普通高等教育“十三五”规划教材
普 通 高 等 院 校 数 学 精 品 教 材

概率论与数理统计

（第三版）

主　编　刘次华
参　编　万建平　李楚进　刘继成
王湘君　胡吉卉　刘小茂
李　萍　胡晓山　叶　鹰
周晓阳　吴　娟

华中科技大学出版社
中国·武汉

目　　录

第一章　随机事件与概率

随机事件对概率论的研究具有重要意义，随机事件的等价表示是概率建模的重要技巧，公理化概率定义下涉及的概率计算公式就是针对随机事件及其运算而设计的.条件概率公式、全概率公式、贝叶斯公式在概率论中的科学地位可与微积分中的分部积分、变量代换媲美，它们构成了分析与应用概率的最常规、最有效工具.

1.1　随机试验与随机事件

掌握随机现象统计规律性的重要手段是重复观测，但在许多场合下这种观测是需要成本的.这些成本可能是时间或其他资源，即人们的这些重复观测是受到约束的.解决这些问题的一条出路就是通过精心设计的随机试验对这种观测进行模型化和简化.随机试验是人们敲开随机现象规律性大门的巧妙工具.例如，人们可通过反复投掷硬币观测正、反两面出现的统计规律来解释人类生育过程中男女性别之比的统计规律.掌握一些经典的随机试验案例及构造设计一些随机试验，对于研究概率论是很有意义的.随着科学研究向深度与广度的发展，我们鼓励读者在生活和工作中，总结经验，设计出一些新的随机试验来更有效地研究随机现象规律性.

1.1.1　随机试验

定义 1.1　设有试验 E，若 E 满足：

(1) 试验之前可知试验的一切可能结果，

(2) 每次试验之前不能确定此次试验的结果，

(3) 试验在相同条件下可以重复进行，

则称试验 E 为随机试验.

例 1.1　表 1.1 中数据记载了几位数学家抛硬币试验的结果.

表 1.1

实　验　者	抛硬币次数 n	出现正面次数 $n_{正}$	出现正面频率 $\frac{n_{正}}{n}$
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

观测上述数据可以发现，当 n 越来越大时，$\frac{n_{正}}{n}$ 有向 $\frac{1}{2}$ 集中的趋势.思考这本质上是一种什么样的趋势.

例 1.2(投针问题)　设平面上画着一族平行线，它们之间的间距相等且都等于 a，向此平

面任投一长度为 $l(l\leqslant a)$ 的针,观测在这种反复的投掷过程中针与平行线相交的次数. 表 1.2 给出了几位数学家实验数据的记录(设 $a=1$).

表 1.2

实 验 者	针 长	投 掷 次 数	相 交 次 数
Wolf	0.8	5000	2532
Smith	0.6	3204	1218.5
De Morgan. C	1.0	600	382.5
Fox	0.75	1030	1489
Lazzerini	0.88	3408	1808
Reina	0.5419	2520	859

有趣的是,通过对针与平行线相交概率的计算得到了一个关于 π 的统计估计方法,由此思路出发诞生了 Monte Carlo 随机模拟方法. Monte Carlo 随机模拟是现代统计计算的重要基础.

例 1.3(高尔顿板) 设板上钉有图 1.1 所示排列的钉子,自上端放入一小球,使其任意自由下落,下落过程中当小球碰到钉子后,它以向左边或向右边相等的机会落下,碰到下一排钉子时情形也是如此. 在底部设有如图所示的格子,进行大量试验后观测各个格子中落入的球的堆积情况. 这样的试验表明各个格子中球的堆积曲线在进行这样的反复试验中形状几乎是一样的. 读者可设想通过这种试验可解释哪些随机现象呢?

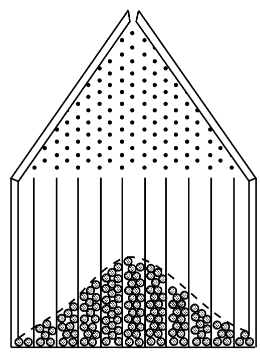

图 1.1

例 1.4 设想某人在平面上从零点出发,其规则是他手持分别标有 1、2、3、4 的一个均匀四面体,他随意抛出后,若标有 1 的面贴地,则他向东走一个单位长度,若标有 2 的面贴地,他向南走一个单位长度,标有 3 的面贴地对应他向西走一个单位长度,标有 4 的面贴地对应他向北走一个单位长度. 若干次后,观测他走过的路线的轨迹. 由此你有什么联想呢?

1.1.2 随机事件与样本空间

定义 1.2 随机试验的每一个可能结果称为一个随机事件,简称事件. 事件一般用 A、B、C 等表示. 不可能再分解的事件称为基本事件,由若干个基本事件组成的事件称为复合事件. 基本事件与复合事件的区分是相对的.

例 1.5 随意向桌面掷一颗骰子,观察出现的点数,一般情形下基本事件为 $\{1\}$,$\{2\}$,$\{3\}$,$\{4\}$,$\{5\}$,$\{6\}$;而出现的点数为偶数 $\{2,4,6\}$ 的事件则为复合事件. 若试验人员关心的仅仅是出现点数的奇偶性,则出现点数为偶数 $\{2,4,6\}$ 的事件的基本条件.

定义 1.3 随机试验 E 产生的所有可能的基本事件的集合称为样本空间,记为 Ω.

样本空间的任一子集即为一个事件,其中必然事件记为 Ω,不可能事件记为 $\varnothing$.

例 1.6 设随机试验 E 为记录两支排球队在一局比赛中某球队的净输球数,则 $\Omega=\{2,3,4,\cdots,25\}$.

例 1.7 设随机试验 E 为任取某地区某年的年降雨量记录,则 $\Omega=\{x:x\geqslant 0,x$ 的单位为 mL,表示该地区年降雨总量$\}$.

1.2　随机事件的关系、运算及其性质

一个随机试验所涉及的事件往往是非常丰富的. 正确地将一个随机试验所产生的所有事件表示出来是有意义的. 这种表示的基础是要建立事件之间的关系及引入一些有意义的事件的运算，为了更好地研究随机事件之间的关系及相关的运算，对这些运算所涉及的性质的研究也是有价值的.

1.2.1　事件的关系及其运算

定义 1.4　设 A,B 表示两个事件，若 A 发生必导致 B 发生，则称 A 包含于 B 或称 B 包含 A，记为 $A\subset B$ 或 $B\supset A$.

例 1.8　A 表示某国家某地区在某年发生了地震，B 表示该国家在该年发生了地震，则有 $A\subset B$.

定义 1.5　若 $A\subset B$ 和 $B\subset A$ 同时成立，则称 A 与 B 相等或称之为等价，记为 $A=B$.

将一个事件有目的地进行等价表示通常是概率论理论研究与实际应用的关键步骤.

定义 1.6　若 A,B 至少一个发生，则称之为 A 与 B 之和(并)，记为 $A\cup B$.

求事件和的运算可推广到可列无限多个事件的场合(一个集合称为可列无限是指该集合中的元素可与自然数集建立一一对应的关系). 设 $A_1,A_2,\cdots$ 为一列事件，则 $\bigcup\limits_{i=1}^{\infty}A_i$ 表示 A_1，$A_2,\cdots$ 至少有一个事件发生.

定义 1.7　若 A 与 B 同时发生，则称之为 A 与 B 之积(交)，记为 $A\cap B$，简记为 AB.

同理，$\bigcap\limits_{i=1}^{\infty}A_i$ 表示 $A_1,A_2,\cdots$ 同时发生.

定义 1.8　若 $AB=\varnothing$，称 A 与 B 互不相容或互斥. 若 $A\cap B=\varnothing$，$A\cup B=\Omega$，则称 A 与 B 互为逆事件，记 $B=\overline{A}$.

显然 A 与 B 互逆，则 A 与 B 必互不相容，但反之不然. 若 A 与 B 互斥，则 A 与 B 之和也记为 $A+B$.

定义 1.9　若 A 发生而 B 不发生，则称之为 A 与 B 之差，记为 $A-B$.

定义 1.10　$A\triangle B=(A-B)\cup(B-A)$ 称为 A 与 B 的对称差.

在建立了上述概念的基础上为了定义概率，有必要引入如下 σ 域的定义.

定义 1.11　设 $\mathscr{F}$ 是由 Ω 中的一些子集组成的集合，具有性质：

(1) $\Omega\in\mathscr{F}$,

(2) 若 $A\in\mathscr{F}$，则 $\overline{A}=\Omega-A\in\mathscr{F}$,

(3) 若 $A_n\in\mathscr{F},n=1,2,\cdots$，则 $\bigcup\limits_{n=1}^{\infty}A_n\in\mathscr{F}$,

则称 $\mathscr{F}$ 是 Ω 中的一个 σ 域(或称为 σ 代数).

定义 1.12　设 $A_n\subset A_{n+1},n=1,2,\cdots$，称事件序列 $\{A_n,n\geqslant 1\}$ 为单调递增序列；设 $A_n\supset A_{n+1},n=1,2,\cdots$，称事件序列 $\{A_n,n\geqslant1\}$ 为单调递减序列。当 $\{A_n,n\geqslant1\}$ 递增时，记 $\lim\limits_{n\to\infty}A_n=\bigcup\limits_{n=1}^{\infty}A_n$，称 $\bigcup\limits_{n=1}^{\infty}A_n$ 为此事件序列的极限. 同理，当 $\{A_n,n\geqslant1\}$ 递减时，记 $\lim\limits_{n\to\infty}A_n=\bigcap\limits_{n=1}^{\infty}A_n$ 为此事件

序列的极限,亦称$\bigcap\limits_{n=1}^{\infty}A_n$为此事件序列的极限.

我们关心的事件往往需通过事件的运算才能表达出来,但若干事件通过这些运算后是否满足某些封闭性呢?这个问题是很尖锐的.σ域的建立就在于防范这种情形的发生.可以直观地认为σ域就是这样一些事件构成的集合,这个集合内的事件关于Ω、关于事件的逆运算及关于事件的可列求和运算封闭.

1.2.2 事件的运算性质

上节介绍的事件的运算具有如下性质.

(1)事件和的运算满足

$$A\cup B=B\cup A\quad(\text{交换律}),\tag{1.1}$$

$$(A\cup B)\cup C=A\cup(B\cup C)\quad(\text{结合律}),\tag{1.2}$$

$$A\cup A=A,\quad A\cup\varnothing=A,\quad A\cup\Omega=\Omega.$$

(2)事件交的运算满足

$$A\cap B=B\cap A\quad(\text{交换律}),\tag{1.3}$$

$$(A\cap B)\cap C=A\cap(B\cap C)\quad(\text{结合律}),\tag{1.4}$$

$$A\cap A=A,\quad A\cap\varnothing=\varnothing,\quad A\cap\Omega=A.$$

(3)事件并与交的运算满足分配律

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)\quad(\text{第一分配律}),\tag{1.5}$$

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)\quad(\text{第二分配律}).\tag{1.6}$$

(4)德·摩根对偶律

$$\overline{A\cup B}=\overline{A}\cap\overline{B},\quad\overline{A\cap B}=\overline{A}\cup\overline{B},\tag{1.7}$$

$$\overline{\bigcup_{n=1}^{\infty}A_n}=\bigcap_{n=1}^{\infty}\overline{A_n},\quad\overline{\bigcap_{n=1}^{\infty}A_n}=\bigcup_{n=1}^{\infty}\overline{A_n}.\tag{1.8}$$

例 1.9 设有事件A_1,A_2,A_3,用事件的运算表示:

(1) $B_1=\{A_1,A_2,A_3$ 中至多发生 2 个$\}$;

(2) $B_2=\{A_1,A_2,A_3$ 中至少发生 2 个$\}$.

解 (1) $B_1=\overline{A}_1\overline{A}_2\overline{A}_3\cup A_1\overline{A}_2\overline{A}_3\cup\overline{A}_1A_2\overline{A}_3\cup\overline{A}_1\overline{A}_2A_3\cup A_1A_2\overline{A}_3\cup A_1\overline{A}_2A_3\cup\overline{A}_1A_2A_3$.

(2) $B_2=A_1A_2A_3\cup A_1A_2\overline{A}_3\cup A_1\overline{A}_2A_3\cup\overline{A}_1A_2A_3$.

例 1.10 设A,B,C表示三个事件,用A,B,C表示如下事件:

(1) A发生且B与C至少有一个发生;

(2) A与B发生而C不发生;

(3) A,B,C中恰有一个发生.

解 (1) $A(B\cup C)$.

(2) $AB\overline{C}$.

(3) $A\overline{B}\overline{C}\cup\overline{A}B\overline{C}\cup\overline{A}\overline{B}C$.

例 1.11 证明:若A,B为两事件,则$A\cup B=A\cup(B-A)$.

证明 $A\cup(B-A)=A\cup(B\overline{A})=(A\cup B)\cap(A\cup\overline{A})=A\cup B$.

例 1.12 把$A_1,A_2,\cdots,A_n$表示成n个互斥事件之和.

解 $A_1\cup(A_2-A_1)\cup(A_3-(A_1\cup A_2))\cup\cdots\cup(A_n-(A_1\cup\cdots\cup A_{n-1}))$.

例 1.13　化简事件$(A\cap B)\cup(A\cap\bar{B})$.

解　$AB\cup A\bar{B}=A(B\cup\bar{B})=A$.

例 1.14　设$(A\cup\bar{B})(\bar{A}\cup\bar{B})\cup(\overline{A\cup B})\cup(\overline{\bar{A}\cup B})=C$，求 B.

解　$B=\bar{C}$.

1.3　事件的概率及其计算

定义 1.13　设$\mathscr{F}$是样本空间 Ω 上的一个 σ 域，$P=P(\cdot)$是$\mathscr{F}$上定义的实函数，且 P 满足：

(1) $P(\Omega)=1$，

(2) $P(A)\geqslant 0$（对一切的 $A\in\mathscr{F}$），

(3) 若 $A_n\in\mathscr{F}$（$n=1,2,\cdots$），且两两互不相容，有 $P(\bigcup\limits_{n=1}^{\infty}A_n)=\sum\limits_{n=1}^{\infty}P(A_n)$ 成立，则称 P 是$\mathscr{F}$上的一个概率，$P(A)$称为事件 A 发生的概率.

通常将$(\Omega,\mathscr{F},P)$称为一个概率空间，性质(3)称为概率的 σ 可加性(可列可加性). 这是 1933 年柯尔莫哥洛夫(Kolmogoroo)建立的概率公理化定义，使得概率成为了一个严谨的数学分支，极大地推动了学科发展. 细心的读者会发现上述定义的基本作用在于判断 P 是否构成$\mathscr{F}$上的一个概率，但对于一个 $A\in\mathscr{F}$ 如何具体地求出 $P(A)$呢？为此我们需要从上述概率的公理化定义出发获得常用的一些概率计算的性质.

定理 1.1　概率具有如下性质：

(1) $P(\varnothing)=0$. $\quad(1.9)$

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

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i). \tag{1.10}$$

(3) 逆事件概率公式：

$$P(\bar{A})=1-P(A). \tag{1.11}$$

(4) 差事件概率公式：若 $B\subset A$，则

$$P(A-B)=P(A)-P(B). \tag{1.12}$$

(5) 概率的单调性：若 $B\subset A$，则

$$P(B)\leqslant P(A).$$

(6) 加法公式：设 A,B,C 为任意三个事件，则

$$P(A\cup B)=P(A)+P(B)-P(AB). \tag{1.13}$$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC). \tag{1.14}$$

对于任意 n 个事件 $A_1,A_2,\cdots,A_n$，下式成立：

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n). \tag{1.15}$$

(7) 设$\{A_n,n\geqslant 1\}$为递增或递减事件序列，则有

$$\lim_{n\to\infty}P(A_n)=P(\lim_{n\to\infty}A_n).$$

此性质表明在前述事件序列极限的定义下，概率是上、下连续的.

证明 (1) 由 $\Omega=\Omega\cup\varnothing\cup\varnothing\cup\cdots$及概率的非负性及 $P(\Omega)=1$(规范性)知$P(\varnothing)=0$.

(2) 只需令 $B_1=A_1,\cdots,B_n=A_n,B_{n+1}=\varnothing,B_{n+2}=\varnothing,\cdots$,可知 $B_1,B_2,\cdots,B_n,B_{n+1},\cdots$两两互不相容,且有

$$\bigcup_{i=1}^{n}A_i=\bigcup_{i=1}^{\infty}B_i,$$

从而有

$$\begin{aligned}P\left(\bigcup_{i=1}^{n}A_i\right)&=P\left(\bigcup_{i=1}^{\infty}B_i\right)=P(B_i)+\cdots+P(B_n)+P(B_{n+1})+\cdots\\&=P(B_1)+\cdots+P(B_n)=P(A_1)+\cdots+P(A_n).\end{aligned}$$

(3) 由 $A\cup\overline{A}=\Omega$ 及有限可加性,有

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

故性质(3)成立.

