{X=x_i\}=p_i(i=1,2,3,\cdots)$ 是分布律的充要条件是 ① $p_i\geqslant 0(i=1,2,3,\cdots)$；② $\sum\limits_{i=1}^{+\infty}p_i=1$.

(2) $f(x)$ 是概率密度的充要条件是 ① $f(x)\geqslant 0$；② $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$.

(3) $F(x)$ 是分布函数的充要条件是 ① $0\leqslant F(x)\leqslant 1$；② $F(-\infty)=0,F(+\infty)=1$；③ 单调不减；④ 右连续.

2）常见的重要分布

(1) (0-1) 分布　若 $X\sim b(1,p)$，则 X 的分布律为

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

(2) 二项分布　若 $X\sim b(n,p)$，则 X 的分布律为

$$P\{X=k\}=\mathrm{C}_n^k p^k(1-p)^{n-k},k=0,1,2,\cdots,n(0<p<1).$$

(3) 泊松分布　若 $X\sim \pi(\lambda)$，则 X 的分布律为

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

(4) 均匀分布　若 $X\sim U(a,b)$，则 X 的概率密度

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

分布函数

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

(5) 指数分布　　若 $X\sim Exp(\lambda)$,则 X 的概率密度

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

分布函数

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

(6) 正态分布　　若 $X\sim N(\mu,\sigma^2)$,则 X 的概率密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},x\in\mathbf{R},$$

分布函数

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

(7) 标准正态分布　　若 $X\sim N(0,1)$,则 X 的概率密度

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

分布函数

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

(8) 若 $X\sim N(\mu,\sigma^2)$,则$\dfrac{X-\mu}{\sigma}\sim N(0,1)$.

(9) 若 $X\sim N(\mu,\sigma^2)$,则 $Y=aX+b\sim N(a\mu+b,(a\sigma)^2)(a\neq 0)$.

3）计算公式

(1) 若 $X\sim F(x)$,则 $P\{a<X\leqslant b\}=F(b)-F(a)$.

(2) 若 $P\{X=x_i\}=p_i,i=1,2,3,\cdots$,则 $P\{a<X\leqslant b\}=\sum\limits_{x_i\in(a,b]}P\{X=x_i\}$.

(3) 若 $X\sim f(x)$, 则 $P\{a<X\leqslant b\}=\int_a^b f(x)dx$.

(4) 若 $X\sim N(0,1)$, 则 $\Phi(x)+\Phi(-x)=1$.

(5) 若 $X\sim N(\mu,\sigma^2)$, 则 $F(x)=\Phi\left(\dfrac{x-\mu}{\sigma}\right)$.

(6) 若 $g(x)$ 单调可导,且 $X\sim f_X(x)$,则 $Y=g(X)$ 的概率密度

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

其中 $h(y)$ 是 $g(x)$ 的反函数，(α,β) 是 $g(x)$ 的值域.

习题 2

(1) 填空题.

① 设随机变量 X 服从参数为 λ 的泊松分布，且 $P\{X=0\}=\dfrac{1}{2}$，则 $\lambda=$ ________，$P\{X>1\}=$ ________.

② 设随机变量 X 服从参数为 $2,p$ 的二项分布，随机变量 Y 服从参数为 $3,p$ 的二项分布，若 $P\{X\geqslant 1\}=\dfrac{5}{9}$，则 $P\{Y\geqslant 1\}=$ ________.

③ 设随机变量 X 服从参数 λ 为 2 的指数分布，a 为任意实数，则 $P\{X>a^2+2\,|\,X>a^2\}=$ ________.

④ 设随机变量 X 服从正态分布 $N(\mu,\sigma^2)(\sigma>2)$，且二次方程 $y^2+4y+X=0$ 无实根的概率为$\dfrac{1}{2}$，则 $\mu=$ ________.

⑤ 设随机变量 X 的概率密度

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

若 k 使得 $P\{X\geqslant k\}=\dfrac{2}{3}$，则 k 的取值范围是________.

(2) 选择题.

① 下列四个函数中，不能作为随机变量的分布函数的是(　　).

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

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

C. $F(x)=\begin{cases}\dfrac{\ln(1+x)}{1+x}, & x\geqslant 0,\\ 0, & x<0\end{cases}$

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

② 设随机变量 X 的概率密度为 $f(x)$，且 $f(-x)=f(x)$，$F(x)$ 是 X 的分布函数，则对任意实数 a，有（　　）.

A. $F(-a)=1-\int_0^a f(x)\mathrm{d}x$　　　　B. $F(-a)=\frac{1}{2}-\int_0^a f(x)\mathrm{d}x$

C. $F(-a)=F(a)$　　　　D. $F(-a)=2F(a)-1$

③ 设随机变量 $X\sim N(\mu,\sigma^2)$，则随着 σ 的增大，概率 $P\{|X-\mu|<\sigma\}$（　　）.

A. 单调增大　　　　B. 单调减小

C. 保持不变　　　　D. 增减不定

④ 设随机变量 X 服从指数分布，则随机变量 $Y=\min(X,2)$ 的分布函数（　　）.

A. 是连续函数　　　　B. 至少有两个间断点

C. 是阶梯函数　　　　D. 恰好有一个间断点

⑤ 设随机变量 $X\sim N(0,1)$，对给定的 $\alpha\in(0,1)$，数 μ_α 满足 $P\{X>\mu_\alpha\}=\alpha$，若 $P\{|X|<x\}=\alpha$，则 x 等于（　　）.

A. $\mu_{\alpha/2}$　　　　B. $\mu_{1-\alpha/2}$

C. $\mu_{1-\alpha/2}$　　　　D. $\mu_{1-\alpha}$

(3) 一袋中装有 5 只球，编号为 1，2，3，4，5. 在袋中同时取 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.

(4) 一批产品共 10 件，其中有 7 件正品，3 件次品，每次从这批产品中任取 1 件，取出的产品仍放回去，求直至取到正品为止所需次数 X 的分布律.

(5) 考虑为期一年的一张保险单，若投保人在投保后一年内因意外死亡，则公司赔付 20 万元，若投保人因其他原因死亡，则公司赔付 5 万元，若投保人在投保期末生存，则公司无需付给任何费用. 若投保人在一年内因意外死亡的概率为 0.000 2，因其他原因死亡的概率为 0.001 0，求公司赔付金额的分布律.

(6) 进行重复独立试验，设每次试验的成功概率为 p，失败概率为 $q=1-p(0<p<1)$.

① 将试验进行到出现一次成功为止，以 X 表示所需的试验次数，求 X 的分布律（此时称 X 服从以 p 为参数的**几何分布**）.

② 将试验进行到出现 r 次成功为止，以 Y 表示所需的试验次数，求 Y 的分布律（此时称 Y 服从以 r,p 为参数的**巴斯卡分布**或**负二项分布**）.

③ 一篮球运动员的投篮命中率为 45%，以 X 表示他首次投中时累计已投篮的次数，写出 X 的分布律，并计算 X 取偶数的概率.

(7) 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 2 次的概率.

(8) 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 次重复独立试验，求指示灯发出信号的概率；

② 进行了 7 次重复独立试验，求指示灯发出信号的概率.

(9) 某地块岩层上有一个 10 m 深的土层，石块随机分布在土层内，建房时设计的桩群要打到岩层. 设土层可以分为 5 个独立层，每层深 2 m，打桩时每一个 2 m 层内碰到一块石头的概率为 0.1(碰到两块或更多块石头的概率忽略不计). 试求：

① 一根桩成功地打到岩层而未碰到任何石块的概率；

② 打到岩层时一根桩最多碰到一块石头的概率；

③ 打到岩层时一根桩恰有两次碰到石块的概率；

④ 一根桩一直打到第四层才第一次碰到石块的概率；

⑤ 假设一座房屋的地基要求有一组 9 根这样的桩打到岩层，各桩打入情况相互独立，求打桩时不碰到石块的概率.

(10) 有甲、乙两种味道和颜色都极为相似的名酒各 4 杯，如果从中挑 4 杯，能将甲种酒全都挑出来，算是试验成功一次.

① 某人随机地猜，问他试验成功一次的概率是多少?

② 某人声称他通过品尝能区分两种酒，他连续试验 10 次，成功了 3 次，试推断他是猜对的，还是他确有区分的能力(设各次试验是相互独立的).

(11) 已知某种疾病的发病率为 0.001，某单位共有 5 000 人，问该单位患有这种疾病的人数不超过 5 人的概率是多少?

(12) 某一城市每天发生火灾的次数 X 服从参数为 $\lambda = 0.8$ 的泊松分布，求该市一天内发生 3 次或 3 次以上火灾的概率.

(13) 在长度为 t 的时间间隔内，某急救中心收到紧急呼叫的次数 X 服从参数为 $\frac{t}{2}$ 的泊松分布，而与时间间隔的起点无关(时间以小时计)，求：

① 某一天从中午 12 时至下午 3 时没有收到紧急呼救的概率；

② 某一天从中午 12 时至下午 5 时至少收到 1 次紧急呼救的概率.

(14) 设在时间 t(分钟) 内，通过某交叉路口的汽车数服从参数与 t 成正比的泊松分布，已知在 1 min 内没有汽车通过的概率为 0.2，求在 2 min 内最多有一辆汽车通过的概率.

(15) 某服务台在长度为 t 的时间间隔内收到呼叫的次数 X 服从参数为 t 的泊松分布，时间 t(单位：min)，与间隔的起点无关，若服务员离开就有可能影响工作，求因服务员离开影响工作的概率.

① 离开 1 min；② 离开 3 min；③ 离开 5 min.

(16) 在保险公司里有 2 500 名同一年龄和同社会阶层的人参加了人寿保险，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 2 000 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 10 000 元和 20 000 元的概率.

(17) 设 $F(x)=\begin{cases}0, & x<0,\\ \dfrac{x}{2}, & 0\leqslant x<1,\\ 1, & x\geqslant 1,\end{cases}$ 问 $F(x)$ 是否为某随机变量的分布函数.

(18) 在区间 $[0,a]$ 上任意投掷一个质点，以 X 表示这个质点的坐标，设这个质点落在 $[0,a]$ 中任意小区间的概率与这个小区间的长度成正比例，试求 X 的分布函数.

(19) 设随机变量 X 的分布律为

X	0	1	2
P	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

求 $F(x)$.

(20) 从一批有 13 个正品和 2 个次品的产品中任意取 3 个，求取得的次品数 X 的分布律和分布函数，并求 $P\{\frac{1}{2}<X\leqslant\frac{5}{2}\}$.

(21) 设离散型随机变量 X 的分布函数

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

试写出 X 的分布律.

(22) 设连续型随机变量 X 的分布函数

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

① 求 $P\{X<2\}$，$P\{0<X<3\}$，$P\{2<X\leqslant 2.5\}$；

② 求随机变量的概率密度 $f(x)$.

(23) 设连续型随机变量 X 的分布函数

$$F(x)=\begin{cases}A, & x<0,\\ Bx^2, & 0\leqslant x<1,\\ Cx-\dfrac{1}{2}x^2-1, & 1\leqslant x<2,\\ 1, & x\geqslant 2,\end{cases}$$

求:① 常数 A,B,C;② X 的概率密度;③ $P\{X>\dfrac{1}{2}\}$.

(24) 已知随机变量 X 的概率密度 $f(x)=\begin{cases}2x, & 0<x<1,\\ 0, & \text{其他},\end{cases}$ 求:① $P\{X\leqslant 0.5\}$;② $P\{X=0.5\}$;③ $F(x)$.

(25) 服从拉普拉斯分布的随机变量 X 的概率密度为 $f(x)=Ae^{-|x|}$,试求:① 系数 A;② $P\{0<X<1\}$;③ X 的分布函数.

(26) 设随机变量 X 的概率密度 $f(x)=\begin{cases}x, & 0\leqslant x<1,\\ 2-x, & 1\leqslant x<2,\\ 0, & \text{其他},\end{cases}$ 求 X 的分布函数 $F(x)$,并画出 $f(x)$ 及 $F(x)$ 的图形.

(27) 设 K 在 $(-3,3)$ 上服从均匀分布,现有方程 $4x^2+4Kx+K+2=0$,求方程有实根的概率.

(28) 某型号晶体管的寿命 X(单位:h) 服从参数为 $\lambda=\dfrac{1}{200}$ 的指数分布,求:

① 晶体管寿命不超过 100 h 的概率;

② 晶体管寿命超过 300 h 的概率;

③ 一个晶体管的寿命不超过 100 h,而另一个晶体管的寿命为 100 ~ 300 h 的概率.

(29) 设顾客在某银行的窗口等待服务的时间 X(单位:min) 服从 $\lambda=\dfrac{1}{5}$ 的指数分布,其概率密度

$$f(x)=\begin{cases}\dfrac{1}{5}e^{-\frac{1}{5}x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

某顾客在窗口等待服务,若超过 10 min,他就离开.

① 设某顾客某天去银行,求他未等到服务就离开的概率;

② 设某顾客一个月要去银行 5 次,求他 5 次中至多有 1 次未等到服务而离开的概率.

(30) 设 $X\sim N(3,2^2)$.