(4) 当 $B\subset A$ 时,有 $A=(A-B)\cup B$,又 $A-B$ 与 B 互不相容,从而由有限可加性即知性质(4)成立.

(5) 由性质(4)知当 $B\subset A$ 时,有 $P(A-B)=P(A)-P(B)$,又 $P(A-B)\geqslant 0$,从而性质(5)成立.

(6) 下面仅就 $P(A\cup B)=P(A)+P(B)-P(AB)$给出证明.

由于 $A\cup B=A\cup(B-AB)$,$AB\subset B$,A 与 $B-AB$ 互不相容,从而

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

(7) 证明略.

下面就两个常见的概型讨论概率公理化定义的性质是否满足的问题.

定义 1.14 古典概型:设随机试验 E 只产生有限个基本事件(也称样本点),此时样本空间中的样本点总数有限,并设每次试验中各个基本事件出现的可能性是相同的.若 A 是由 m 个基本事件组成的事件,则 A 的概率定义为

$$P(A)=\frac{A\text{ 中所含样本点数}}{\Omega\text{ 中样本点总数}}.\tag{1.16}$$

读者在利用此公式计算 A 的概率时一定要验证 Ω 中样本点总数有限及每一个样本点出现机会均等的条件.

由于 $P(A)$的定义式中分子、分母所涉及的数均非负,故对任意的 $A\in\Omega$,有$P(A)\geqslant 0$,$P(\Omega)=1$ 是显然的.由于 Ω 中的样本点总数有限,故由 Ω 的样本点构成的所有子集的个数是有限的,从而概率公理化定义中的 σ 可加性此时对应的为特殊情形,即有限可加性.事实上,设 $A=\bigcup_{i=1}^{n}A_i$,$A_i(i=1,2,\cdots,n)$两两不相容,从而 A 中的样本点数为 $A_1,A_2,\cdots,A_n$ 中样本点数之和,所以

$$P(A)=P(A_1\cup A_2\cup\cdots\cup A_n)=\frac{A_1,A_2,\cdots,A_n\text{ 中样本点数之和}}{\Omega\text{ 中样本点总数}}=\sum_{i=1}^{n}P(A_i).$$

古典概率的计算涉及如下两条原理.

定义 1.15 加法原理:设事件 A 有 n 类方法出现,并设第 i 类方法由 m_i 种方式组成,则 A 的出现方式共有 $m_1+m_2+\cdots+m_n$ 种.

定义 1.16 乘法原理:若事件 A 有 m 种不同方式出现,设另有事件 B 对 A 的每一种出现方式有 n 种出现方式对应,那么 AB 就应以 nm 种不同方式出现.

加法原理体现了并行的理念,乘法原理体现分步机理.

此外,古典概型往往涉及排列组合问题,笔者认为这些知识读者已经熟悉,故不作介绍.

例 1.15　(1) $A=\{$一批产品共 N 件,其中有 M 件次品,从中任取一件,这件产品恰为次品$\}$,求 $P(A)$.

(2) $B=\{$一批产品共 N 件,其中 M 件次品,从中任取 n 件,这 n 件中恰有 l 件次品$\}$,求 $P(B)$.

(3) $C=\{$一批产品共 N 件,分成 $1,2,\cdots,k$ 个等级,第 i 个等级中有 M_i 件产品,$i=1,2,\cdots,k$,$M_1+M_2+\cdots+M_k=N$. 从中任取 n 件,这 n 件中恰有第 i 个等级的产品 l_i 件,$i=1,2,\cdots,k\}$,求 $P(C)$.

解　显然此时符合古典概型条件.

(1) $P(A)=\dfrac{M}{N}$.

(2) 由乘法原理得

$$P(B)=\frac{C_M^l C_{N-M}^{n-l}}{C_N^n}. \tag{1.17}$$

(3) 由乘法原理得

$$P(C)=\frac{C_{M_1}^{l_1} C_{M_2}^{l_2}\cdots C_{M_k}^{l_k}}{C_N^n}.$$

这里涉及的概率可抽象出一个称之为超几何分布的概率模型,在第二章中将进行介绍.

定义 1.17　几何概型:设样本空间中样本点的集合与平面(或一维、三维空间)某区域 G 一一对应,即 Ω 可认为是 G,并设 Ω 中的样本点出现的机会相等,设 $A\subset\Omega$,对应地有区域 $A_0\subset G$,不妨设 $A_0=A$,称

$$P(A)=\frac{A\text{ 的面积(体积)}}{G\text{ 的面积(体积)}} \tag{1.18}$$

为事件 A 的几何概型下定义的几何概率,简称 $P(A)$ 为 A 的概率.

几何概型可认为是古典概型的推广,一方面几何概型将古典概型中的样本空间中的点数从有限多个推广到无穷多个,另一方面古典概型中的点可看成是一些孤立的点,几何概型中 Ω 中的点为充满了一个平面(或空间)区域中的点. 几何概型的计算主要涉及将所关心的概率计算问题,在满足几何概型的条件下,转化为对线段长度或平面区域面积或空间区域体积的度量问题. 解题的关键是能根据问题的本质,作出正确的相应的几何图形,再进行度量.

例 1.16　在一张画了小方格的纸上随机地投一枚直径为1 cm的圆片,方格为多大时才能使圆片与线不相交的概率小于 1%?(设方格边长为 a(cm).)

解　方格边长为 a (cm),当圆片圆心落入图 1.2 中阴影部分时才与边界不相交. 由几何概型有

$$P(\text{圆片不与线相交})=\frac{\text{阴影部分面积}}{\text{方格面积}}=\frac{(a-1)^2}{a^2}.$$

令 $\dfrac{(a-1)^2}{a^2}<0.01$,当 $a\leqslant 1$ (cm)时,圆片必与线相交,只需考虑 $a>1$,故

$$\frac{a-1}{a}<0.1,\quad a<\frac{10}{9},$$

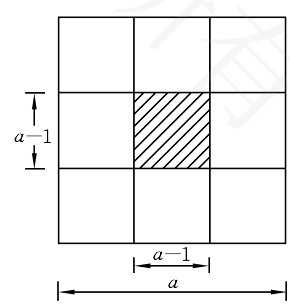

图 1.2

所以 $1<a<\frac{10}{9}$时可达到要求.

读者可自行验证几何概型满足概率公理化定义中的全部条件.

1.4　条件概率事件独立性

回顾以上概率的计算,其基本方法是利用概率的运算性质计算概率,当问题符合古典概型和几何概型条件时,可利用这两个模型计算概率.概率计算性质的实质是讨论概率计算与事件和、差等运算交换后所满足的性质.下面我们要从另外的途径发掘事件概率计算的新方法,其方向是从事件发生之间的影响关系、从事件局部到全局的关系出发,以此获得概率计算新方法的突破口.

定义 1.18　设有随机试验 E 及事件 A,B,若 $P(B)>0$,则称

$$P(A|B)=\frac{P(AB)}{P(B)} \tag{1.19}$$

为 A 在 B 发生的条件下的条件概率.注意到概率为 0 的事件也有可能发生,读者可思考如何刻画 $P(B)=0$ 时的条件概率 $P(A|B)$.

同理,可定义

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

可以验证这样定义的条件概率满足概率公理化定义中关于概率的三个条件.条件概率引进的角度是基于事件发生之间的影响关系构造的.该定义同时也给出了条件概率的计算方法,事实上该定义式可启发人们从两个方面求 $P(A|B)$.一方面从基于 A,B 原有的样本空间出发,通过计算 $P(AB),P(B)$,再由$\frac{P(AB)}{P(B)}$求出 $P(A|B)$.另一方面就 A,B 而言,可视 B 为一个缩小的样本空间,记为 Ω_B.在此样本空间上,只要求出 $P(A)$即为原来定义下的 $P(A|B)$.

考虑由

$$P(A|B)=\frac{P(AB)}{P(B)}\quad 及\quad P(B|A)=\frac{P(AB)}{P(A)},$$

可得乘法公式

$$P(AB)=P(A|B)P(B)=P(B|A)P(A). \tag{1.20}$$

这个公式可推广到 n 个事件 $A_1,A_2,\cdots,A_n$ 的场合:

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1}). \tag{1.21}$$

事实上条件概率可认为是一种相对概率,在 $P(A|B)$中当 $B=\Omega$ 时,$P(A|B)=P(A)$,故通常事件的概率可看成是条件概率的特例.

例 1.17　设加工产品 20 件,其中有 15 件一等品、5 件二等品,一等品、二等品混放.现随机取两件,每次取一件,共两次,不放回抽样,求在第一次取到一等品的条件下第二次仍取到一等品的概率.

解　$A=\{$第一次取到一等品$\}$,$B=\{$第二次取到一等品$\}$,在原样本空间下

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{\frac{15\times14}{20\times19}}{\frac{15}{20}}=\frac{14}{19}.$$

在缩减的样本空间下计算,得

$$P(B|A)=\frac{14}{19}.$$

定义 1.19　设 Ω 为样本空间,若存在 $A_i\in\Omega$,且 $A_iA_j=\varnothing(i\neq j;i,j=1,2,\cdots,n)$ 及 $\bigcup_{i=1}^{n}A_i=\Omega$,则称 $A_1,A_2,\cdots,A_n$ 构成对 Ω 的一个划分.

定理 1.2　全概率公式:设 $A_1,A_2,\cdots,A_n$ 为 Ω 的一个划分且 $P(A_i)>0,i=1,2,\cdots,n$,则对于任意的 $B\subset\Omega$,有

$$P(B)=\sum_{i=1}^{n}P(B\mid A_i)P(A_i).\tag{1.22}$$

全概率公式的意义在于当 B 的构成较复杂时,可通过对 Ω 引入某种合适的划分,在这种划分下可将 B 分解为一系列局部的较为简单的事件,在计算出这些局部的事件的概率后,再通过全概率公式求出 $P(B)$.当然,若有 $B\subset A$,又找到 A 的有效划分 $A=\sum_{i=1}^{n}A_i$,则仍成立上述全概率公式.

例 1.18　设有一矿工被困井下,他可以等可能性地选择三个通道之一逃生.若矿工通过第一个通道逃生成功的可能性为$\frac{2}{3}$,通过第二个通道逃生成功的可能性为$\frac{4}{5}$,通过第三个通道逃生成功的可能性为$\frac{1}{6}$.问此矿工逃生成功的可能性是多少?

解　设 $A=\{$矿工逃生成功$\}$;$B_i=\{$矿工选择第 i 个逃生通道$\}$,$i=1,2,3$,则

$$P(A)=\sum_{i=1}^{3}P(A\mid B_i)P(B_i)=\frac{2}{3}\times\frac{1}{3}+\frac{4}{5}\times\frac{1}{3}+\frac{1}{6}\times\frac{1}{3}=\frac{49}{90}.$$

人们通常见到诸如这样的广告:购买某产品,它将会给你带来某效用;或听到这样的承诺:投资某产品或投资某地,你的回报率将是多少.但如果人们问一个反问题:我要求回报率达到多少应投资何产品或在何地投资呢?或者我要求达到某效用应购买何产品呢?显然这样的反问题是很有意义的.下面介绍的贝叶斯公式就是这类问题在概率中的相应模型.

定理 1.3　贝叶斯公式:设 $A_1,A_2,\cdots,A_n$ 为样本空间的一个划分,且 $P(A_i)>0,i=1,2,\cdots,n$,对任意随机事件 $B\subset\Omega$,当 $P(B)>0$ 时,则有

$$P(A_i\mid B)=\frac{P(B\mid A_i)P(A_i)}{\sum_{k=1}^{n}P(B\mid A_k)P(A_k)}.\tag{1.23}$$

一般称 $P(A_i)$为先验概率,$P(A_i|B)$为后验概率.

例 1.19　设某医生面临某患者,该患者呈状态 N.医生试图据此判断该患者是否患疾病 C.医生有如下经验认识:$P(N|C)=0.8$,人群中患疾病 C 的概率 $P(C)=0.005$.医生所接触的患者中有状况 N 的概率 $P(N)=0.1$.现要求 $P(C|N)$.

解　$P(C|N)=\frac{P(N|C)P(C)}{P(N)}=\frac{0.8\times0.005}{0.1}=0.04.$

细心的读者可看出 $P(C|N)=0.04$ 比 $P(C)=0.005$ 大 7 倍.这又从概率方面给人们认识某些事件以何启示呢?

例 1.20　在例 1.18 中,求 $P(B_2|A)$.

解 $P(B_2 \mid A) = \dfrac{P(A \mid B_2)P(B_2)}{\sum\limits_{i=1}^{n} P(A \mid B_i)P(B_i)} = \dfrac{\frac{4}{5} \times \frac{1}{3}}{\frac{49}{90}} = \dfrac{24}{49}$.

条件概率揭示了事件发生之间的关系在概率计算方面的影响,这种关系对于概率模型及实际问题的分析研究是很有意义的.但有些情形,随机事件的发生相互之间没有影响关系或者影响关系很弱也是有价值的.例如,某时刻在甲国某地区发生了一次汽车交通事故,这个事故对乙国的火车交通的安全性的影响可能就很弱.再如某国在 1000 年前的某月某日在某地区下了一场雨,那么这场雨对 1000 年后该地区的某月某日的气象不可能造成什么影响.这样的问题我们引入事件独立性的概念来刻画.

定义 1.20 设事件 A,B 满足

$$P(AB)=P(A)P(B), \tag{1.24}$$

则称 A 与 B 独立.

由定义可知,当 A 与 B 独立时,B 与 A 也独立,故称之为 A 与 B 相互独立,简称 A 与 B 独立.读者可能马上想象,若 $P(ABC)=P(A)P(B)P(C)$,那么 A,B,C 也应该相互独立,但这是不全面的,不准确的.事实上,A,B,C 相互独立应同时满足

$$P(ABC)=P(A)P(B)P(C),$$
$$P(AB)=P(A)P(B), \quad P(AC)=P(A)P(C),$$
$$P(BC)=P(B)P(C).$$

定义 1.21 对于 n 个事件 $A_1,A_2,\cdots,A_n$,若同时满足

$$P(A_{i_1}A_{i_2}\cdots A_{i_k})=P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}) \ (2\leqslant k\leqslant n), \tag{1.25}$$

则称 $A_1,A_2,\cdots,A_n$ 相互独立.

注意到 n 个事件独立的刻画很繁琐,有什么简洁有效的方法刻画呢?进而读者还可以思考如何使两个随机试验独立?

另外,由独立性的概念可得如下定理成立.

定理 1.4 设 $P(B)>0$,则 A 与 B 相互独立的充要条件为 $P(A|B)=P(A)$.

利用事件的独立性概念可知,当事件独立时可大大简化积事件概率的计算.例如,我们要求 $P(\bigcup\limits_{i=1}^{n} A_i)$ 且 n 较大时,直接计算是困难的.若 $A_1,A_2,\cdots,A_n$ 相互独立,则 $P(A_1\cup A_2\cup\cdots\cup A_n)=1-P(\overline{A_1\cup A_2\cup\cdots\cup A_n})=1-P(\bar{A}_1\cap\bar{A}_2\cap\cdots\cap\bar{A}_n)=1-P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n)$.其实在使用该方法时需如下定理:

定理 1.5 $\{A,B\},\{A,\bar{B}\},\{\bar{A},B\},\{\bar{A},\bar{B}\}$ 四对事件中若有一对相互独立,则其余三对相互独立.

本定理的证明对读者来说是不难实现的.

例 1.21 一个家庭中有若干小孩,假设生男和生女等可能,令

$A=\{$一个家庭中有男孩,又有女孩$\}$,

$B=\{$一个家庭中最多有一个女孩$\}$,

在家中分别有两个或三个小孩时,讨论 A 与 B 的独立性。

解 1°家庭中有两个小孩:$\Omega=\{MM,MF,FM,FF\}$,$A=\{FM,MF\}$,$B=\{MM,MF,FM\}$,$AB=\{FM,MF\}$,$P(A)=\dfrac{1}{2}$,$P(B)=\dfrac{3}{4}$,$P(AB)=\dfrac{1}{2}$,$P(AB)\neq P(A)P(B)$.A 与 B 不

独立.