① 求 $P\{2<X\leqslant 5\}$,$P\{-4<X\leqslant 10\}$,$P\{|x|>2\}$,$P\{X>3\}$;

② 确定 c，使得 $P\{X>c\}=P\{X\leqslant c\}$；

③ 设 d 满足 $P\{X>d\}\geqslant 0.9$，问 d 至多为多少？

(31) 公共汽车车门高度，是按男子与车门碰头机会在 0.01 以下来设计的. 设男子身高 X 服从 $\mu=168$ cm，$\sigma=7$ cm 的正态分布，即 $X\sim N(168,7^2)$，问车门的高度应如何确定.

(32) 某企业准备通过招聘考试招收职工，根据考试分数，从高分到低分分别录取正式工 280 人，临时工 20 人，报考的人数是 1 657 人，考试满分是 400 分. 考试后得知，考试总平均成绩 $\mu=166$ 分，360 分以上的高分考生 31 人，某考生得 256 分，问他能否被录取？能否被聘为正式工？

(33) 设某种晶体管的寿命 X(单位：h) 的概率密度

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

① 若一个晶体管在使用 150 h 后仍完好，那么该晶体管使用时间少于 200 h 的概率是多少？

② 若一个电子仪器中装有 3 个独立工作的这样的晶体管，在使用 150 h 内恰有一个晶体管损坏的概率是多少？

(34) 在电源电压不超过 200 V、为 200 ～ 240 V 和超过 240 V 这 3 种情况下，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；

② 该电子元件损坏时，电源电压为 220 ～ 240 V 的概率.

(35) 设 $f(x)$，$g(x)$ 都是概率密度，求证

$$h(x)=\alpha f(x)+(1-\alpha)g(x),0\leqslant\alpha\leqslant 1$$

也是一个概率密度.

(36) 设随机变量 X 的分布律为

X	-1	0	1	2	$\frac{5}{2}$
P	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{3}{10}$

① 求 $2X$ 的分布律；

② 求 X^2 的分布律.

(37) 设 X 的分布律为 $P\{X=k\}=\dfrac{1}{2^k}$，$k=1,2,3,\cdots$，求 $Y=\sin\left(\dfrac{\pi}{2}X\right)$ 的分布律.

(38) 设随机变量 X 的概率密度 $f_X(x)=\dfrac{1}{\pi(1+x^2)}$，求随机变量 $Y=1-\sqrt[3]{X}$ 的概率密度 $f_Y(y)$.

(39) 随机变量 X 的概率密度为 $f(x)$，求 $Y=\sqrt{X}$ 的概率密度 $f_Y(y)$.

(40) 设随机变量 X 的概率密度 $f_X(x)=\begin{cases}x^3\mathrm{e}^{-x^2}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$ 求 $Y=2X+3$ 的概率密度.

(41) 设随机变量 X 的概率密度 $f(x)=\begin{cases}\mathrm{e}^{-x}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$

① 求 $Y=\mathrm{e}^X$ 的概率密度；② 求 $Y=X^2$ 的概率密度.

(42) 设随机变量 X 的概率密度 $f(x)=\begin{cases}\dfrac{2x}{\pi^2}, & 0<x<\pi,\\ 0, & \text{其他},\end{cases}$ 求 $Y=\sin X$ 的概率密度.

(43) 设电流 I 是一个随机变量，它均匀分布在 $9\sim 11$ 安培之间，若此电流通过 2 欧姆的电阻，在其上消耗的功率 $W=2I^2$，求 W 的概率密度.

(44) 设随机变量 X 的分布函数 $F(x)$ 是严格单调的连续函数，证明 $Y=F(X)$ 服从 $[0,1]$ 上的均匀分布.

第3章

多维随机变量及其分布

在实际应用中，有些随机试验的结果需要同时用两个或两个以上的随机变量来描述. 例如，在医疗诊所中，通常涉及几个试验指标，每一个指标都是一个随机变量；某地区气候状况需要考虑温度、湿度、风力等多个随机变量. 又如，研究某股票的投资价值，要考虑股票的市盈率、市净率、资本报酬率、净值周转率等指标，它们也都是一些随机变量. 往往这些随机变量并非彼此孤立地存在，在这种情况下，不但要研究多个随机变量各自的统计规律，还要研究它们之间的相依关系. 因此，还需要考察它们联合取值的统计规律，即多维随机变量的分布. 由于二维与 $n(n\geqslant 3)$ 维随机变量的研究方法和所得结果没有本质的区别，为简单起见，本章重点讨论二维随机变量，其结果也适用于$n(n\geqslant 3)$维的情形.

3.1　二维随机变量的分布函数

3.1.1　二维随机变量及其分布函数

定义 3.1　设 E 是一个随机试验，它的样本空间为 $S=\{e\}$，而 $X=X(e)$，$Y=Y(e)$ 是定义在 S 上的两个随机变量，称(X,Y) 为定义在 S 上的**二维随机变量**或**二维随机向量**.

注：① 一般地，称 n 个随机变量的整体 $X=(X_1,X_2,\cdots,X_n)$ 为 n 维随机变量或 n 维随机向量.

② 多维随机变量的关键是定义在同一样本空间上，对于不同样本空间上的多个随机变量，本书不讨论.

二维随机变量(X,Y)的性质不仅与X及Y有关，还依赖于这两个随机变量的相互关系，故需将(X,Y)作为一个整体来研究．类似一维随机变量，下面将引入分布函数的概念来研究二维随机变量．

定义 3.2　设(X,Y)是二维随机变量，对任意实数x,y，二元函数

$$F(x,y)=P\{(X\leqslant x)\cap(Y\leqslant y)\}\xlongequal{\text{记为}}P\{X\leqslant x,Y\leqslant y\}$$

称为二维随机变量(X,Y)的**分布函数**或X与Y的**联合分布函数**，记为$(X,Y)\sim F(x,y)$．

若将二维随机变量(X,Y)视为平面上随机点的坐标，则分布函数$F(x,y)$在(x,y)处的函数值，就是随机点(X,Y)落在无穷矩形域$\{(s,t)\mid s\leqslant x,t\leqslant y\}$内的概率，如图 3-1 所示．

有了分布函数$F(x,y)$，借助图 3-2，容易算出随机点落在矩形域$\{(x,y)\mid x_1<x\leqslant x_2,y_1<y\leqslant y_2\}$内的概率

$$P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1).$$

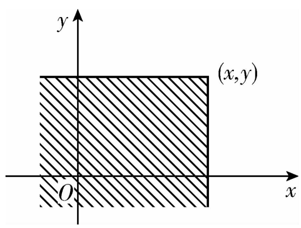


图 3-1

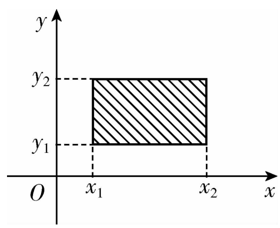


图 3-2

分布函数$F(x,y)$的基本性质：

(1) **有界性**　对任意的x和y，有$0\leqslant F(x,y)\leqslant 1$，且对任意固定的$y$，$F(-\infty,y)=0$，对任意固定的$x$，$F(x,-\infty)=0$，$F(-\infty,-\infty)=0$，$F(+\infty,+\infty)=1$．

(2) **单调性**　$F(x,y)$关于x和y均为单调不减函数，即对任意固定的y，当$x_2>x_1$时，$F(x_2,y)\geqslant F(x_1,y)$；对任意固定的$x$，当$y_2>y_1$时，$F(x,y_2)\geqslant F(x,y_1)$．

(3) **右连续**　$F(x,y)$关于x和y均为右连续，即$F(x,y)=F(x+0,y)$，$F(x,y)=F(x,y+0)$．

(4) **非负性**　对任意的$x_1<x_2$，$y_1<y_2$，有

$$P\{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\geqslant 0.$$

注：根据这些性质可以确定$F(x,y)$中的未知常数，也可以验证某二元函数是否为二维随机变量的分布函数．若某二元实值函数$F(x,y)$满足上述四个性质，则必存在二维随机变量(X,Y)以$F(x,y)$为分布函数．

3.1.2　二维随机变量的边缘分布函数

二维随机变量(X,Y)作为一个整体，具有分布函数$F(x,y)$，而分量X和Y都是一维随机变量，也各有对应的分布函数$F_X(x)$和$F_Y(y)$.

定义 3.3　称二维随机变量(X,Y)中X(或Y)的分布函数$F_X(x)$(或$F_Y(y)$)为(X,Y)关于X(或Y)的**边缘分布函数**，即

$$F_X(x)=P\{X\leqslant x\}=P\{X\leqslant x,Y<+\infty\}=F(x,+\infty),$$
$$F_Y(y)=P\{Y\leqslant y\}=P\{X<+\infty,Y\leqslant y\}=F(+\infty,y).$$

边缘分布函数具有一维分布函数的性质，由二维分布函数可唯一确定边缘分布函数，反之不然.

例 3.1　设(X,Y)的分布函数为

$$F(x,y)=\frac{1}{\pi^2}\left(A+\arctan\frac{x}{2}\right)\left(B+\arctan\frac{y}{3}\right),-\infty<x,y<+\infty.$$

(1) 求常数A,B.

(2) 求$F(x,y)$关于X和Y的边缘分布函数$F_X(x),F_Y(y)$.

解　(1) 利用分布函数$F(x,y)$的性质得

$$F(+\infty,+\infty)=1,$$
$$F(-\infty,-\infty)=0,$$

即

$$\begin{aligned}\lim_{\substack{x\to+\infty\\y\to+\infty}}F(x,y)&=\lim_{\substack{x\to+\infty\\y\to+\infty}}\frac{1}{\pi^2}\left(A+\arctan\frac{x}{2}\right)\left(B+\arctan\frac{y}{3}\right)\\&=\frac{1}{\pi^2}\left(A+\frac{\pi}{2}\right)\left(B+\frac{\pi}{2}\right)=1,\end{aligned}$$

$$\begin{aligned}\lim_{\substack{x\to-\infty\\y\to-\infty}}F(x,y)&=\lim_{\substack{x\to-\infty\\y\to-\infty}}\frac{1}{\pi^2}\left(A+\arctan\frac{x}{2}\right)\left(B+\arctan\frac{y}{3}\right)\\&=\frac{1}{\pi^2}\left(A-\frac{\pi}{2}\right)\left(B-\frac{\pi}{2}\right)=0,\end{aligned}$$

解得

$$A=B=\frac{\pi}{2}.$$

故

$$F(x,y)=\frac{1}{\pi^2}\left(\frac{\pi}{2}+\arctan\frac{x}{2}\right)\left(\frac{\pi}{2}+\arctan\frac{y}{3}\right).$$

(2) 由定义知

$$F_X(x)=F(x,+\infty)=\lim_{y\to+\infty}F(x,y)$$

$$= \lim_{y \to +\infty} \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right)\left(\frac{\pi}{2} + \arctan\frac{y}{3}\right)$$
$$= \frac{1}{\pi^2}\left(\frac{\pi}{2} + \arctan\frac{x}{2}\right) \cdot \pi$$
$$= \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2}, -\infty < x < +\infty.$$

同理可得

$$F_Y(y) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{3}, -\infty < y < +\infty.$$

3.2 二维离散型随机变量

3.2.1 二维离散型随机变量及其分布律与边缘分布律

定义 3.4 若二维随机变量(X,Y)的所有可能取值为有限对或可列无限多对，则称(X,Y)为**二维离散型随机变量**.

显然(X,Y)为二维离散型随机变量，当且仅当X和Y均为一维离散型随机变量.

定义 3.5 设二维离散型随机变量(X,Y)所有可能取值为$(x_i, y_j)(i,j=1,2,3,\cdots)$，称

$$P\{X = x_i, Y = y_j\} = p_{ij}, i,j = 1,2,3,\cdots$$

为二维离散型随机变量(X,Y)的**分布律**或X与Y的**联合分布律**.

由概率的性质，易得分布律具有以下性质：

(1) **非负性** $p_{ij} \geqslant 0, i,j = 1,2,3,\cdots$.

(2) **归一性** $\sum\limits_{i=1}^{+\infty}\sum\limits_{j=1}^{+\infty} p_{ij} = 1$.

注：① 上述两条性质常用于检验(X,Y)的分布律正确性及确定(X,Y)分布律中所含的未知参数.

② 二维离散型随机变量(X,Y)作为一个整体具有分布律，而分量X和Y都是一维随机变量，也各有对应的分布律.

定义 3.6 称二维离散型随机变量(X,Y)中X(或Y)的分布律为(X,Y)关于X(或Y)的**边缘分布律**，即

$$P\{X=x_i\}=\sum_{j=1}^{+\infty}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{+\infty}p_{ij},i=1,2,3,\cdots$$

$$P\{Y=y_j\}=\sum_{i=1}^{+\infty}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{+\infty}p_{ij},j=1,2,3,\cdots$$

通常记 $P\{X=x\}=p_{i\cdot},i=1,2,3,\cdots;P\{Y=y_j\}=p_{\cdot j},j=1,2,3,\cdots$.

二维离散型随机变量(X,Y)的分布律和边缘分布律也可列成表格的形式

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

上表中最后一列是(X,Y)关于 X 的边缘分布律，$p_{i\cdot}$ 是表中第 i 行各数之和；最后一行是(X,Y)关于 Y 的边缘分布律，$p_{\cdot j}$ 是表中第 j 列各数之和.

边缘分布律具有一维分布律的性质，由二维随机变量的分布律可唯一确定边缘分布律，反之不然.