2°家庭中有三个小孩：$\Omega=\{MMM,MMF,MFM,FMM,MFF,FMF,FFM,FFF\}$，$\overline{A}=\{MMM,FFF\}$，$B=\{MMM,MMF,MFM,FMM\}$，$AB=\{MMF,MFM,FMM\}$，$P(A)=\dfrac{3}{4}$，$P(B)=\dfrac{1}{2}$，$P(AB)=\dfrac{3}{8}$，$P(AB)=P(A)P(B)$. A 与 B 独立.

习　题　一

1.1　建立如下问题的样本空间：

(1) 某支股票两期的价格波动.设每期上涨记为 u，下跌记为 d，每期价格要么涨要么跌；

(2) 观测某保险公司某险种的索赔次数；

(3) 随机抽查某河流在某年某月内的月流量；

(4) 2007 年全年世界范围内动物种类灭绝数.

1.2　填空题.

(1) 随机试验的要求为______.

(2) A,B 互不相容是指______.

(3) 事件的和、积、差、逆与集合运算的并、交、差、补的对应关系为______.

(4) 列举某三个事件运算的性质______.

1.3　连续进行三次射击，设 $A_i=\{$第 i 次射击命中目标$\}(i=1,2,3)$，$B_j=\{$三次射击恰好命中 j 次$\}(j=0,1,2,3)$，$C_k=\{$三次射击至少命中 k 次$\}$ $(k=0,1,2,3)$.

(1) 通过 A_1,A_2,A_3 表示 B_j 和 $C_k(j,k=0,1,2,3)$；

(2) 通过 B_j 表示 $C_k(j,k=0,1,2,3)$.

1.4　设随机事件 A,B 及和事件 $A\cup B$ 的概率分别为 0.4，0.3，0.6，求 $P(A\overline{B})$.

1.5　填空题.

(1) 记 $P(AB)=P(\overline{A}\,\overline{B})$，$P(A)=p$，$P(B)=$______.

(2) $AB=\varnothing$，$P(A)=p$，$P(B)=q$，$P(A\cup B)=$______.

　　$P(\overline{A}\cup B)=$______，$P(\overline{A}B)=$______，$P(\overline{A}\,\overline{B})=$______.

(3) $P(A-B)=P(A)-P(B)$成立的充要条件是______.

(4) 若 $P(AB)=P(A)P(B)$，且 $P(B)=0.5$，$P(A-B)=0.3$，则 $P(B-A)=$______.

1.6　求 n 阶行列式的展开式中任取一项，这项至少含一个主对角线元素的概率.

1.7　利用概率论想法证明下列恒等式：

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots+\frac{(A-a)\cdot\cdots\cdot 2\cdot 1}{(A-1)\cdot\cdots\cdot(a+1)a}=\frac{A}{a},$$

其中，A,a 为正整数，$A>a$.

1.8　设有函数 $f(x_1,x_2)=x_1^2+x_2^2$，s. t. $x_1+x_2=1$，在平面上任取一点(x_1^α,x_2^α)，使得 $x_1^\alpha+x_2^\alpha=1$，且 $f(x_1^\alpha,x_2^\alpha)=\min\limits_{(x_1,x_2)\in\mathbf{R}^2}\{f(x_1,x_2)\}$，求此事件的概率.

1.9　在平面上画一些平行线，它们之间的距离都为 a，向此平面随意投一长度为 l $(l<a)$的针，求此针与任一平行线相交的概率.

1.10　在长为 k 的线段上取任两点，从而将此线段分为三截，问这三截线段构成一个三角形的概率是多少？

1.11　设 A,B 是两个随机事件，$0<P(A)<1$，$P(B)>0$，$P(B|A)=P(B|\overline{A})$，则必有(　　).

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

(C) $P(AB)=P(A)P(B)$　　(D) $P(AB)\neq P(A)P(B)$

1.12　设 A,B 为随机事件,且 $P(B)>0$,$P(A|B)=1$,则必有(　　).

(A) $P(A\cup B)>P(A)$　　(B) $P(A\cup B)>P(B)$

(C) $P(A\cup B)=P(A)$　　(D) $P(A\cup B)=P(B)$

1.13　某人向同一目标重复射击,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) $6p^2(1-p)^2$

1.14　袋中有 50 个乒乓球,其中 20 个黄球、30 个白球.两人依次随机地从袋中各取一球,取后不放回,则第二人取得黄球的概率是多少?

1.15　设有车间 A,B,次品率分别为 1%,2%,现从由 A 和 B 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,问此次品为 A 生产的概率是多少?

1.16　甲、乙两人独立对同一目标射击一次,其命中率分别为 0.6 和 0.5,现已知目标被命中,则它是由甲命中的概率为多少?

1.17　设袋中有 M 个球,其中有 N 个白球,从中任意取 n 个球,一次取一球,不放回抽样,A_i 表示第 i 次取到白球这一事件,B_k 表示在取出的 n 个球中恰好有 k 个白球这一事件,求 $P(A_i|B_k)$.

1.18　设 $A_1,A_2,\cdots,A_n$ 相互独立,证明:$P\left(\bigcup\limits_{k=1}^{\infty}A_k\right)=1-\prod\limits_{k=1}^{n}P(\bar{A}_k)$.

1.19　设有一个 n 阶行列式,A 表示"任取这个行列式的展开式中的一项,其中包含了主对角线上的元素"这一事件,问当 n 很大后 $P(A)=P(\bar{A})$ 是否成立?

1.20　构造例子,使得:

(1) $P(A|B)<P(A)$;　　(2) $P(A|B)=P(A)$;　　(3) $P(A|B)>P(A)$.

第二章 随机变量及其分布

在本章，我们引入随机变量的概念. 随机变量建立了样本空间与实数的联系，用它的取值范围可以表示所有感兴趣的随机事件，从而将计算事件的概率转化为求它的分布函数. 随机变量的概念是现代概率论的核心，在此基础上建立的概率论，其内容更加丰富，与其他数学分支的联系更加紧密. 随机变量可以分为离散型随机变量、连续型随机变量、其他类型的随机变量. 我们着重介绍前两类，适当涉及离散型和连续型的混合型随机变量.

2.1 随机变量及其分布函数

在随机现象中，随机试验的结果经常是用数值来描述的，例如，掷一个骰子所得的点数，统计某时间段登陆某网站的人数，预报明天天气的最高气温，等等. 有的随机试验结果虽然不是数值，但可以通过定义函数建立它们与数值之间的联系. 例如，考虑抛一枚硬币的随机试验，样本空间 $\Omega=\{$正面,反面$\}$，如果定义

$$X(\omega)=\begin{cases}1, & \omega=\text{正面},\\ 0, & \omega=\text{反面},\end{cases}$$

则可以用 Ω 上有两个取值的函数 $X(\omega)$ 表示抛一枚硬币的结果，即

$$\{\omega: X(\omega)=1\}=\{\text{正面}\}, \quad \{\omega: X(\omega)=0\}=\{\text{反面}\}.$$

X 取值为 1 或者 0 要视抛硬币的结果是正面还是反面来确定. 因而，在抛硬币的结果明确之前，函数 X 可能取值为 1，也可能取值为 0.

也经常有下面的情形出现. 尽管随机试验的结果是用数值描述的，但我们关心的不是试验结果本身，而是试验结果的某个函数. 例如，掷两个骰子，通常感兴趣的是它们的点数之和，而不管这两个骰子各是几点. 比如，试验结果(1,4)，(2,3)，(3,2)，(4,1)对应的点数之和都等于5，其他结果也可以根据点数之和类似地定义. 这个点数之和就是试验结果的函数.

这两类问题我们都定义了样本空间上的函数，我们称它为随机变量. 设$(\Omega,\mathscr{F},P)$为一概率空间. 从上述例子知道，随机变量的定义与概率 P 无关. 然而，通常概率 P 不一定对所有 Ω 的子集都有定义，而只是对 $\mathscr{F}$ 中的集合才有定义. 为使随机变量的取值范围所表示的集合总有概率可言，就需要对随机变量附加要求，即对任意实数 x，$\{\omega: X(\omega)\leqslant x\}\in\mathscr{F}$. 由此，我们引入如下定义.

定义 2.1 设 E 为随机试验，Ω 为其样本空间，$\mathscr{F}$ 为 Ω 上的 σ 域，称 Ω 上满足：对任意实数 x，$\{\omega: X(\omega)\leqslant x\}\in\mathscr{F}$ 的实值函数 X 为**随机变量**.

注意到，随机变量的随机性只是来源于随机试验结果的不确定性. 也就是说，在试验结果揭晓之前我们不知道随机变量的取值，因此是不能完全预言的，是随机的. 但随机变量本身的确就是通常的映射，样本空间与实数的对应关系是确定的. 今后时常省略 ω，将 $X(\omega)$，$\{\omega: X(\omega)\leqslant x\}$，$P(\{\omega: X(\omega)\leqslant x\})$等分别简记为 X，$\{X\leqslant x\}$，$P(X\leqslant x)$等.

随机变量的引入是一件方便的事情.我们可以用随机变量的取值范围表示事件.例如,设 X 为随机变量,$a<b$ 为实数,则$\{X\leqslant a\}$,$\{a<X\leqslant b\}$,$\{X=a\}$,…都表示事件.自然地,能用随机变量的取值范围表示事件后,我们更想知道这些事件的概率.为此,我们引入如下定义.

定义 2.2　设 X 为随机变量,对任意实数 x,定义 $F_X(x)=P(X\leqslant x)$.称函数$F_X(x)$为随机变量 X 的**分布函数**.

改写分布函数为 $F_X(x)=P(\omega:X(\omega)\in(\infty,x])$.如图 2.1 所示.如果将 X 看做数轴上随机点的坐标,$F_X(x)$可以直观理解为随机点 X 落在区间$(-\infty,x]$内的概率.如果不会引起混淆,常常将 $F_X(x)$简记为 $F(x)$.如下定理给出了分布函数的一般性质.

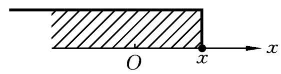

图 2.1

定理 2.1　设 $F(x)$为 X 的分布函数,则 $F(x)$具有性质:

(1) 单调非降性:对任意的 $x_1<x_2$,$F(x_1)\leqslant F(x_2)$;

(2) 右连续性:对任意的 a,$\lim\limits_{x\downarrow a}F(x)=F(a)$;

(3) 规范性:$F(-\infty)=0$,$F(+\infty)=1$,其中

$$F(-\infty)=\lim_{x\to-\infty}F(x),\quad F(+\infty)=\lim_{x\to+\infty}F(x).$$

证明　(1) 当 $x_1<x_2$ 时有$\{X\leqslant x_1\}\subset\{X\leqslant x_2\}$,由概率的单调性,得 $F(x_1)\leqslant F(x_2)$.

(2) 对任意的 $x_n\downarrow a$,有

$$\bigcap_n\{X\leqslant x_n\}=\{X\leqslant a\},$$

由概率的上连续性,有$\lim\limits_{x_n\downarrow a}F(x_n)=F(a)$.

(3) 与(2)的证明类似.因为

$$\bigcap_n\{X\leqslant -n\}=\varnothing,\quad \bigcup_n\{X\leqslant n\}=\Omega,$$

故分别由概率的上、下连续性得证.

由随机变量的分布函数,我们能计算很多事件的概率.例如:

(1) $P(X\leqslant b)=F(b)$;

(2) $P(X>b)=1-P(X\leqslant b)=1-F(b)$;

(3) $P(a<X\leqslant b)=F(b)-F(a)$;

(4) $P(X<b)=F(b^-)$,其中 $F(b^-)=\lim\limits_{x\uparrow b}F(x)$;

(5) $P(X=b)=F(b)-F(b^-)$.

其中,(4)是由于

$$\{X<b\}=\bigcup_n\left\{X\leqslant b-\frac{1}{n}\right\},$$

然后利用概率的下连续性得到,其中极限的存在性是因为 $F(x)$为单调非降的.其他等式都是显然的.

例 2.1　考虑某样本空间 Ω 上的随机事件 A,定义随机变量

$$I_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A,\end{cases}$$

则当且仅当 A 发生时 $I_A=1$,否则 $I_A=0$.I_A 称为 A 的**示性函数**.它可以看做是随机事件 A 所对应的随机变量.因此,随机变量可以看做事件的自然推广.显然,

$$\{I_A \leqslant x\}=\begin{cases}\varnothing, & x<0,\\ \overline{A}, & 0\leqslant x<1,\\ \Omega, & x\geqslant 1.\end{cases}$$

若 $P(A)=p$,则有 $P(I_A=1)=p$,$P(I_A=0)=1-p$. I_A 的分布函数为

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

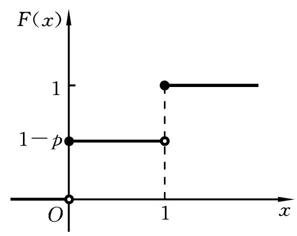

图 2.2

图 2.2 所示是一个阶梯函数,在 0,1 处依次有跃度为 $1-p$ 和 p 的跳跃.

反过来,根据分布函数 $F(x)$,由(5)容易知道 $P(I_A=1)=p$,$P(I_A=0)=1-p$. 对任意的 $x\neq 0$ 或 1,$P(I_A=x)=0$. 再结合 I_A 的定义知,$P(A)=p$.

例 2.2　网络电视播放软件 PPTV 在剧场频道每天循环播放电影. 据悉某天将放映电影《速 8》,片长共 150 分钟. 晚饭后,你随意打开电脑观看,问电影刚开始从头放映不超过 15 分钟的概率?

这是一个一维几何型概率的问题. 以分钟(min)为时间单位,可取 $\Omega=[0,150]$,则 $A=\{$电影刚开始从头放映不超过 15 分钟$\}=[0,15]$,则

$$P(A)=\frac{|A|}{|\Omega|}=\frac{1}{10}.$$

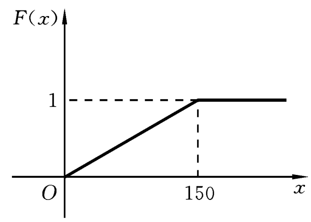

图 2.3

因为该试验的结果本身就是数值,不妨直接定义 $X(\omega)=\omega$,$\omega\in[0,150]$. 则 X 表示打开电脑观看的时间,$A=\{X\leqslant 15\}$. 由几何型概率的定义,可计算 X 的分布函数(见图 2.3)为

$$F(x)=\begin{cases}0, & x<0,\\ \dfrac{x}{150}, & 0\leqslant x<150,\\ 1, & x\geqslant 150.\end{cases}$$

因此,$P(A)=F(15)=\dfrac{1}{10}$. 注意到,$F(x)$是一个连续函数,所以对所有的实数 x,$P(X=x)=0$. 这表明,概率为 0 的事件未必是不可能事件.

显然,对任意的 $x\neq y$,$\{X=x\}$与$\{X=y\}$互不相容. $P(\bigcup_{x\in[0,150]}\{X=x\})=P(\Omega)=1$. 这说明区间$[0,150]$上有不可列多个值,而且概率公理化定义中可列可加性的假设是本质的.

可以证明,由分布函数可以决定所有用随机变量取值范围所表示的事件的概率. 因此,有关随机变量的概率问题都可以借助它的分布函数来回答. 这样,我们研究概率的问题就转化为对随机变量的研究,有关随机变量的概率问题可由分布函数来刻画. 同时也可以证明,满足定理 2.1 中三个性质的函数一定为某个随机变量的分布函数. 因此,原则上我们能够把概率论限制在对由随机变量所确定的分布函数的研究上. 这样就可避免涉及抽象的概率空间,也避免"随机试验"、"试验的结果"等术语. 把概率论简化为随机变量的理论使得分析的知识可以派上用场,而且也简化了理论的诸多方面. 然而,美中不足的是,这使概率论的背景晦涩不明,不能体现概率论生动而直观的一面.

基于此,理论上我们常常通过分布函数来引入随机变量. 但应注意到,分布函数只是实直线上的函数,所以,同一个分布函数可以联系不同的概率空间,同一概率空间上不同的随机变

量也可以有相同的分布函数.因此,通常用分布函数所引入的只是具有相同分布函数的一类随机变量.如果还需要关心相应的概率空间,以及其上的随机变量,就要视具体的应用问题而确定.