注：① 由离散型随机变量(X,Y)的分布律可以确定(X,Y)的分布函数

$$F(x,y)=\sum_{x_i\leqslant x}\sum_{y_j\leqslant y}p_{ij},$$

其中和式是对一切满足 $x_i\leqslant x,y_j\leqslant y$ 的 i,j 来求和的.

② 对离散型随机变量而言，(X,Y)的分布律不仅比分布函数更直观，还能够容易地求出(X,Y)取值于任何区域 D 上的概率，即

$$P\{(X,Y)\in D\}=\sum_{(x_i,y_j)\in D}\sum p_{ij},$$

其中和式是对一切满足$(x_i,y_j)\in D$ 的 i,j 来求和的.

③ 求二维离散型随机变量(X,Y)的分布律，首先确定(X,Y)的所有可能取值，然后计算取各对可能值的概率.

例 3.2　从含有 3 个正品，2 个次品的 5 个产品中依次抽取 2 个，采用两种方式：(1) 放回抽样；(2) 不放回抽样. 设 X 表示第一次取到的次品的个数，Y 表示第二次取到的次品的个数，求(X,Y)的联合分布律及其边缘分布律.

解 由题意知,X 及 Y 各自的可能取值均为 0,1,于是(X,Y)只能取$(0,0)$,$(0,1)$,$(1,0)$和$(1,1)$,而其可能取值的概率可由乘法原理得到.

(1) 放回抽样:

$$P\{X=0,Y=0\}=\frac{3}{5}\times\frac{3}{5}=\frac{9}{25},P\{X=0,Y=1\}=\frac{3}{5}\times\frac{2}{5}=\frac{6}{25},$$

$$P\{X=1,Y=0\}=\frac{2}{5}\times\frac{3}{5}=\frac{6}{25},P\{X=1,Y=1\}=\frac{2}{5}\times\frac{2}{5}=\frac{4}{25}.$$

于是(X,Y)的分布律及其边缘分布律为

X \ Y	0	1	$P\{X=x_i\}$
0	$\frac{9}{25}$	$\frac{6}{25}$	$\frac{3}{5}$
1	$\frac{6}{25}$	$\frac{4}{25}$	$\frac{2}{5}$
$P\{Y=y_j\}$	$\frac{3}{5}$	$\frac{2}{5}$	1

(2) 不放回抽样:

$$P\{X=0,Y=0\}=\frac{3}{5}\times\frac{2}{4}=\frac{3}{10},P\{X=0,Y=1\}=\frac{3}{5}\times\frac{2}{4}=\frac{3}{10}.$$

$$P\{X=1,Y=0\}=\frac{2}{5}\times\frac{3}{4}=\frac{3}{10},P\{X=1,Y=1\}=\frac{2}{5}\times\frac{1}{4}=\frac{1}{10}.$$

于是(X,Y)的分布律及边缘分布律为

X \ Y	0	1	$P\{X=x_i\}$
0	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{3}{5}$
1	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{5}$
$P\{Y=y_j\}$	$\frac{3}{5}$	$\frac{2}{5}$	1

从例 3.2 可看出,对于两种情况,X,Y 的边缘分布律是相同的,但(X,Y)的分布律不同.这就是说,由联合分布律可得到边缘分布律,但由边缘分布律却不一定能确定联合分布律.

常常将边缘分布律写在联合分布律表格的边缘上,这就是"边缘分布律"这个名词的由来.

例 3.3　盒子里有 2 个黑球、2 个红球、2 个白球，在其中任取 2 个球，以 X 表示取得的黑球的个数，以 Y 表示取得的红球的个数，试写出 (X,Y) 的分布律，并求事件 $\{X+Y\leqslant 1\}$ 的概率.

解　X,Y 各自的可能取值均为 0,1,2. 由题设知，(X,Y) 取 $(1,2)$，$(2,1)$ 和 $(2,2)$ 均不可能，取其他值的概率可由古典概率计算.

从 6 个球中任取 2 个一共有 $C_6^2=15$ 种取法. (X,Y) 取 $(0,0)$ 表示取得的 2 个球是白球，其取法只有 1 种，所以其概率

$$P\{X=0,Y=0\}=\frac{1}{15}.$$

类似地，(X,Y) 取其他几对数组的概率

$$P\{X=0,Y=1\}=P\{X=1,Y=0\}=P\{X=1,Y=1\}=\frac{2\times 2}{15}=\frac{4}{15},$$

$$P\{X=2,Y=0\}=P\{X=0,Y=2\}=\frac{1}{15}.$$

因此，(X,Y) 的分布律为

X \ Y	0	1	2
0	$\frac{1}{15}$	$\frac{4}{15}$	$\frac{1}{15}$
1	$\frac{4}{15}$	$\frac{4}{15}$	0
2	$\frac{1}{15}$	0	0

由于事件 $\{X+Y\leqslant 1\}$ 包含三个基本事件，分别对应着 $(0,0)$，$(0,1)$ 和 $(1,0)$，所以

$$\begin{aligned}P\{X+Y\leqslant 1\}&=P\{X=0,Y=0\}+P\{X=0,Y=1\}+P\{X=1,Y=0\}\\&=\frac{1}{15}+\frac{4}{15}+\frac{4}{15}=\frac{3}{5}.\end{aligned}$$

3.2.2　条件分布

对于二维随机变量 (X,Y)，有时需要考虑其中一个随机变量取定某个值时，另一个随机变量的概率分布，通常称为条件分布，条件分布刻画了随机变量 X 与 Y 之间的相依性，借助于随机事件条件概率的概念，下面引入随机变量的条件分布.

定义 3.7　设二维离散型随机变量 (X,Y) 的分布律为

$$P\{X=x_i,Y=y_j\}=p_{ij},i,j=1,2,3,\cdots$$

对于固定的 j，若 $P\{Y=y_j\}=p_{\cdot j}>0$，称

$$P\{X=x_i \mid Y=y_j\}=\frac{p\{X=x_i,Y=y_j\}}{p\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},i=1,2,3,\cdots$$

为 $Y=y_j$ 条件下 X 的**条件分布律**,简称为**条件分布**.

同样,对于固定的 i,若 $P\{X=x_i\}=p_{i\cdot}>0$,称

$$P\{Y=y_j \mid X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\frac{p_{ij}}{p_{i\cdot}},j=1,2,3,\cdots$$

为 $X=x_i$ 条件下 Y 的**条件分布律**.

注:① 条件分布是一种概率分布,它具有概率分布的一切性质.

② 条件分布的定义表明,二维离散型随机变量的分布不但确定了其边缘分布,而且确定了其条件分布.

例 3.4 设(X,Y)的分布律为

X \ Y	-1	0	2
0	0.1	0.2	0
1	0.3	0.05	0.1
2	0.15	0	0.1

求 $Y=0$ 时,X 的条件分布律.

解 由题意知

$$P\{Y=0\}=0.2+0.05+0=0.25,$$

因此,在 $Y=0$ 时,X 的条件分布律

$$P\{X=0 \mid Y=0\}=\frac{P\{X=0,Y=0\}}{P\{Y=0\}}=\frac{0.2}{0.25}=0.8,$$

$$P\{X=1 \mid Y=0\}=\frac{P\{X=1,Y=0\}}{P\{Y=0\}}=\frac{0.05}{0.25}=0.2,$$

$$P\{X=2 \mid Y=0\}=\frac{P\{X=2,Y=0\}}{P\{Y=0\}}=\frac{0}{0.25}=0,$$

也可写成

$X \mid Y=0$	0	1	2
P	0.8	0.2	0

例 3.5 设某公交车起始站上车人数 X 服从参数为$\lambda(\lambda>0)$的泊松分布,每位乘客在中途下车的概率为 $p(0<p<1)$,且中途下车与否相互独立,以 Y 表示中途下车的人数,求:

(1) 在发车时有 n 个乘客的条件下，中途有 m 人下车的概率.

(2) 二维随机变量 (X,Y) 的分布律.

(3) 中途下车人数 Y 的分布律.

解　(1) 由题意知，$\{X=n\}$ 表示发车时有 n 个乘客，$\{Y=m\}$ 表示中途有 m 个人下车，则发车时有 $X=n$ 个乘客的条件下，中途下车的人数 Y 的条件分布为二项分布 $b(n,p)$，即当 $n=0,1,2,\cdots$ 时，

$$P\{Y=m \mid X=n\}=\mathrm{C}_n^m p^m(1-p)^{n-m}, m=0,1,2,\cdots,n.$$

(2) 由题意知，$X\sim\pi(\lambda)$，则有

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

于是 (X,Y) 的分布律

$$\begin{aligned}P\{X=n,Y=m\}&=P\{X=n\}P\{Y=m \mid X=n\}\\&=\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}\cdot\mathrm{C}_n^m p^m(1-p)^{n-m}(n=0,1,2,\cdots;m=0,1,2,\cdots,n).\end{aligned}$$

$$\begin{aligned}(3)\ P\{Y=m\}&=\sum_{n=m}^{+\infty}P\{X=n,Y=m\}\\&=\sum_{n=m}^{+\infty}\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}\cdot\mathrm{C}_n^m p^m(1-p)^{n-m}\\&=\sum_{n=m}^{+\infty}\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}\frac{n!}{m!(n-m)!}p^m(1-p)^{n-m}\\&=\mathrm{e}^{-\lambda}\frac{(\lambda p)^m}{m!}\sum_{n=m}^{+\infty}\frac{[(1-p)\lambda]^{n-m}}{(n-m)!}\\&=\mathrm{e}^{-\lambda}\frac{(\lambda p)^m}{m!}\mathrm{e}^{(1-p)\lambda}\\&=\frac{(\lambda p)^m}{m!}\mathrm{e}^{-\lambda p}(m=0,1,2,\cdots),\end{aligned}$$

即 $Y\sim\pi(\lambda p)$.

3.3　二维连续型随机变量

3.3.1　二维连续型随机变量及其分布

定义 3.8　设 $F(x,y)$ 为二维随机变量 (X,Y) 的分布函数，若存在非负可积函

数 $f(x,y)$ 使得对于任意实数 x,y 有

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

则称 (X,Y) 为**二维连续型随机变量**，并称 $f(x,y)$ 为 (X,Y) 的**概率密度**或随机变量 X 和 Y 的**联合概率密度**，记为 $(X,Y)\sim f(x,y)$.

概率密度 $f(x,y)$ 具有以下性质：

(1) **非负性**　$f(x,y)\geqslant 0$.

(2) **归一性**　$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1$.

注：若一个二元函数具有以上两条性质，则此二元函数可作为某二维随机变量的概率密度.

(3) 若 D 为 xOy 平面上的一个区域，则 (X,Y) 落入 D 的概率为

$$P\{(X,Y)\in D\}=\iint_{D}f(x,y)\mathrm{d}x\mathrm{d}y.$$

(4) 若 $f(x,y)$ 在点 (x,y) 处连续，则有

$$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y).$$

在几何上，$Z=f(x,y)$ 表示空间的一个曲面，由性质(2)知，介于它和 xOy 平面之间的空间区域的体积为1. 由性质(3)知，$P\{(X,Y)\in D\}$ 的值等于以 D 为底，以曲面 $Z=f(x,y)$ 为顶的柱体体积.

例 3.6　设二维连续型随机变量 (X,Y) 的概率密度

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

求：(1) 系数 A；(2) $P\left\{X\geqslant\frac{3}{4}\right\}$；(3) $P\left\{X<\frac{1}{4},Y<\frac{1}{2}\right\}$.

解　(1) 由 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1$，即 $\int_{0}^{1}\mathrm{d}x\int_{0}^{x}Ax\mathrm{d}y=1$，求得 $A=3$.

(2) $P\left\{X\geqslant\frac{3}{4}\right\}=\int_{\frac{3}{4}}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=\int_{\frac{3}{4}}^{1}\mathrm{d}x\int_{0}^{x}3x\mathrm{d}y=\frac{37}{64}$.

(3) $P\left\{X<\frac{1}{4},Y<\frac{1}{2}\right\}=\int_{-\infty}^{\frac{1}{4}}\int_{-\infty}^{\frac{1}{2}}f(x,y)\mathrm{d}x\mathrm{d}y=\int_{0}^{\frac{1}{4}}\mathrm{d}x\int_{0}^{x}3x\mathrm{d}y=\frac{1}{64}$.

例 3.7　设二维连续型随机变量 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}\dfrac{1}{4x^2y^3}, & x>\dfrac{1}{2},y>\dfrac{1}{2},\\ 0, & \text{其他}.\end{cases}$$

求：(1) (X,Y) 的分布函数 $F(x,y)$；(2) $P\{XY<1\}$.