根据随机变量的取值情况,将取有限个或者可列无限个值的随机变量称为**离散型随机变量**,其余的称为**非离散型随机变量**.在非离散型随机变量中,有一类重要的特殊类型的随机变量,称为**连续型随机变量**.例 2.1 中的随机变量为离散型的,例2.2中的随机变量是连续型的.事实上,由上面解释的随机变量与其分布函数的对应关系可以知道,不同类型随机变量对应着不同类型的分布函数,相反,不同类型的分布函数自然对应不同类型的随机变量.亦即离散型随机变量与阶梯型的分布函数相对应,连续型随机变量与绝对连续型的分布函数相对应,其他的分布函数则对应着其他类型的随机变量.下面的两节我们将分别着重讨论离散型和连续型这两种最重要的特殊类型的随机变量,在 2.3 节也将提及离散型和连续型的混合型随机变量.

2.2 离散型随机变量

2.2.1 离散型随机变量及其分布列

若随机变量 X 仅取有限个值或者可列无限个值,则称 X 为**离散型随机变量**.

设 X 是离散型随机变量,不妨假定 X 可能的取值为 $x_1, x_2, \cdots$,X 取这些值的概率为

$$p_i = P(X = x_i), \quad i = 1, 2, \cdots. \tag{2.1}$$

称(2.1)式为随机变量 X 的**概率分布**,简称**分布列**.它包含两个方面的内容:X 可能取什么值与取这些值的概率.

X 的分布列也可以用如下的表格来表示:

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

显然,分布列有以下性质:

(1) 非负性:$p_i \geqslant 0, i = 1, 2, \cdots$;

(2) 规范性:$\sum_i p_i = 1$.

若已知随机变量 X 的分布列 p_i,则它的分布函数为

$$F(x) = P(X \leqslant x) = \sum_{x_i \leqslant x} P(X = x_i) = \sum_{x_i \leqslant x} p_i.$$

相反地,若 X 的分布函数为 $F(x)$,则它的分布列为

$$p_i = P(X = x_i) = F(x_i) - F(x_i^-), \quad i = 1, 2, \cdots.$$

亦即离散型随机变量的分布列与分布函数相互决定.因此,如分布函数一样,利用分布列可以计算 X 所有取值的概率,满足上述两个性质的数列也必然为某个离散型随机变量的分布列.因为分布列比分布函数更直观、更简单,所以我们常常不用分布函数而是用分布列来刻画离散型随机变量.而且,通常在定义离散型随机变量时,不具体明确随机试验、样本空间等概念,只给出满足上面两个性质的分布列.同样,与此分布列相联系的样本空间、随机变量总是存在的,它们依赖于具体的随机试验模型.

由于我们常常关心的是与随机变量相联系的概率,所以仅考虑总概率为 1 的那些取值就

足够了.因此,离散型随机变量的概念可以自然地拓广到以概率1取至多可列个值的随机变量.另外,对于离散型随机变量也往往忽略概率为0的取值.

2.2.2　常见的离散型分布

离散型随机变量总是根据其分布列来归类的.下面介绍几种常见的离散型随机变量的分布列,它们不仅在应用中经常用到,而且在理论中也有着特殊的重要性.

1. 单点分布

定义 2.3　若随机变量 X 的分布列为

$$P(X=C)=1,$$

其中 C 为常数,则称 X 服从**单点分布**.

自然地,常数是服从单点分布的随机变量.

2. 两点分布

定义 2.4　若随机变量 X 的分布列为

X	0	1
P	$1-p$	p

其中 $0<p<1$,则称 X 服从 **0-1 分布**,它是一个两点分布.

任何试验,当只考虑某事件 A 是否出现时,可看成是只有两种结果 $\{A,\overline{A}\}$ 的试验.只有两种可能结果的试验称为**伯努利试验**.如例2.1那样定义 A 的示性函数 I_A,它服从两点分布,恰好描述了伯努利试验.

3. 二项分布

定义 2.5　若随机变量 X 的分布列为

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

其中 $0<p<1$,$q=1-p$,则称 X 服从参数为 (n,p) 的**二项分布**.记为 $X\sim B(n,p)$.

由二项式定理知

$$1=(p+q)^n=\sum_{k=0}^{n}\mathrm{C}_n^k p^k q^{n-k}.$$

因此,该分布列满足应当具备的两条性质,定义是合理的.这也是称它为二项分布的原因.显然,$B(1,p)$ 即为两点分布.

在相同条件下,将同一试验重复进行,且各次试验的结果是相互独立的,这样的试验称为**独立重复试验**.将伯努利试验独立重复 n 次,称为 n **重伯努利试验**.

例 2.3　在 n 重伯努利试验中,设每次试验事件 A 发生的概率为 p.以 X 记 n 重伯努利试验中事件 A 发生的次数.令 $A_i=\{$第 i 次试验时事件 A 发生$\}$,$i=1,2,\cdots,n$,则

$$\{X=k\}=(A_1\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n)\cup\cdots\cup(\overline{A}_1\cdots\overline{A}_{n-k}A_{n-k+1}\cdots A_n),$$

其中等式右边的并共有 C_n^k 项,而且是互不相容、等概率的.由试验的独立性得

$$P(A_1\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n)=p^k(1-p)^{n-k}=p^k q^{n-k},$$

其中 $q=1-p$.因此,$P(X=k)=\mathrm{C}_n^k p^k q^{n-k}$,亦即 $X\sim B(n,p)$.

对 $i=1,2,\cdots,n$,有 $I_{A_i}\sim B(1,p)$,且 $X=\sum\limits_{i=1}^{n}I_{A_i}$.

例2.3表明,二项分布描述了 n 重伯努利试验中某事件 A 发生次数的概率分布.二项分

布是重要的离散型分布之一,它以 n 重伯努利试验为模型,在实践中有着广泛的应用.

下面的定理表明二项分布的分布列先递增后递减.

定理 2.2 设随机变量 $X \sim B(n,p)$,则

$$P(X=[(n+1)p])=\max_{0\leqslant k\leqslant n}\{P(X=k)\},$$

其中$[(n+1)p]$为$(n+1)p$ 的整数部分.特别地,若$(n+1)p$ 本身为整数,则

$$P(X=(n+1)p)=P(X=(n+1)p-1)=\max_{0\leqslant k\leqslant n}\{P(X=k)\}.$$

证明 对 $0\leqslant k\leqslant n$,计算

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

因此 $P(X=k)\geqslant P(X=k-1)$,当且仅当 $k\leqslant(n+1)p$.定理得证.

称$[(n+1)p]$为二项分布 $B(n,p)$的最可能出现次数,$P(X=[(n+1)p])$为二项分布的中心项.递推式 $P(X=k)=\dfrac{(n-k+1)p}{kq}P(X=k-1)$可用来编写计算机程序计算概率.

例 2.4 设每台自动机床在运行过程中需要维修的概率均为 $p=0.01$,并且各机床需要维修相互独立,如果(1) 每名维修工人负责看管 20 台机床;(2) 3 名维修工人共同看管 80 台机床,求不能及时维修的概率;(3) 假定有这种机床 200 台,为使不能及时维修的概率在 0.02 以下,问至少需要安排多少名维修工人共同看管这些机床?

解 设 X 为需要维修的机床数.

(1) $X\sim B(20,0.01)$,

$$P(X>1)=1-0.99^{20}-20\times0.01\times0.99^{19}\approx0.0169.$$

(2) $X\sim B(80,0.01)$,

$$P(X>3)=1-\sum_{k=0}^{3}C_{80}^k0.01^k\times0.99^{80-k}\approx0.0087.$$

可以看出,(2)的工作效率高.直观上这是容易理解的,因为在(2)的情形,当同时出现多台需要维修的机床时维修工人之间可以相互协助,以此提高及时维修的概率.当然,我们没有考虑维修需要的时间、管理成本等因素,因此在实际应用中还应该作更细致的分析.

(3) 因为 $X\sim B(200,0.01)$,所以问题化为求满足如下不等式的最小的 r,即

$$P(X>r)=\sum_{k=r+1}^{200}C_{200}^k0.01^k\times0.99^{200-k}\leqslant0.02.$$

而

$$P(X>5)=1-\sum_{k=0}^{5}C_{200}^k0.01^k\times0.99^{200-k}\approx0.0160,$$

即 $r=5$.

注意到,n 越大二项分布的计算量越大,直接计算是很麻烦的.

4. 泊松分布

定义 2.6 若随机变量 X 的分布列为

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

其中 $\lambda>0$,则称 X 服从参数为 λ 的**泊松分布**,记为 $X\sim P(\lambda)$.

因为 $\sum\limits_{k=0}^{\infty}\dfrac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum\limits_{k=0}^{\infty}\dfrac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$,所以定义是合理的.

设随机变量 $X\sim B(n,p)$,记 $\lambda=np$,则

$$C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}\prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right).$$

对充分大的 n 与适当大小的 λ，有

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda},\quad \left(1-\frac{\lambda}{n}\right)^k \approx 1,\quad \prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right)\approx 1.$$

因此，对充分大的 n 与适当大小的 λ，有

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

这表明，当 n 很大，p 很小，而 $\lambda=np$ 不太大时，可用泊松分布来近似二项分布. 由于泊松分布计算相对比较容易，可供查阅的表也更多，所以在具体应用中，若 $n\geqslant 20, p\leqslant 0.05$，取 $\lambda=np$，则常用如下的近似计算公式

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

第五章的中心极限定理将表明，当 $np(1-p)$ 较大时二项分布还可以由正态分布来近似.

在应用中，也常常遇到 n 无法确定的二项分布的情形，此时就直接使用泊松分布来描述. 例如，来到某售票窗口的人数，进入商场的顾客数，书中印刷错误的个数，布匹上的疵点数，纱锭上面纱断头次数，某放射性物质放射出的 α 粒子数，热电子的发射数，显微镜下在某观察范围内的微生物数等等.

类似于二项分布，泊松分布也有中心项. 计算

$$\frac{P(X=k)}{P(X=k-1)}=\frac{\lambda}{k},\quad k\geqslant 1, \tag{2.2}$$

知 $P(X=k)\geqslant P(X=k-1)$ 当且仅当 $k\leqslant\lambda$. 所以 $P(X=[\lambda])$ 为泊松分布的**中心项**. (2.2)式可以用来编写计算机程序计算概率. 顺便指出，(2.2)式也是泊松分布的充分条件，其证明留作习题.

重新考虑例 2.4 的(3)，可取 $\lambda=200\times 0.01=2$，查泊松分布表可得

$$P(X>5)\approx 1-\sum_{k=0}^{5}\frac{\lambda^k e^{-\lambda}}{k!}=\sum_{k=6}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}\approx 0.0166.$$

即 $r=5$，结果没有改变.

例 2.5　周末到达某商场的顾客数 N 服从参数为 $\lambda=1500$(人)的泊松分布，其中女性顾客占 $p=70\%$. 求到达商场的女性顾客数 N_F 的分布列.

解　对自然数 k，由全概率公式得

$$\begin{aligned}P(N_F=k) &= \sum_{n=0}^{\infty}P(N=n)P(N_F=k\mid N=n)=\sum_{n=k}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda}\cdot C_n^k p^k(1-p)^{n-k}\\ &= \sum_{n=k}^{\infty}\frac{[(1-p)\lambda]^{n-k}}{(n-k)!}\frac{1}{k!}e^{-\lambda}(p\lambda)^k=\frac{(p\lambda)^k e^{-p\lambda}}{k!}.\end{aligned}$$

即到达商场的女性顾客数仍服从泊松分布，参数变为 $p\lambda=1050$(人). 事实上，还可以类似证明，到达商场的男性顾客数 N_M 服从参数为 $(1-p)\lambda=450$(人)的泊松分布. 一般的泊松分布都可以这样来进行分解. 由归纳法泊松分布的随机变量可以分解为任意 n 个随机变量的和.

5. 超几何分布

定义 2.7　若随机变量 X 的分布列为

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

其中 $n\leqslant N, M\leqslant N$,则称 X 服从**超几何分布**.

利用组合的性质 $\sum\limits_{k=0}^{n} C_M^k C_{N-M}^{n-k}=C_N^n$,易知它满足分布列应当具备的性质,因此定义是合理的.注意到,上面的分布列中只有 k 满足 $n-(N-M)\leqslant k\leqslant\min\{n,M\}$ 的项是非 0 的.约定 $k<0$ 或 $n<k$ 时 $C_n^k\equiv 0$,则上式总成立.

例 2.6 N 件产品中有次品 M 件,从中任取 n 件,以 X 记这 n 件产品中的次品数,则 X 服从超几何分布.

如例 2.6 一样,通常的抽样是无放回的,超几何分布正是无放回抽样的概率模型.如果是有放回抽样,可以看作是独立重复试验,因此服从二项分布.但当$\frac{n}{N}$很小时,有放回与无放回的差别是很小的.而二项分布比超几何分布计算要简单些,因此常用二项分布近似超几何分布.具体地,当 N 很大,n 很小时,令 $p=\frac{M}{N}$,则

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}\approx C_n^k p^k(1-p)^{n-k},\quad k=0,1,2,\cdots,n.$$

6. 几何分布

定义 2.8 若随机变量 X 的分布列为

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

其中 $0<p<1, q=1-p$,则称 X 服从**几何分布**.

因为

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

所以定义是合理的.容易计算

$$P(X>n)=\sum_{k=n+1}^{\infty}pq^{k-1}=\frac{pq^n}{1-q}=q^n.$$

例 2.7 在独立重复试验中,设每次试验事件 A 发生的概率为 p.以 X 记事件 A 首次发生时所需的试验次数,则

$$\{X=k\}=\{\text{前 } k-1 \text{ 次试验 } A \text{ 不发生,第 } k \text{ 次 } A \text{ 发生}\}.$$

由独立性知

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

即 X 服从几何分布.

几何分布有下面的性质.

定理 2.3 取自然数值的随机变量 X 服从几何分布的充要条件是 X 具有**无记忆性**:

$$P(X>m+n\mid X>m)=P(X>n),\text{对任意的自然数 } m,n\geqslant 1. \tag{2.3}$$

证明 设 X 服从几何分布,则对任意的 $m,n\geqslant 1$,有

$$P(X>m+n\mid X>m)=\frac{P(X>m+n,X>m)}{P(X>m)}=\frac{P(X>m+n)}{P(X>m)}=\frac{q^{m+n}}{q^m}$$
$$=q^n=P(X>n).$$

另一方面,记 $g(n)=P(X>n)$.由(2.3)式知,对任意的 $m,n\geqslant 1$,有 $g(n)>0$,且

$$g(m+n)=g(m)g(n).$$

解该方程得 $g(n)=g(1)^n$. 由 $g(+\infty)=0$ 知 $g(1)<1$. 记 $q=g(1)$, $p=1-q$, 则对任意的 $k\geqslant 1$, 有

$$P(X=k)=P(X>k-1)-P(X>k)=q^{k-1}-q^k=pq^{k-1}.$$

得证 X 服从几何分布.

这表明, 在做了 m 次试验事件 A 未发生的条件下, 再做 n 次试验事件 A 仍未发生的概率等于从开始算起做 n 次试验事件 A 未发生的概率. 也就是说, 前面做的 m 次试验被忘记了. 这主要是由于是独立重复试验, 前面试验的结果对后面试验结果的概率没有影响造成的.

2.3　连续型随机变量

2.3.1　连续型随机变量及其概率密度

非离散型随机变量有不可列个取值. 若随机变量 X 的分布函数是连续的, 则当 $h\to 0$ 时, 有

$$P(X\in(a,a+h])=F(a+h)-F(a)\to 0.$$

若此概率收敛到 0 的速度也是知道的, 亦即 $P(X\in(a,a+h])\approx f(a)h+o(h)$, 则

$$f(a)=\lim_{h\to 0}\frac{F(a+h)-F(a)}{h}=F'(a).$$

若 $F(x)$ 还有更好的性质, 比如连续可微, 或者绝对连续, 则有

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

这种特殊类型的随机变量, 我们称之为连续型随机变量, 其准确定义如下.

定义 2.9　设随机变量 X 的分布函数为 $F(x)$. 若存在非负可积函数 $f(x)$, 使得对任意实数 x, 有

$$F(x)=\int_{-\infty}^{x} f(t)\,\mathrm{d}t, \tag{2.4}$$

则称 X 为连续型随机变量. $f(x)$ 称为 X 的**概率密度**.

需要说明的是, 在某些集合上改变 $f(x)$ 的值不会影响(2.4)式中积分的值, 所以在此意义下概率密度是不唯一的. 我们称这些不影响积分值的点的集合为**零测集**. 显然, 有限个点和可列无限个点的集合是零测集. 因此, 关于概率密度的结论总是除开一个零测集成立的, 除开一个零测集外的集合上函数值相等的概率密度视为同一函数. 但是, 在概率密度的众多版本中往往存在一个有较好性质的版本, 比如连续或分段连续的函数. 选取这样的版本对我们来说通常是方便的.