解　(1) $F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v=\begin{cases}\int_{\frac{1}{2}}^{x}\mathrm{d}u\int_{\frac{1}{2}}^{y}\frac{1}{4u^2v^3}\mathrm{d}v, & x>\frac{1}{2},y>\frac{1}{2},\\ 0, & \text{其他}\end{cases}$

$$=\begin{cases}\left(1-\frac{1}{2x}\right)\left(1-\frac{1}{4y^2}\right), & x>\frac{1}{2},y>\frac{1}{2},\\ 0, & \text{其他.}\end{cases}$$

(2) 记 $G=\{(x,y)\mid xy<1\}$，$D=\left\{(x,y)\mid x>\frac{1}{2},y>\frac{1}{2}\right\}$(见图 3-3)，因 $f(x,y)$ 仅在 D 内取非零值，故

$$P\{XY<1\}=\iint_{G}f(x,y)\mathrm{d}x\mathrm{d}y=\iint_{G\cap D}\frac{1}{4x^2y^3}\mathrm{d}x\mathrm{d}y$$

$$=\int_{\frac{1}{2}}^{2}\mathrm{d}x\int_{\frac{1}{2}}^{\frac{1}{x}}\frac{1}{4x^2y^3}\mathrm{d}y=\frac{9}{16}.$$

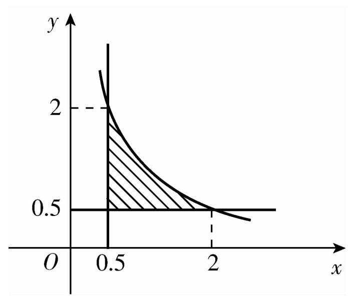


图 3-3

3.3.2　二维连续型随机变量的边缘分布

定义 3.9　设 $f(x,y)$ 为二维连续型随机变量 (X,Y) 的概率密度，称

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\right]\mathrm{d}x$$

为 (X,Y) 关于 X 的**边缘分布函数**，称

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$$

为 (X,Y) 关于 X 的**边缘概率密度**.

同样可定义 (X,Y) 关于 Y 的边缘分布函数及边缘概率密度，它们分别是

$$F_Y(y)=F(+\infty,y)=\int_{-\infty}^{y}\left[\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\right]\mathrm{d}y,$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x.$$

例 3.8 设二维随机变量(X,Y)的概率密度

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

求边缘概率密度 $f_X(x)$ 和 $f_Y(y)$.

解 $f(x,y)$ 的非零区域如图 3-4 所示.

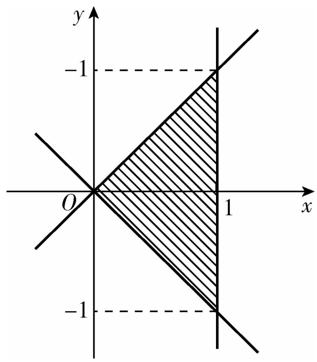


图 3-4

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\int_{-x}^{x}\mathrm{d}y, & 0<x<1,\\ 0, & \text{其他}\end{cases}=\begin{cases}2x, & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\begin{cases}\int_{-y}^{1}\mathrm{d}x, & -1<y<0,\\ \int_{y}^{1}\mathrm{d}x, & 0\leqslant y<1,\\ 0, & \text{其他}\end{cases}=\begin{cases}1+y, & -1<y<0,\\ 1-y, & 0\leqslant y<1,\\ 0, & \text{其他}.\end{cases}$$

3.3.3 常用二维连续型随机变量的分布

1）二维均匀分布

定义 3.10 设 D 为 xOy 平面上的有界区域，面积为 S_D，若二维随机变量(X,Y)的概率密度

$$f(x,y)=\begin{cases}\dfrac{1}{S_D}, & (x,y)\in D,\\ 0, & \text{其他},\end{cases}$$

则称(X,Y)在 D 上服从**均匀分布**，记为$(X,Y)\sim U(D)$.

若(X,Y)在区域 D 上服从二维均匀分布，则对于任一平面区域 G 有

$$P\{(X,Y)\in G\}=\iint\limits_{G} f(x,y)\mathrm{d}x\mathrm{d}y=\iint\limits_{G\cap D}\frac{1}{S_D}\mathrm{d}x\mathrm{d}y$$
$$=\frac{1}{S_D}\iint\limits_{G\cap D}\mathrm{d}x\mathrm{d}y=\frac{S_{G\cap D}}{S_D},$$

其中 $S_{G\cap D}$ 为平面区域G 与 D 公共部分的面积(见图 3-5).

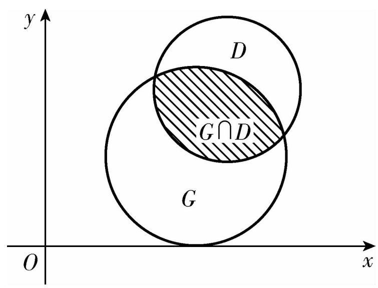


图 3-5

特别地,当 $G\subset D$ 时,有 $P\{(X,Y)\in G\}=\dfrac{S_G}{S_D}$.

这说明 D 上二维均匀分布随机变量(X,Y) 落入 D 内任意子区域G 内的概率只与区域G 的面积有关,而与区域G 的形状及位置无关. 因此,二维均匀分布常用作随机数点的数学模型.

例如,向平面上有界区域 D 上任投一质点,若质点落在 D 内任一小区域G 的概率与小区域的面积成正比,而与 G 的位置无关,则质点的坐标(X,Y) 在 G 上服从均匀分布.

例 3.9　设 D 为由曲线 $y=x^2$ 和 $y=\sqrt{x}$ 围成的平面区域(见图 3-6),(X,Y) 在 D 上服从均匀分布.

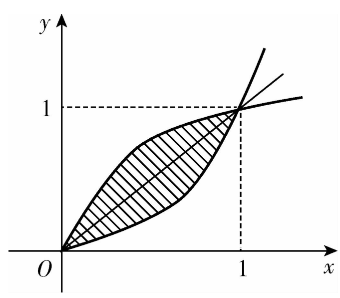


图 3-6

求:(1) $P\{X>Y\}$;(2) (X,Y) 的边缘概率密度.

解 区域 D 的面积

$$S_D=\int_0^1(\sqrt{x}-x^2)\mathrm{d}x=\frac{1}{3},$$

故 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}3, & (x,y)\in D,\\ 0, & \text{其他}.\end{cases}$$

(1) 设 $G=\{(x,y)\mid x>y\}$，则

$$P\{X>Y\}=P\{(X,Y)\in G\}=\frac{S_{G\cap D}}{S_D}=\frac{\frac{1}{6}}{\frac{1}{3}}=\frac{1}{2}.$$

(2) $$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\int_{x^2}^{\sqrt{x}}3\mathrm{d}y, & 0<x<1,\\ 0, & \text{其他}\end{cases}=\begin{cases}3(\sqrt{x}-x^2), & 0<x<1,\\ 0, & \text{其他}.\end{cases}$$

由 X 和 Y 在问题中地位的对称性，将上式中的 x 改成 y，即得到 Y 的边缘概率密度

$$f_Y(y)=\begin{cases}3(\sqrt{y}-y^2), & 0<y<1,\\ 0, & \text{其他}.\end{cases}$$

注：① 在例 3.9 中，二维均匀分布随机变量 (X,Y) 的两个边缘分布都不再是均匀分布.