若随机变量 X 存在概率密度 $f(x)$, 则形式上有

$$f(x)=F'(x).$$

注意到, $F(x)$ 是连续函数, 但不是对所有的点都是可微的, 上式在除去一个零测集外是对的. 而且, 对 $f(x)$ 连续的点总是成立的. 亦即 X 为连续型随机变量, 当且仅当其分布函数 $F(x)$ 满足

$$F(x)=\int_{-\infty}^{x} F'(t)\,\mathrm{d}t,$$

且 $F'(x)$ 为 X 的概率密度. 由分布函数的性质可知, $f(x)$ 满足

$$\begin{cases} f(x) \geqslant 0, \\ \int_{-\infty}^{+\infty} f(x)\mathrm{d}t = 1. \end{cases} \tag{2.5}$$

事实上,给定一个满足(2.5)式的函数 $f(x)$,容易知道由(2.4)式定义的函数$F(x)$是一个分布函数,所以必存在概率空间及其上以 $f(x)$为概率密度的随机变量 X.因此,概率密度完整地描述了连续型随机变量.例如,由(2.4)式知,可以用概率密度来计算

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

而且,与离散型随机变量的分布列类似,我们通常直接给出概率密度来定义连续型随机变量.

由分布函数 $F(x)$的连续性知,对任意的 x,$P(X=x)=0$.因此,由概率的可列可加性,在计算连续型随机变量取值的概率时,我们总是可以忽略可列个取值.特别地,

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

如图 2.4 所示,X 落在区间$[a,b]$的概率等于概率密度$f(x)$在区间$[a,b]$的图形与水平轴所围区域的面积.更一般地,对 $\mathbf{R}$ 的所有子集 B(严格地说,B 为 $\mathbf{R}$ 的 Borel 子集),有

$$P(X \in B) = \int_B f(x)\mathrm{d}x.$$

由此知道,零测集上的取值也是可以忽略的.

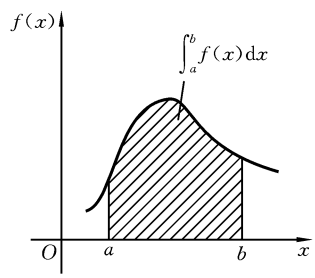

图 2.4

例 2.8　设连续型随机变量 X 的分布函数为

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

求:(1) 常数 A;(2) $P\left(|X|<\dfrac{\pi}{6}\right)$;(3) X 的概率密度 $f(x)$.

解　(1) 因为连续型随机变量的分布函数是连续的,所以

$$1=\lim_{x\uparrow\frac{\pi}{2}}F(x)=\lim_{x\uparrow\frac{\pi}{2}}A\sin x=A.$$

(2)
$$P\left(|X|<\frac{\pi}{6}\right)=P\left(-\frac{\pi}{6}<X<\frac{\pi}{6}\right)=F\left(\frac{\pi}{6}\right)-F\left(-\frac{\pi}{6}\right)=\frac{1}{2}.$$

(3)
$$f(x)=F'(x)=\begin{cases} \cos x, & 0<x<\dfrac{\pi}{2}, \\ 0, & 其他. \end{cases}$$

注意到,该分布函数只是一个分段连续可导的函数,在点 $x=0$ 处的左导数为 0,右导数为 1,因此是不可导的.但我们允许这样的情形发生,此时,$f(0)$点的值可以任意定义,所有这些函数都是 X 的概率密度.实际上,我们的做法通常只是分段求导,随手把那些不可导的点归为其中一类,而不必作特殊的定义.

应当指出,连续型随机变量的本质在于它有概率密度,即有函数 f 满足(2.4)式.因此理论上把它作为连续性随机变量的定义.至于它在一个区间连续取值倒不是本质的,甚至也是不确切的.因为我们总是可以忽略至少可列个概率为 0 的取值.

2.3.2　常见的连续型分布

与离散型随机变量类似，连续型随机变量是以概率密度来归类的. 下面将介绍几种重要的连续性随机变量的概率密度.

1. 均匀分布

定义 2.10　若随机变量 X 的概率密度为

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

其中 $a<b$，则称 X 服从区间$[a,b]$上的**均匀分布**，记为 $X\sim U(a,b)$.

如图 2.5 所示，容易验证该函数为一概率密度. 若 $X\sim U(a,b)$，容易计算其分布函数为

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t=\begin{cases}0, & x<a,\\ \dfrac{x-a}{b-a}, & a\leqslant x<b,\\ 1, & x\geqslant b.\end{cases}$$

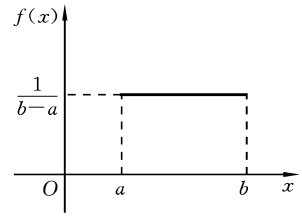

图 2.5

对任意的$[x,y]\subset[a,b]$，有

$$P(X\in[x,y])=F(y)-F(x)=\frac{y-x}{b-a}.$$

这说明 X 取值于区间$[a,b]$的任意子区间的概率只与该区间的长度有关. 这就是均匀分布的概率意义. 同时，可以证明具备该性质的随机变量必为均匀分布，这将作为习题留给读者. 几何型概率的样本点是在其样本空间中均匀分布的，例如，在例 2.2 中，$X\sim U(0,150)$. 另外，在定点计算中的舍入误差被认为是服从均匀分布的例子，比如，运算的数据只保留一位小数，小数点第一位以后的数字按四舍五入处理，则每次运算的舍入误差 $\varepsilon\sim U(-0.05,0.05)$. 利用此可进行误差分析.

均匀分布的重要性还体现在下面的结论. 令 $F(x)$为一分布函数. 对 $0<x<1$，记

$$F^{-1}(x)=\inf\{y:F(y)\geqslant x\},$$

称 $F^{-1}(x)$为 $F(x)$的**广义反函数**. 显然，若 $F(x)$严格单调增，$F^{-1}(x)$即为 $F(x)$的反函数(见图2.6). 由分布函数的性质(3)知，对每个 $0<x<1$，有

$$\{y:F(y)\geqslant x\}\neq\varnothing.$$

因此，$F^{-1}(x)$为取有限实数值的函数，即定义是合理的. 而且，它有下面的性质.

(1) $F^{-1}(x)$是单调非降的. 这是因为，对 $x_1<x_2$，有

$$\{y:F(y)\geqslant x_1\}\supset\{y:F(y)\geqslant x_2\}.$$

(2) $F(F^{-1}(x))\geqslant x$，即该下确界是可以取到的. 由此，

$$\{F(x)\geqslant y\}=\{x\geqslant F^{-1}(y)\}.$$

这是因为存在 y_n 满足 $y_n\downarrow F^{-1}(x)$，$F(y_n)\geqslant x$. 由分布函数 $F(x)$的右连续性有$F(y_n)\downarrow F(F^{-1}(x))$. 所以 $F(F^{-1}(x))\geqslant x$.

(3) 若 $F(x)$在 $F^{-1}(x)$处连续，则 $F(F^{-1}(x))=x$. 因为 $F\left(F^{-1}(x)-\dfrac{1}{n}\right)<x$ 且

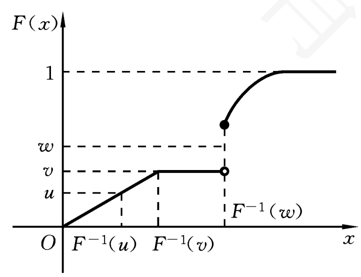

图 2.6

$$F\left(F^{-1}(x)-\frac{1}{n}\right)\uparrow F(F^{-1}(x)),$$

则 $F(F^{-1}(x))\leqslant x$.

设随机变量 X 的分布函数 $F(x)$ 是连续函数,令 $Y=F(X)$,则对每个 $0<y<1$,有

$$P(Y<y)=P(F(X)<y)=P(X<F^{-1}(y))=F(F^{-1}(y))=y,$$

及

$$P(Y\leqslant y)=\lim_{n\to\infty}P\left(F(X)<y+\frac{1}{n}\right)=\lim_{n\to\infty}\left(y+\frac{1}{n}\right)=y,$$

即 $Y\sim U(0,1)$. 证明过程中用到了 $F(x)$ 的连续性. 容易找到 $F(x)$ 不连续,$F(X)$ 不服从区间 $[0,1]$ 上均匀分布的反例.

设 $U\sim U(0,1)$,$F(x)$ 为一分布函数. 记 $X=F^{-1}(U)$,则

$$P(X\leqslant x)=P(F^{-1}(U)\leqslant x)=P(U\leqslant F(x))=F(x),$$

即 X 的分布函数为 $F(x)$. 这表明,任何分布的随机变量都是均匀分布随机变量的函数,而且有具体的函数表达式. 理论上,若要证明具有某种分布函数的随机变量是存在的,只需证明区间 $[0,1]$ 上均匀分布的存在性即可. 在随机模拟里,重要的是如何模拟随机数. 计算机的编程语言里都有一个产生区间 $[0,1]$ 上均匀分布随机数的函数,在MATLAB中是“rand”,称为伪随机数,利用它就能够容易模拟任何需要分布的随机数. 作为例子,若 $X\sim U(0,1)$,由区间 $[a,b]$ 上均匀分布函数的广义反函数可得

$$a+X(b-a)\sim U(a,b).$$

2. 指数分布

定义 2.11　若随机变量 X 的概率密度为

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

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

容易验证该概率密度的定义是合理的(见图 2.7). 设 $X\sim E(\lambda)$,则当 $x\geqslant 0$ 时,有

$$P(X>x)=\int_x^{+\infty}f(t)\,dt=e^{-\lambda x},$$

当 $x<0$ 时,$P(X>x)=1$. 因此其分布函数为

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

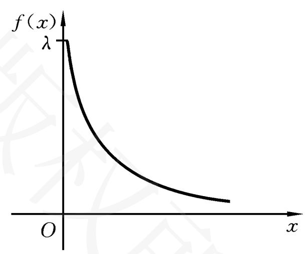

图 2.7

指数分布具有下面类似几何分布的无记忆性(见定理 2.4).

定理 2.4　取非负实值的随机变量 X 服从指数分布当且仅当它有下面的**无记忆性**:

$$P(X>x+y\mid X>x)=P(X>y),\quad x,y>0.$$

证明　设 $X\sim E(\lambda)$,则

$$P(X>x+y\mid X>x)=\frac{P(X>x+y,X>x)}{P(X>x)}=\frac{P(X>x+y)}{P(X>x)}=\frac{e^{-\lambda(x+y)}}{e^{-\lambda x}}$$
$$=e^{-\lambda y}=P(X>y).$$

另一方面,令 $g(x)=P(X>x)$,则

$$g(x+y)=g(x)g(y),\quad \forall\, x,y\geqslant 0. \tag{2.6}$$

由于 $g(x)=1-F_X(x)$,所以 $g(x)$ 单调非增,右连续,$g(0^+)=1$,$g(+\infty)=0$. 容易证明,满足

这些性质的函数只有 $g(x)=\mathrm{e}^{-\lambda x}$，其中 $\lambda>0$.

事实上，反复用(2.6)式得 $g(m/n)=g^m(1/n)$. 令 $m=n$ 得 $g(1)=g^n(1/n)$. 因此 $g(m/n)=(g(1))^{m/n}$. 由右连续性得对任意的 x，有 $g(x)=(g(1))^x$. 又由 $g(0^+)=1$ 得 $g(1)>0$，由 $g(+\infty)=0$ 得 $g(1)<1$. 最后记 $\lambda=-\ln(g(1))>0$. 得证.

设 X 表示某仪器的寿命. 无记忆性是说，该仪器使用了 x 小时后再继续使用 y 小时的概率等于该仪器从刚开始算起能使用 y 小时的概率. 也就是说，该仪器只要没有损坏将永远是“年轻”的，忘记了之前被使用过 x 小时. 在实际中，“永远年轻”是不可能的，因此只是一种近似. 例如，寿命长的电子元件在使用初期阶段本身老化的现象可以忽略不计，造成损坏的原因往往是高电压等因素. 因此在这一阶段指数分布比较确切地描述了其寿命的分布情况. 实际应用中，瓷盘、电子元件、计算机软件的寿命，电话的通话时间，服务系统的服务时间、等待时间，从现在开始到发生一次地震、爆发一场战争、收到一个错发的短信的时间等都认为是具有无记忆性，因此服从指数分布. 另外，指数分布也在连续时间 Markov 链中扮演重要的角色.

例 2.9　考虑有两个自动取款机的自助银行. 假定自动取款机为每位顾客的服务时间服从参数为 λ 的指数分布. 当你进入银行的时候，看到刚好有两个顾客在使用这两台取款机. 求在这三个顾客中你最后被服务完离开银行的概率，以及分别求另两个顾客最后离开银行的概率.

当你有取款机使用的时候，他们两个人中只有一人结束服务离开银行，另一人仍在接受服务. 由指数分布的无记忆性，从现在起仍在接受服务的这个人还需被服务的时间是参数为 λ 的指数分布，即与你接受服务的时间是同分布的. 因此，你最后被服务完离开银行的概率是 1/2. 由对称性知，他们两人最后离开的概率相同，都为 1/4.

若 $X\sim U(0,1)$，则

$$F^{-1}(X)=-\frac{1}{\lambda}\ln(1-X)$$

服从参数为 λ 的指数分布. 注意到也有 $1-X\sim U[0,1]$，所以 $-\frac{1}{\lambda}\ln X\sim E(\lambda)$. 设 $F(x)$ 是随机变量 X 的分布函数且连续，我们已经知道 $F(X)\sim U(0,1)$，记

$$R(x)=-\ln(1-F(x)),$$

则 $R(X)\sim E(1)$. 基于此，由于 $R(x)$ 在可靠性理论及金融保险等领域的应用，参数为 1 的指数分布在这些领域也有着重要的应用.

3. 正态分布

定义 2.12　若随机变量 X 的概率密度为

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

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

显然，$f(x)>0$. 记

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

为说明 $f(x)$ 确是一个概率密度，还需证明 $I=1$. 注意到，因为 $f(x)$ 的原函数不是通常的初等函数，所以该积分不能直接用求原函数的方法来计算. 下面的方法是通过计算该积分的平方来实现的. 关键的技巧是将累次积分转化为重积分，再利用极坐标计算.

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(y-\mu)^2}{2\sigma^2}} \mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{1}{2\pi} \mathrm{e}^{-\frac{(s^2+t^2)}{2}} \mathrm{d}s\mathrm{d}t \quad \left(s=\frac{x-\mu}{\sigma}, t=\frac{y-\mu}{\sigma}\right) \\
&= \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\infty} \mathrm{e}^{-\frac{r^2}{2}} r\mathrm{d}r = 1.
\end{aligned}
$$

由 $I \geqslant 0$ 知 $I=1$.

其分布函数为

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

如图 2.8 所示,概率密度 $f(x)$ 的图形是钟形曲线,且有下面的性质.

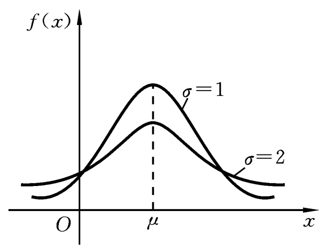

图 2.8

(1) $f(\mu-x)=f(\mu+x)$,所以图形关于 $x=\mu$ 对称. μ 仅影响图形的位置,不影响图形的形状.

(2) 当 $x=\mu$ 时,$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}$ 达到最大值. 因为图形与 x 轴围成的总面积等于 1,所以 σ 越大图形越扁平.

(3) $x=\mu\pm\sigma$ 是 $f(x)$ 的拐点,且当 $x\to\pm\infty$ 时均以 x 轴为渐近线.

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

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

如图 2.9 所示. 由于它不是一个初等函数,其值通常是近似计算得到的. 为方便应用,书后附表 2 编制了对非负 x 的 $\Phi(x)$ 的值. 由被积函数关于原点的对称性知,对负的 x 有 $\Phi(-x)=1-\Phi(x)$.

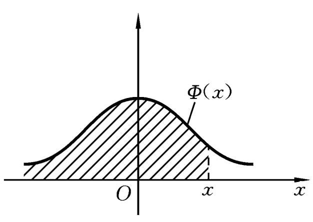

图 2.9

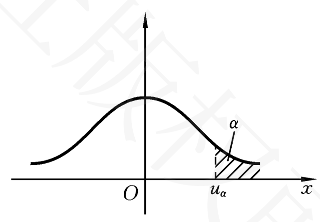

图 2.10

任给 $0<\alpha<1$,称满足等式

$$
P(X>u_\alpha)=\alpha
$$

的 u_α 为标准正态分布的上侧 α 分位点. 如图 2.10 所示,其直观意义是使得概率密度图形的右边与 x 轴围成的面积等于 α 所对应的横坐标的值. 易知

$$
\Phi(u_\alpha)=1-P(X>u_\alpha)=1-\alpha.
$$

注意到 $\Phi(x)$ 是严格单调增的函数,其反函数 $\Phi^{-1}(x)$ 存在且 $u_\alpha=\Phi^{-1}(1-\alpha)$. 因此,给定 $0<\alpha\leqslant 0.5$,查标准正态分布表 $1-\alpha$ 对应的值即为 u_α. 若 $\alpha>0.5$,查标准正态分布表 α 对应的值得到 $u_{1-\alpha}$. 由标准正态概率密度的对称性知

$$
u_\alpha=-u_{1-\alpha}.
$$

若 $X\sim N(\mu,\sigma^2)$，其分布函数

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

即由 $\Phi(x)$ 的值可以得到一般正态分布函数的值. 因此

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

例 2.10（3σ **原则**）　设随机变量 $X\sim N(\mu,\sigma^2)$，求 $P(|X-\mu|<3\sigma)$.

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

注意到，尽管在某一区间 $f(x)$ 恒严格为正，但由上面计算知，只计算 $(\mu-3\sigma,\mu+3\sigma)$ 中的值所引起的误差不到千分之三. 因此，在应用中通常只取 3σ 以内的值，落在此范围之外是个小概率事件，这可以作为可靠性、满意度的指标，称为"3σ 原则".

因自身有良好的性质及中心极限定理，正态分布无论在理论上还是应用中都是至关重要的连续型分布. 中心极限定理表明，如果一随机现象是许许多多偶然因素共同作用的总和，各偶然因素所起的作用势均力敌，没有哪个起主导作用，那么这个随机现象的概率模型就近似服从正态分布. 这在第五章中将会详细介绍. 另外，一些重要的分布也由正态分布导出，而且在正态分布的假设下可以得到丰富的结果. 这在数理统计部分将会有深刻体会. 应用中，例如，测量的误差，农作物的产量，工厂产品的尺寸、含量、强度各类质量指标，生物学中同一群体的身高、体重等形态指标，等等，都认为服从或近似服从正态分布.

2.3.3　混合型随机变量

在应用中也常常遇到一类离散型和连续型的混合型随机变量，它们既不是离散型也不是连续型的随机变量. 对于混合型随机变量，可以计算其分布函数，并由此容易计算感兴趣事件的概率.

例 2.11　设某电路受外界刺激电压 V 随机波动且 $V\sim E(\lambda)$. 现用电压表进行测量，电压表的最大读数为 V_0. 以 X 记电压表的读数，则

$$X=\min\{V,V_0\}.$$

注意到 $P(X=V_0)=P(V\geqslant V_0)=\mathrm{e}^{-\lambda V_0}$，即 $0<P(X=V_0)<1$. 而对所有的 $x\neq V_0$，$P(X=x)=0$. 因此 X 不是离散型也不是连续型，其分布函数（见图 2.11）为

$$F_X(x)=P(X\leqslant x)=\begin{cases}0, & x\leqslant 0,\\ 1-\mathrm{e}^{-\lambda x}, & 0<x<V_0,\\ 1, & x\geqslant V_0.\end{cases}$$

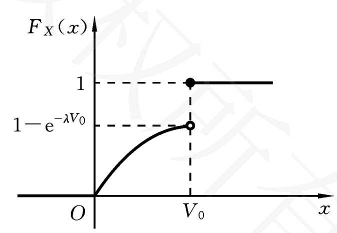

图 2.11

它也可以表示为

$$F_X(x)=(1-\mathrm{e}^{-\lambda V_0})F_1(x)+\mathrm{e}^{-\lambda V_0}F_2(x),$$

其中

$$F_1(x)=\begin{cases}0, & x<0,\\ \dfrac{1-\mathrm{e}^{-\lambda x}}{1-\mathrm{e}^{-\lambda V_0}}, & 0\leqslant x<V_0,\\ 1, & x\geqslant V_0,\end{cases}\quad F_2(x)=\begin{cases}0, & x<V_0,\\ 1, & x\geqslant V_0\end{cases}$$

均为分布函数，$F_1(x)$为连续型随机变量的分布函数，$F_2(x)$为离散型随机变量的分布函数. $F_X(x)$为离散型和连续型分布函数的凸组合，因此称 $F_X(x)$为离散型和连续型的混合型随机变量.

到此为止，我们可以处理三类随机变量，连续型、离散型、离散型和连续型的混合型. 这对我们实际应用的需要已经足够了. 因为我们漏掉的是一类被称为奇异型的随机变量，而目前奇异型的随机变量只在理论上有意义，在应用中还没有遇到过. 因此，在解决实际问题中，我们自然可以将目标随机变量与这三种类型对号入座，先确定类型，然后利用各自的特点研究之.

我们常说某随机变量的分布，其中术语“分布”是可以准确定义的. 由于要涉及测度论的知识，所以我们不给出其定义. 本书中的“分布”是粗略的术语，它通指该随机变量的分布函数，如果是离散型或连续型随机变量还分别可以指它的分布列与概率密度. 称某些随机变量“同分布”是指它们的分布函数相同.

2.4　随机变量函数的分布

无论在理论上还是在应用中，都常常需要处理是某个随机变量 X 的函数的随机变量 Y，我们想通过 X 的分布得到 Y 的分布. 这类问题既普遍又重要，本节将讨论这个问题.

设 $X(\omega)$为随机变量，是 $\Omega \mapsto \mathbf{R}$ 的映射，$g(x)$为通常的实函数，是 $\mathbf{R} \mapsto \mathbf{R}$ 的映射，则 $g(x)$与 X 的复合 $Y=g(X(\omega))$是 $\Omega \mapsto \mathbf{R}$ 的映射. 若想让 Y 为随机变量，则应当对所有的 x，使$\{Y\leqslant x\}$为事件. 为此，需要假定 $g(x)$有适当的性质，比如 $g(x)$为 Borel 函数. 幸运的是，这类函数相当广泛，我们通常遇到的函数都能满足该性质. 所以 $Y=g(X)$为随机变量.

自然地，我们会问：$Y=g(X)$是什么类型的随机变量呢？已知 X 的分布如何求 Y 的分布呢？一般而言，我们只需由 X 的分布求出 Y 的分布函数便同时回答了这两个问题. 也就是说，根据 X 的分布，具体计算 $F_Y(y)=P(Y\leqslant y)=P(g(X)\leqslant y)=P(X\in g^{-1}((-\infty,y]))$的值. X 为离散型的情形是较简单的，此时可以直接求 Y 的分布列. X 为连续型的情形相对复杂些. 下面我们将分别对 X 是离散型和连续型两类随机变量来回答这个问题.

2.4.1　离散型随机变量函数的分布

设 X 为离散型随机变量. 由函数的定义知道，随机变量 $Y=g(X)$取值的个数不会多于 X 取值的个数. 因此 Y 必定为离散型随机变量. 若已知 X 的分布列，则求 Y 的分布列，即求 Y 的所有可能的取值与取每个值的概率. 先看下面的例子.

例 2.12　设 X 的分布列为

X	-2	-1	0	1	2
P	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

求：(1) $Y_1=2X+1$ 的分布列；(2) $Y_2=X^2$ 的分布列.

解　(1) Y_1 的取值为$-3,-1,1,3,5$，而且

$$P(Y_1=-3)=P(2X+1=-3)=P(X=-2)=\frac{1}{10},$$

$$P(Y_1=-1)=P(2X+1=-1)=P(X=-1)=\frac{1}{5},$$

$$P(Y_1=1)=P(2X+1=1)=P(X=0)=\frac{2}{5},$$

$$P(Y_1=3)=P(2X+1=3)=P(X=1)=\frac{1}{5},$$

$$P(Y_1=5)=P(2X+1=5)=P(X=2)=\frac{1}{10}.$$

因此，Y_1 的分布列为

Y_1	-3	-1	1	3	5
P	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$

(2) Y_2 的取值为 4，1，0，而且

$$P(Y_2=4)=P(X^2=4)=P(X=-2)+P(X=2)=\frac{1}{5},$$

$$P(Y_2=1)=P(X^2=1)=P(X=1)+P(X=-1)=\frac{2}{5},$$

$$P(Y_2=0)=P(X^2=0)=P(X=0)=\frac{2}{5}.$$

因此，Y_2 的分布列为

Y_2	0	1	4
P	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

正如在例 2.12 中看到的那样，求离散型随机变量函数的分布列是不困难的. 一般地，设 X 的分布列为

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

则 $Y=g(X)$ 的分布列为

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

当然，这里可能有某些 $g(x_i)$ 相等，把它们做适当并项即可，或许在并项时的具体运算并不简单.

2.4.2　连续型随机变量函数的分布

连续型的情形就稍复杂一些. 注意到，若 X 为连续型随机变量，但 $Y=g(X)$ 可能不是连续型的随机变量. 如下例所示.

例 2.13　设 X 的概率密度为 $f(x)$，$g(x)=I_{[0,+\infty)}(x)$，则 $Y=g(X)$ 为两点分布，且

$$P(Y=0)=P(X<0)=\int_{-\infty}^{0} f(x)\mathrm{d}x,$$

$$P(Y=1)=P(X\geqslant 0)=\int_{0}^{+\infty} f(x)\mathrm{d}x.$$

从例 2.13 可看出，Y 为离散型随机变量的情形也是不复杂的. 和前面 X 为离散型的情形

的不同之处在于，前面并项的结果是至多可列项分布列的和，而这里由于为不可列项，所以是概率密度在某区间上的积分.

理论上，由于 $g(x)$ 的多样化，$g(X)$ 可以是任何一种类型的随机变量. 例如，例2.11为混合型的，例 2.13 为离散型的. 下面我们的兴趣是 Y 也是连续型随机变量的情形. 此时，设 X 的概率密度为 $f_X(x)$，欲求 Y 的概率密度 $f_Y(y)$. 为此，我们先求 Y 的分布函数 $F_Y(y)$.

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

然后再对 y 求导数得

$$f_Y(y) = F'_Y(y).$$

该方法称为**分布函数法**. 这种求分布函数的方法也适用于 Y 是所有其他类型的情形.

例 2.14 设 $X \sim N(\mu,\sigma^2)$，求 $Y = ax+b$ 的概率密度.

解 不妨设 $a>0$. 对任意的 y，有

$$F_Y(y) = P(Y\leqslant y) = P(ax+b\leqslant y) = P\left(X\leqslant \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

因此
$$f_Y(y) = F'_Y(y) = F'_X\left(\frac{y-b}{a}\right)\left[\frac{y-b}{a}\right]' = \frac{1}{\sqrt{2\pi}\sigma}\cdot \mathrm{e}^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{a}$$
$$= \frac{1}{\sqrt{2\pi}(\sigma a)}\cdot \mathrm{e}^{-\frac{[y-(b+a\mu)]^2}{2a^2\sigma^2}},$$

即 $Y \sim N(a\mu+b,\sigma^2a^2)$.

例 2.14 表明正态分布随机变量的线性函数仍为正态分布. 特别地，取 $a=\frac{1}{\sigma}$，$b=-\frac{\mu}{\sigma}$，得 $a\mu+b=0$，$\sigma^2a^2=1$. 因此 $\frac{X-\mu}{\sigma}\sim N(0,1)$.

例 2.15 设 $X\sim N(0,1)$，求 $Y=X^2$ 的概率密度.

解 如图 2.12 所示. 显然，当 $y\leqslant 0$ 时，$F_Y(y)=0$. 当 $y>0$ 时，

$$F_Y(y) = P(X^2\leqslant y) = P(-\sqrt{y}\leqslant X\leqslant\sqrt{y})$$
$$= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1.$$

因此，当 $y\leqslant 0$ 时，$f_Y(y)=0$，当 $y>0$ 时，

$$f_Y(y) = F'_Y(y) = 2\Phi'(\sqrt{y})\cdot\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}}\mathrm{e}^{-\frac{y}{2}}.$$

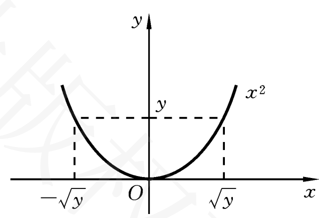

图 2.12

有时，我们也可以通过对积分作变量替换，将得到的分布函数直接改写为某函数的不定积分. 由概率密度的定义，该函数即为概率密度. 例如，在例 2.14 的后续的积分计算中令 $x=\frac{t-b}{a}$，有

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} f_X(x)\mathrm{d}x = \int_{-\infty}^{y} f_X\left(\frac{t-b}{a}\right)\cdot\frac{1}{a}\mathrm{d}t$$
$$= \int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{\left(\frac{t-b}{a}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{a}\mathrm{d}t = \int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}(\sigma a)}\mathrm{e}^{-\frac{[t-(b+a\mu)]^2}{2a^2\sigma^2}}\mathrm{d}t.$$

亦即 Y 的概率密度为

$$f_Y(y)=\frac{1}{\sqrt{2\pi}(\sigma a)}\mathrm{e}^{-\frac{[y-(b+a\mu)]^2}{2a^2\sigma^2}}.$$

在例 2.15 中，对 $y>0$，令 $t=x^2$，有

$$F_Y(y)=2\Phi(\sqrt{y})-1=2\int_0^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=\int_0^{y}\frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-\frac{t}{2}}\mathrm{d}t.$$

所以，Y 的概率密度为

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

例 2.16　设 $X\sim U\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$，求 $Y=\cos X$ 的概率密度.

解　如图 2.13 所示. 显然，当 $y<0$ 时，$F_Y(y)=0$，当 $y\geqslant 1$ 时 $F_Y(y)=1$. 当 $0\leqslant y<1$ 时，

$$\begin{aligned}F_Y(y)&=P(Y\leqslant y)=P(\cos X\leqslant y)=P\left(\left(-\frac{\pi}{2}\leqslant X\leqslant-\arccos y\right)\cup\left(\arccos y\leqslant X\leqslant\frac{\pi}{2}\right)\right)\\&=F_X(-\arccos y)-F_X\left(-\frac{\pi}{2}\right)+F_X\left(\frac{\pi}{2}\right)-F_X(\arccos y).\end{aligned}$$

因此

$$f_Y(y)=F_Y'(y)=f_X(-\arccos y)\cdot\frac{1}{\sqrt{1-y^2}}+f_X(\arccos y)\cdot\frac{1}{\sqrt{1-y^2}}=\frac{2}{\pi\sqrt{1-y^2}}.$$

Y 的概率密度为

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi\sqrt{1-y^2}}, & 0\leqslant y<1,\\ 0, & \text{其他}.\end{cases}$$

利用分布函数法我们能够证明下面的定理.

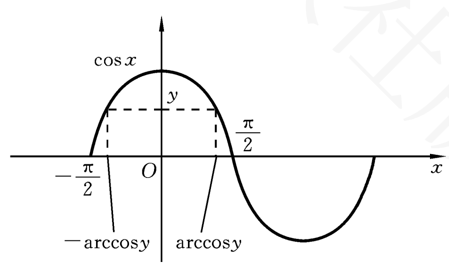

图 2.13

定理 2.5　设随机变量 X 的概率密度为 $f_X(x)$. 假设 $y=g(x)$ 是严格单调的函数，其反函数 $x=g^{-1}(y)$ 有连续导数，则 $Y=g(X)$ 也为连续型随机变量，且概率密度为

$$f_Y(y)=\begin{cases}f_X(g^{-1}(y))\left|[g^{-1}(y)]'\right|, & y\in(\alpha,\beta),\\ 0, & \text{其他},\end{cases}$$

其中 (α,β) 为 $g(x)$ 的值域.

证明　如果能证明 Y 的概率密度有此表达式，则说明 Y 有概率密度. 由连续型随机变量的定义，则 Y 为连续型随机变量.