② 容易得到，服从矩形域 $a<x<b,c<y<d$ 上的均匀分布 (X,Y) 的两个边缘分布仍为均匀分布，且分别为

$$f_X(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其他}\end{cases}\quad \text{和}\quad f_Y(y)=\begin{cases}\dfrac{1}{d-c}, & c<y<d,\\ 0, & \text{其他}.\end{cases}$$

2）二维正态分布

定义 3.11 若二维随机变量 (X,Y) 的概率密度

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},$$

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 均为常数，且 $-\infty<\mu_1,\mu_2<+\infty$，$-\infty<x,y<+\infty$，$\sigma_1,\sigma_2>0$，$|\rho|<1$，则称 (X,Y) 服从**二维正态分布**，记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

$f(x,y)$ 的图形如图 3-7 所示.

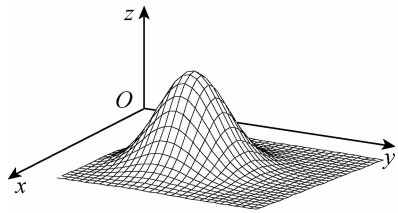


图 3-7

例 3.10　设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，试求二维正态分布的两个边缘概率密度 $f_X(x)$ 和 $f_Y(y)$.

解　由于$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，所以

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$$
$$=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}\mathrm{d}y,$$

通过配方可得

$$\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}$$
$$=\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2+(1-\rho^2)\frac{(x-\mu_1)^2}{\sigma_1^2},$$

于是

$$f_X(x)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2}\mathrm{d}y$$
$$=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2}\mathrm{d}y.$$

令 $t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)$，则有

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},$$

即

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},-\infty<x<+\infty.$$

同理可得

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},-\infty<y<+\infty.$$

注：上述结果表明，二维正态随机变量的两个边缘分布都是一维正态分布，且都不依赖于参数 ρ，即对给定的 $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$，不同的 ρ 对应不同的二维正态分布，但它们的边缘分布都是相同的. 因此，由关于 X 和关于 Y 的边缘分布，一般来说不能确定二维随机变量 (X,Y) 的分布.

例 3.11 设二维随机变量 (X,Y) 的概率密度

$$f(x,y)=\frac{1}{2\pi}\mathrm{e}^{-\frac{1}{2}(x^2+y^2)}(1+x^3y),$$

试求关于 X 和关于 Y 的边缘概率密度.

解 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$,

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y^2}{2}}.$$

此例说明，边缘分布均为正态分布的二维随机变量，其联合分布不一定是二维正态分布.

3.3.4 条件分布

与二维离散型随机变量一样，二维连续型随机变量也需要考虑条件分布问题. 但是，因为连续型随机变量取某个值的概率为零，即对任意 x,y，有 $P\{X=x\}=0$，$P\{Y=y\}=0$，所以无法直接用条件概率的公式得到连续型随机变量的条件分布.

设 (X,Y) 的概率密度为 $f(x,y)$，边缘概率密度为 $f_X(x)$，$f_Y(y)$. 给定 y，对任意给定的 $\varepsilon>0$，对于任意的 x，将 $P\{X\leqslant x\mid Y=y\}$ 看成是 $\varepsilon\to0^+$ 时 $P\{X\leqslant x\mid y\leqslant Y\leqslant y+\varepsilon\}$ 的极限.

设 $P\{y\leqslant Y\leqslant y+\varepsilon\}>0$，则有

$$\begin{aligned}P\{X\leqslant x\mid Y=y\}&=\lim_{\varepsilon\to0^+}P\{X\leqslant x\mid y\leqslant Y\leqslant y+\varepsilon\}\\&=\lim_{\varepsilon\to0^+}\frac{P\{X\leqslant x,y\leqslant Y\leqslant y+\varepsilon\}}{P\{y\leqslant Y\leqslant y+\varepsilon\}}\\&=\lim_{\varepsilon\to0^+}\frac{\int_{-\infty}^{x}\left[\int_{y}^{y+\varepsilon}f(x,y)\mathrm{d}y\right]\mathrm{d}x}{\int_{y}^{y+\varepsilon}f_Y(y)\mathrm{d}y}.\end{aligned}$$

当 $f_Y(y)$，$f(x,y)$ 在 y 处连续时，由积分中值定理得

$$\int_{y}^{y+\varepsilon}f_Y(y)\mathrm{d}y=\varepsilon f_Y(\xi_1),$$

$$\int_{y}^{y+\varepsilon}f(x,y)\mathrm{d}y=\varepsilon f(x,\xi_2),$$

其中 $y \leqslant \xi_1, \xi_2 \leqslant y+\varepsilon$，于是

$$P\{X \leqslant x \mid Y=y\}=\lim_{\varepsilon \to 0^{+}} \frac{\int_{-\infty}^{x} f(x, \xi_2) \mathrm{d} x}{f_Y(\xi_1)}=\int_{-\infty}^{x} \frac{f(x, y)}{f_Y(y)} \mathrm{d} x .$$

至此，可以定义连续型随机变量的条件分布.

定义 3.12　设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$，(X,Y) 关于 Y 的边缘概率密度为 $f_Y(y)$，若对于固定的 y，$f_Y(y)>0$，则称

$$F_{X|Y}(x \mid y)=P\{X \leqslant x \mid Y=y\}=\int_{-\infty}^{x} \frac{f(x, y)}{f_Y(y)} \mathrm{d} x$$

为在 $Y=y$ 条件下 X 的**条件分布函数**，称

$$f_{X|Y}(x \mid y)=\frac{f(x, y)}{f_Y(y)}$$

为在 $Y=y$ 条件下 X 的**条件概率密度**.

同理，对固定的 x，若 $f_X(x)>0$，则称

$$F_{Y|X}(y \mid x)=P\{Y \leqslant y \mid X=x\}=\int_{-\infty}^{y} \frac{f(x, y)}{f_X(x)} \mathrm{d} y$$

为在 $X=x$ 条件下 Y 的**条件分布函数**，称

$$f_{Y|X}(y \mid x)=\frac{f(x, y)}{f_X(x)}$$

为在 $X=x$ 条件下 Y 的**条件概率密度**.

例 3.12　设 (X,Y) 的概率密度为

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

(1) 求条件概率密度；(2) 计算 $P\{0.1<Y \leqslant 0.4 \mid X=1.5\}$.

解　(1) $f(x,y)$ 的非零区域如图 3-8 所示.

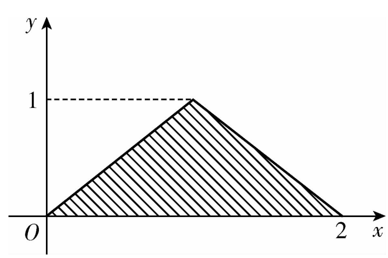


图 3-8