不妨设 $g(x)$ 是单调增函数(单调减的情形是类似的)，此时 $g^{-1}(y)$ 也单调增，所以 $[g^{-1}(y)]'>0$. 记 $F_X(x)$ 为 X 的分布函数. 显然，当 $y\leqslant\alpha$ 时，$F_Y(y)=0$；当 $y\geqslant\beta$ 时，$F_Y(y)=1$；当 α

$<y<\beta$ 时,

$$F_Y(y)=P(Y\leqslant y)=P(g(X)\leqslant y)=P(X\leqslant g^{-1}(y))=F_X(g^{-1}(y)).$$

所以
$$f_Y(y)=F'_Y(y)=\begin{cases} f_X(g^{-1}(y))[g^{-1}(y)]', & y\in(\alpha,\beta),\\ 0, & \text{其他}\end{cases}$$

再看例 2.14. 不妨设 $a>0$. $g(x)=ax+b$ 严格单调增, $g^{-1}(y)=\dfrac{y-b}{a}$, $[g^{-1}(y)]'=\dfrac{1}{a}$ 连续, $(\alpha,\beta)=(-\infty,+\infty)$.

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

所以,对所有的 y,有

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}}\cdot\frac{1}{a}=\frac{1}{\sqrt{2\pi}(\sigma a)}\mathrm{e}^{-\frac{[y-(b+a\mu)]^2}{2a^2\sigma^2}}.$$

即与前面的结果一致.

推论 2.1　设随机变量 X 的概率密度为 $f_X(x)$. 假设 $g(x)$ 在不相重叠的区间 $I_1,I_2,\cdots$ 上逐段满足定理 2.5 的条件,且分别以 $g_i^{-1}(y)$, $[g_i^{-1}(y)]'(i=1,2,\cdots)$ 记 $y=g(x)$ 在各段上的反函数及其导数. 则 $Y=g(X)$ 也为连续型随机变量,且概率密度为

$$f_Y(y)=\begin{cases} \displaystyle\sum_{\{i:y\in g(I_i)\}} f_X(g_i^{-1}(y))\,|[g_i^{-1}(y)]'|, & y\in(\alpha,\beta),\\ 0, & \text{其他}.\end{cases}$$

证明　对 $\alpha<y<\beta$, 记 $I_i(y)=\{x\in I_i:g(x)\leqslant y\}$, 有

$$\begin{aligned} F_Y(y)&=P(Y\leqslant y)=P(g(X)\leqslant y)=P\Big(\bigcup_i(I_i\cap(g(X)\leqslant y))\Big)\\ &=\sum_i\int_{-\infty}^{y} I_{\{t\in g(I_i)\}}f_X(g_i^{-1}(t))\,|[g_i^{-1}(t)]'|\,\mathrm{d}t. \end{aligned}$$

对 y 求导即完成了证明.

该推论表明,我们可以逐段应用定理 2.5,然后再对每段求和.

再看例 2.15. $y=g(x)=x^2$, $I_1=(-\infty,0]$, $I_2=(0,+\infty)$ 满足推论 2.1 的条件.

$$g_1^{-1}(y)=-\sqrt{y},\quad g_2^{-1}(y)=\sqrt{y},$$

$$[g_1^{-1}(y)]'=-\frac{1}{2\sqrt{y}},\quad [g_2^{-1}(y)]'=\frac{1}{2\sqrt{y}}$$

连续. 因此,当 $y\geqslant 0$ 时,

$$f_Y(y)=\Phi'(-\sqrt{y})\left|-\frac{1}{2\sqrt{y}}\right|+\Phi'(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|=\frac{1}{\sqrt{2\pi y}}\mathrm{e}^{-\frac{y}{2}},$$

当 $y<0$ 时, $f_Y(y)=0$. 与前面的结果一致.

尽管遇到的很多情形都可以应用推论 2.1,但是我们仍建议读者采用最一般的分布函数法. 一方面,事先常常我们并不知道 Y 的类型,但求分布函数的方法总是适用的,因此用此方法总可以求出分布函数. 其他类型的情形无非是求导后的函数不是一个概率密度而已,此时的随机变量正是由分布函数来描述的. 另一方面,我们观察到定理 2.5 及推论 2.1 的证明也只是分布函数法的简单应用,甚至不易比较出定理的证明与叙述哪个更简洁. 通过做习题容易掌握分布函数法,这样可以养成一个良好的思维习惯.

习　题　二

2.1　填空题.

(1) 设随机变量 X 的分布函数为

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

则 X 的分布列为______.

(2) 设随机变量 $X\sim B(2,p)$,$Y\sim B(3,p)$,$P(X\geqslant 1)=\dfrac{5}{9}$,则 $P(Y\geqslant 1)=$______.

(3) 设随机变量 X 的概率密度为

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

以 Y 表示对 X 的三次独立重复观察中事件 $\left\{X\leqslant\dfrac{1}{2}\right\}$ 出现的次数,则 $P(Y=2)=$______.

(4) 设随机变量 X 的概率密度为

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

若使得 $P(X\geqslant k)=\dfrac{2}{3}$,则 $k=$______.

(5) 若随机变量 X 服从$[1,6]$上的均匀分布,则方程 $y^2+Xy+1=0$ 有实根的概率是______.

(6) 若随机变量 X 服从正态分布 $N(2,\sigma^2)$,$P(2<X<4)=0.3$,则 $P(X<0)=$______.

(7) 设随机变量 Y 服从参数为 1 的指数分布,a 为常数且大于零,则 $P\{Y\leqslant a+1\mid Y>a\}=$______.

2.2　选择题.

(1) 设 $F_1(x)$ 与 $F_2(x)$ 分别为随机变量 X_1 与 X_2 的分布函数.为使 $F(x)=aF_1(x)-bF_2(x)$ 是某一随机变量的分布函数,在下列给定的各组数值中应取(　　).

(A) $a=\dfrac{3}{5},b=-\dfrac{2}{5}$　(B) $a=\dfrac{2}{3},b=\dfrac{2}{3}$　(C) $a=-\dfrac{1}{2},b=\dfrac{3}{2}$　(D) $a=\dfrac{1}{2},b=-\dfrac{3}{2}$

(2) 设随机变量 X 服从指数分布,则随机变量 $Y=\min\{X,2\}$ 的分布函数的间断点个数为(　　).

(A) 0　(B) 1　(C) 2　(D) 大于 2

(3) 设随机变量 X 服从正态分布 $N(\mu,\sigma^2)$,则随 σ 的增大,概率 $P(|X-\mu|<1)$(　　).

(A)单调增　(B) 单调减　(C) 保持不变　(D) 增减不定

(4) 设随机变量 X 服从标准正态分布,u_α 为其上侧 α 分位点.若 $P(|X|<x)=\alpha$,则 $x=$(　　).

(A) $u_{\frac{\alpha}{2}}$　(B) $u_{1-\frac{\alpha}{2}}$　(C) $u_{\frac{1-\alpha}{2}}$　(D) $u_{1-\alpha}$

(5)设随机变量 X 与 Y 相互独立,且分别服从参数为 1 与参数为 4 的指数分布,则 $P\{X<Y\}=$(　　).

(A) $\dfrac{1}{5}$　(B) $\dfrac{1}{3}$　(C) $\dfrac{2}{5}$　(D) $\dfrac{4}{5}$

2.3　已知离散型随机变量 X 的分布列为

$$P(X=1)=0.2,\quad P(X=2)=0.3,\quad P(X=3)=0.5.$$

求 X 的分布函数 $F(x)$.

2.4　某保险公司多年的统计资料表明,在索赔户中被盗索赔户占 20%.以 X 表示在随意抽查的 100 个

索赔户中因被盗向保险公司索赔的户数,求 X 的分布列.

2.5　假设有 10 只同种电子元件,其中有 2 只废品.装配仪器时,从这批元件中任取一只,如是废品,则扔掉重新任取一只;如仍是废品,则扔掉再取一次.求在取到正品之前已取出的废品只数 X 的分布列.

2.6　从学校乘汽车到火车站的途中有 3 个十字路口,假设在每个十字路口遇到红灯的事件是相互独立的,并且概率都是 $\frac{2}{5}$.用 X 记途中遇到红灯的次数,求 X 的分布列和分布函数.

2.7　已知甲、乙两箱中装有同种产品,其中甲箱中装有 3 件合格品和 3 件次品,乙箱中仅装有 3 件合格品.从甲箱中任取 3 件产品放入乙箱后,求乙箱中次品件数 X 的分布列.

2.8　已知随机变量 X 的概率密度

$$f(x)=\frac{1}{2}\mathrm{e}^{-|x|},\quad x\in\mathbf{R},$$

求随机变量 X 的分布函数 $F(x)$.

2.9　假设一大型设备在任何长为 t 的时间内发生故障的次数 $N(t)$ 服从参数为 λt 的泊松分布.求相继两次故障之间时间间隔 T 的概率分布,以及在设备已无故障工作 8 h 的情形下,再无故障运行 8 h 的概率 Q.

2.10　设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\frac{1}{2}\cos\frac{x}{2}, & 0\leqslant x\leqslant\pi,\\ 0, & \text{其他},\end{cases}$$

对 X 独立地重复观察 4 次,用 Y 记观察值大于 $\frac{\pi}{3}$ 的次数,求随机变量 Y 的分布列.

2.11　设随机变量 X 和 Y 同分布,X 的概率密度为

$$f(x)=\begin{cases}\frac{3}{8}x^2, & 0<x<2,\\ 0, & \text{其他}.\end{cases}$$

已知事件 $A=\{X>a\}$ 和 $B=\{Y>a\}$ 独立,且 $P(A\cup B)=\frac{3}{4}$,求常数 a.

2.12　假定随机变量 X 只在区间 $[0,1]$ 上取值,且对任意的 $[x,y]\subset[0,1]$,$P(X\in[x,y])$ 只与 $y-x$ 有关.证明 $X\sim U[0,1]$.

2.13　设随机变量 X 的绝对值不大于 1,$P(X=-1)=\frac{1}{8}$,$P(X=1)=\frac{1}{4}$.在事件 $\{-1<X<1\}$ 出现的条件下,X 在 $(-1,1)$ 内的任一子区间上取值的条件概率与该子区间的长度成正比,试求 X 的分布函数 $F(x)$ 及 X 求负值的概率 p.

2.14　某仪器装有 3 只独立工作的同型号电子元件,其寿命(单位:h)都服从参数为 $\frac{1}{600}$ 的指数分布.求在仪器使用的最初 200 h 内,至少有一只电子元件损坏的概率 α.

2.15　假设一设备开机后无故障工作的时间 X 服从参数为 $\frac{1}{5}$ 的指数分布.设备定时开机,出现故障时自动关机,而在无故障的情况下工作 2 h 便关机.试求该设备每次开机无故障工作的时间 Y 的分布函数 $F(y)$.

2.16　在电源电压不超过 200 V、200～240 V 和超过 240 V 三种情况下,某种电子元件损坏的概率分别为 0.1,0.001,0.2.假设电源电压 X 服从正态分布 $N(220,25^2)$.求该电子元件损坏的概率 α 和该电子元件损坏时电源电压在 200～240 V 的概率 β.

2.17　假设测量的随机误差 X 服从 $N(0,10^2)$.试求在 100 次独立重复测量中,至少有三次测量误差的绝对值大于 19.6 的概率 α,并利用泊松分布求出 α 的近似值.

2.18　设 $X\sim N(\mu,\sigma^2)$,方程 $y^2+4y+X=0$ 无实根的概率是 0.5,求 μ.

2.19　设随机变量 X 的概率密度为 $f(x)=\frac{1}{\pi(1+x^2)}$,求随机变量 $Y=1-\sqrt[3]{X}$ 的概率密度 $f_Y(y)$.

2.20　设随机变量 X 服从 $[1,2]$ 上的均匀分布,求随机变量 $Y=\mathrm{e}^{2X}$ 的概率密度 $f_Y(y)$.

2.21　设随机变量 X 的概率密度为

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

求随机变量 $Y=\mathrm{e}^{X}$ 的概率密度 $f_Y(y)$.

2.22　设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{2}, & -1<x<0,\\ \dfrac{1}{4}, & 0\leqslant x<2,\\ 0, & \text{其他},\end{cases}$$

求随机变量 $Y=X^2$ 的概率密度 $f_Y(y)$.

2.23　设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{3\sqrt[3]{x^2}}, & x\in[1,8],\\ 0, & \text{其他},\end{cases}$$

$F(x)$是 X 的分布函数,求随机变量 $Y=F(X)$的分布函数.

2.24　设随机变量 X 服从参数为 2 的指数分布. 证明 $Y=1-\mathrm{e}^{-2X}$服从$[0,1]$上的均匀分布.

2.25　甲向一个目标射击,直到击中 r 次为止. 用 X 表示射击停止时的射击次数. 如果甲每次射击击中目标的概率是 $p\in(0,1)$,证明 X 的分布列为

$$P(X=k)=\mathrm{C}_{k-1}^{r-1}p^r(1-p)^{k-r},\quad k=r,r+1,\cdots,$$

这时称 X 服从帕斯卡分布. 记 $Y=X-r$,Y 为射击停止时射击失败的次数. 证明 Y 的分布列为

$$P(Y=k)=\mathrm{C}_{k+r-1}^{r-1}p^r(1-p)^k,\quad k=0,1,\cdots,$$

这时称 Y 服从负二项分布.

2.26　设 T 是表示寿命的非负随机变量,有连续的概率密度 $f(x)$. $S(x)=P(T>x)$,$\lambda(t)=\dfrac{f(t)}{S(t)}$分别称为生存函数和失效率函数. 证明

$$S(x)=\exp\left\{-\int_0^x\lambda(t)\mathrm{d}t\right\}.$$

2.27　若 X 为取非负整数值的随机变量,分布列为 $P(X=k)=p_k$,$k=0,1,\cdots$,满足

$$\frac{p_{k+1}}{p_k}=\frac{\lambda}{k},$$

其中 $\lambda>0$,则 $X\sim P(\lambda)$.

2.28　设 X 为随机变量,$\ln X\sim N(\mu,\sigma^2)$,证明 X 的概率密度为

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma x}\mathrm{e}^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},\quad x>0.$$

这时称 X 服从对数正态分布.

2.29　设随机变量 X 的概率密度为 $f(x)=\begin{cases}\dfrac{1}{a}x^2, & 0<x<3,\\ 0, & \text{其他}\end{cases}$,令随机变量 $Y=\begin{cases}2, & X\leqslant 1,\\ X, & 1<X<2,\\ 1, & X\geqslant 2.\end{cases}$

(1)求 Y 的分布函数;(2)求概率 $P(X\leqslant Y)$.

第三章　多维随机变量及其分布

在实际应用中，我们往往需要多个随机变量来共同描述一个随机试验的结果．例如，在某平面区域中随机地取点，需要用横坐标和纵坐标表示随机点所处的位置；再如，研究 n 种股票的价格波动情况，通常用成交量和即时价格来描述．这些随机变量是相互联系的，必须把它们作为一个整体来研究，讨论它们统计规律性及各变量之间的相互联系．

3.1　多维随机变量

3.1.1　多维随机变量

我们先引入多维随机变量的定义．

定义 3.1　设 $X_1,X_2,\cdots,X_n$ 是定义在给定概率空间$(\Omega,\mathscr{F},P)$上的 n 个随机变量，则称$(X_1,X_2,\cdots,X_n)$为 n 维随机变量(或随机向量)．

为记号简单起见，我们一般只讨论二维随机变量，多维情形的讨论可类似进行．

定义 3.2　设(X,Y)为二维随机变量，称 $F(x,y)=P(X\leqslant x,Y\leqslant y)$为$(X,Y)$的联合分布函数，而称 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．

实际上，联合分布函数 $F(x,y)$就是随机点(X,Y)落在以(x,y)为右上顶点的无穷矩形$(-\infty,x]\times(-\infty,y]$内的概率，如图 3.1 所示．

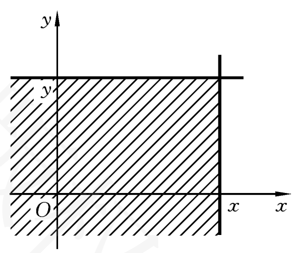

图 3.1

与一维情形类似，我们容易得到联合分布函数的性质．

定理 3.1　设 $F(x,y)$为(X,Y)的联合分布函数，则有

(1) 对任意 x,y，$0\leqslant F(x,y)\leqslant 1$；

(2) $F(x,y)$分别关于 x 和 y 单调非降；

(3) $F(x,y)$分别关于 x 和 y 右连续；

(4) 对任意 x,y，$F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0$，$F(+\infty,+\infty)=1$；

(5) 对任意 $x_1<x_2$，$y_1<y_2$，有

$$F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\geqslant 0; \tag{3.1}$$

(6) 对任意 x,y，$F(+\infty,y)=F_Y(y)$，$F(x,+\infty)=F_X(x)$．

证明　只证(5)．如图 3.2 所示，我们有

$$\begin{aligned}0 &\leqslant P(x_1<X\leqslant x_2,y_1<Y\leqslant y_2)\\ &=P(X\leqslant x_2,Y\leqslant y_2)-P(X\leqslant x_1,Y\leqslant y_2)\\ &\quad -P(X\leqslant x_2,Y\leqslant y_1)+P(X\leqslant x_1,Y\leqslant y_1)\\ &=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1).\end{aligned}$$

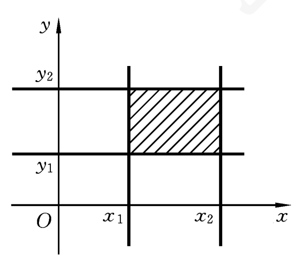

图 3.2

可以证明，具有上述性质(1)～(5)的二元函数$F(x,y)$一定是

某二维随机变量(X,Y)的联合分布函数.

例 3.1　定义二元函数

$$F(x,y)=\begin{cases}0, & x<0 \text{ 或 } y<0,\\ \dfrac{1}{2}xy+\dfrac{1}{2}y, & 0\leqslant y\leqslant x<1,\\ \dfrac{1}{2}xy+\dfrac{1}{2}x, & 0\leqslant x\leqslant y<1,\\ x, & 0\leqslant x<1,y\geqslant 1,\\ y, & 0\leqslant y<1,x\geqslant 1,\\ 1, & x\geqslant 1,y\geqslant 1,\end{cases}$$

容易验证$F(x,y)$满足定理 3.1 的(1)～(5),因此,$F(x,y)$可以看做某二维随机变量(X,Y)联合分布函数.由定理 3.1 中(6)可知,X,Y为一维同分布的连续型随机变量,其分布函数为

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

3.1.2　二维离散型随机变量

二维随机变量与一维随机变量类似,常用的有离散型与连续型两类,我们先考虑离散情形.

定义 3.3　若二维随机变量(X,Y)只取有限个或可列个值,则称(X,Y)为二维离散型随机变量,称$P(X=x_i,Y=y_j)=p_{ij}$为(X,Y)的联合分布列,称$P(X=x_i)=p_{i\cdot}$(相应地$P(Y=y_j)=p_{\cdot j}$)为(X,Y)关于X(相应地Y)的边缘分布列.

易见:

(1) $p_{ij}\geqslant 0$;

(2) $\sum\limits_{i,j}p_{ij}=1$;

(3) $p_{i\cdot}=\sum\limits_{j}p_{ij},p_{\cdot j}=\sum\limits_{i}p_{ij}$;　(3.2)

(4) 若对任意$B\subset\mathbf{R}^2$,$\{(X,Y)\in B\}\in\mathscr{F}$,则有

$$P((X,Y)\in B)=\sum_{(x_i,y_j)\in B}p_{ij}. \tag{3.3}$$

通常我们用如下联立表直观地表示联合分布列与边缘分布列的情况.

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

一般地,当$p_{i\cdot}\neq 0$或$p_{\cdot j}\neq 0$时我们可以由条件概率的乘积公式来计算p_{ij},即

$$p_{ij}=P(X=x_i)P(Y=y_j\,|\,X=x_i)=P(Y=y_j)P(X=x_i\,|\,Y=y_j). \tag{3.4}$$

例 3.2　设 X,Y 的分布列分别为

X	-1	0	1
P	$1/4$	$1/2$	$1/4$

，

Y	0	1
P	$1/2$	$1/2$

，

且 $P(XY=0)=1$，求 (X,Y) 的联合分布列.

解　由于 $P(XY=0)=1$，所以 $P(XY\neq 0)=0$，再由联合分布列与边缘分布列的关系得到

$X \backslash Y$	0	1	
-1	$\frac{1}{4}$	0	$\frac{1}{4}$
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$	

例 3.3　设袋中有 a 个红球 b 个白球，采用(1) 有放回方式；(2) 无放回方式取两次球，定义

$$X=\begin{cases}1, & \text{第一次取红球},\\ 0, & \text{第一次取白球},\end{cases}\quad Y=\begin{cases}1, & \text{第二次取红球},\\ 0, & \text{第二次取白球},\end{cases}$$

求 (X,Y) 的联合分布列及边缘分布列.

解　(1) 在有放回情形.

(X,Y) 的联合分布列及边缘分布列如下所示.

$X \backslash Y$	1	0	
1	$\frac{a^2}{(a+b)^2}$	$\frac{ab}{(a+b)^2}$	$\frac{a}{a+b}$
0	$\frac{ba}{(a+b)^2}$	$\frac{b^2}{(a+b)^2}$	$\frac{b}{a+b}$
	$\frac{a}{a+b}$	$\frac{b}{a+b}$	

(2) 在无放回情形.

(X,Y) 的联合分布列及边缘分布列如下所示.

$X \backslash Y$	1	0	
1	$\frac{a(a-1)}{(a+b)(a+b-1)}$	$\frac{ab}{(a+b)(a+b-1)}$	$\frac{a}{a+b}$
0	$\frac{ba}{(a+b)(a+b-1)}$	$\frac{b(b-1)}{(a+b)(a+b-1)}$	$\frac{b}{a+b}$
	$\frac{a}{a+b}$	$\frac{b}{a+b}$	

可见，两种情形有相同的边缘分布列，但联合分布列不同.

3.1.3　二维连续型随机变量

我们再来考虑连续情形.

定义 3.4　若存在二元非负函数 $f(x,y)$ 使得二维随机变量 (X,Y) 的分布函数为

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

则称 (X,Y) 为一二维连续型随机变量，称 $f(x,y)$ 为 (X,Y) 的联合概率密度（简称概率密度）.

下面的定理给出二维连续型随机变量的有关性质.

定理 3.2　设 (X,Y) 为二维连续型随机变量，$f(x,y)$，$F(x,y)$ 分别为其联合概率密度与联合分布函数，则有

(1) $F(+\infty,+\infty)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=1$；　(3.5)

(2) 对任意 $B\subset\mathbf{R}^2$，$\{(X,Y)\in B\}\in\mathscr{F}$，则有 $P((X,Y)\in B)=\iint\limits_{B}f(x,y)\mathrm{d}x\mathrm{d}y$；　(3.6)

(3) 在 $f(x,y)$ 的连续点上，有 $f(x,y)=\dfrac{\partial^2F(x,y)}{\partial x\partial y}$；　(3.7)

(4) X,Y 为一维连续型随机变量，且它们的概率密度分别为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y,\quad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x.\tag{3.8}$$

我们称 $f_X(x)$（或 $f_Y(y)$）为 (X,Y) 关于 X（或 Y）的边缘概率密度.

证明　只证性质(4). 由于

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty}f(u,v)\mathrm{d}v\right)\mathrm{d}u,$$

所以

$$f_X(x)=F_X'(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y.$$

例 3.4　设

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

求：(1) 常数 a；(2) $f_X(x)$，$f_Y(y)$；(3) $F(x,y)$；(4) $P(X+Y\geqslant 1)$.

解　(1) 由定理 3.2 性质(1)，有

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

得到 $a=\dfrac{1}{3}$.

(2) 当 $0\leqslant x\leqslant 1$ 时，

$$f_X(x)=\int_0^2\left(x^2+\frac{1}{3}xy\right)\mathrm{d}y=2x^2+\frac{2}{3}x;$$

当 $x>1$ 或 $x<0$ 时，

$$f_X(x)=0;$$

当 $0\leqslant y\leqslant 2$ 时，

$$f_Y(y)=\int_0^1\left(x^2+\frac{1}{3}xy\right)\mathrm{d}x=\frac{1}{3}+\frac{1}{6}y;$$

当 $y>2$ 或 $y<0$ 时，

$$f_Y(y)=0.$$

(3) 如图 3.3 所示,共分 5 种情形:

对 $x<0$ 或 $y<0$(见图 3.3(a)),有

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}0\mathrm{d}x\mathrm{d}y=0;$$

对 $0\leqslant x<1,0\leqslant y<2$(见图 3.3(b)),有

$$F(x,y)=\int_{0}^{x}\int_{0}^{y}\left(u^2+\frac{1}{3}uv\right)\mathrm{d}u\mathrm{d}v=\frac{1}{3}x^3y+\frac{1}{12}x^2y^2;$$

对 $0\leqslant x<1,y\geqslant 2$(见图 3.3(c)),有

$$F(x,y)=\int_{0}^{x}\int_{0}^{2}\left(u^2+\frac{1}{3}uv\right)\mathrm{d}u\mathrm{d}v=\frac{2}{3}x^3+\frac{1}{3}x^2;$$

对 $x\geqslant 1,0\leqslant y<2$(见图 3.3(d)),有

$$F(x,y)=\int_{0}^{1}\int_{0}^{y}\left(u^2+\frac{1}{3}uv\right)\mathrm{d}u\mathrm{d}v=\frac{1}{3}y+\frac{1}{12}y^2;$$

对 $x\geqslant 1,y\geqslant 2$(见图 3.3(e)),有

$$F(x,y)=1.$$

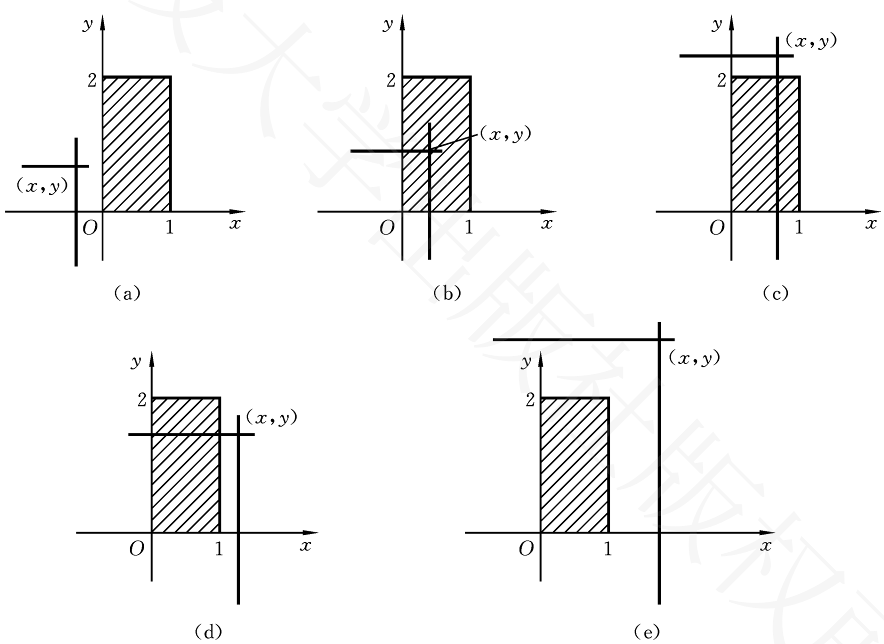

图 3.3

(4) 积分区域如图 3.4 中阴影部分所示,于是

$$P(X+Y\geqslant 1)=\int_{0}^{1}\left(\int_{1-x}^{2}\left(x^2+\frac{1}{3}xy\right)\mathrm{d}y\right)\mathrm{d}x=\frac{65}{72}.$$

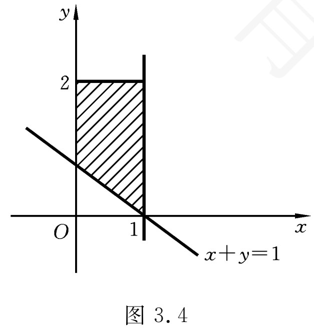

图 3.4

注意　与离散型随机变量不同,两个一维连续型随机变量 X,Y 并不一定构成一个二维连续型随机变量 (X,Y),如 (X,X) 及例 3.1 对应的二维随机变量 (X,Y).

下面介绍两个常用的二维连续型随机变量.

1. 二维均匀分布

若 $D\subset\mathbf{R}^2$,且其面积 $m(D)$ 满足 $0<m(D)<+\infty$,则称具

有概率密度

$$f(x,y)=\begin{cases}1/m(D), & (x,y)\in D,\\ 0, & (x,y)\notin D\end{cases} \tag{3.9}$$

的二维连续型随机变量(X,Y)服从区域D上的均匀分布,记为$(X,Y)\sim U(D)$.

例 3.5　设D是由$y=x$与$y=x^2$所围区域,$(X,Y)\sim U(D)$,求边缘概率密度$f_X(x)$,$f_Y(y)$.

解　区域D如图3.5所示,其面积$m(D)=\int_0^1(x-x^2)\mathrm{d}x\mathrm{d}y=\dfrac{1}{6}$,所以,

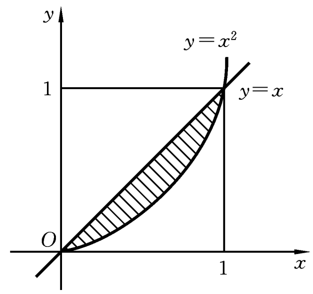

图 3.5

对$0\leqslant x\leqslant 1$,有

$$f_X(x)=\int_{x^2}^{x}6\mathrm{d}y=6(x-x^2);$$

对$x>1$或$x<0$,有

$$f_X(x)=0;$$

对$0\leqslant y\leqslant 1$,有

$$f_Y(y)=\int_{y}^{\sqrt{y}}6\mathrm{d}x=6(\sqrt{y}-y);$$

对$y>2$或$y<0$,有

$$f_Y(y)=0.$$

可见,多维均匀分布的边缘分布不一定是一维均匀分布.

2. 二维正态分布

若随机变量(X,Y)的密度

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \tag{3.10}$$

则称二维连续型随机变量(X,Y)服从参数为$(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$的二维正态分布,记为$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,其中五个参数的取值范围分别是:$\mu_1,\mu_2\in\mathbf{R}$;$\sigma_1,\sigma_2>0$;$|\rho|<1$.它们的具体含义将在第四章介绍.二维正态分布的概率密度的图形如图3.6所示.

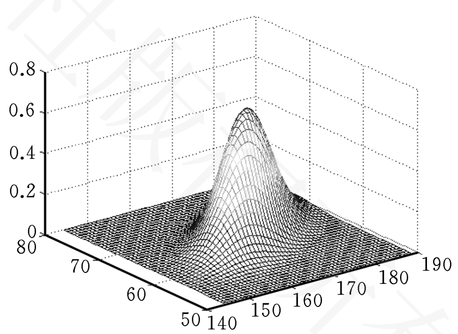

图 3.6

如果记向量$\boldsymbol{x}=(x,y)$,$\boldsymbol{\mu}=(\mu_1,\mu_2)$及矩阵$\boldsymbol{B}=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$,则

$$f(\boldsymbol{x})=\frac{1}{2\pi|\boldsymbol{B}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right\}, \tag{3.11}$$

记为$(X,Y)\sim N(\boldsymbol{\mu},\boldsymbol{B})$.这种记法可以很容易地推广到$n$维正态分布.

下面来看二维正态分布的边缘分布,设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,令

$$u=\frac{x-\mu_1}{\sigma_1},\quad v=\frac{y-\mu_2}{\sigma_2},$$

(X,Y)关于X的边缘概率密度

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[(v-\rho u)^2+(1-\rho^2)u^2]\right\}\mathrm{d}v \\ &= \frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{u^2}{2}}\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right\}\mathrm{d}v \\ &= \frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{u^2}{2}} = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}, \end{aligned}$$

所以 $X\sim N(\mu_1,\sigma_1^2)$,同理 $Y\sim N(\mu_2,\sigma_2^2)$,可见多维正态分布的边缘分布仍为正态分布.

3.2 条件分布

3.2.1 条件分布

在第一章中我们知道,对于概率空间$(\Omega,\mathscr{F},P)$中任意给定的事件 B,$P(\cdot|B)$为$(\Omega,\mathscr{F})$上的一个新的概率.若 X 为一随机变量,我们可以定义其在这一新的概率下的分布函数 $F(x|B)=P(X\leqslant x|B)$,称之为 X 在事件 B 下的条件分布函数.特别地,对另一随机变量 Y,若取 $B=\{Y=y\}$,则称 $F(x|y)=F(x|Y=y)$为已知 $Y=y$ 时,X 的条件分布函数.

3.2.2 离散情形

对于离散型随机变量,可直接用条件概率计算.

定义 3.5 若(X,Y)的联合分布列为 $P(X=x_i,Y=y_j)=p_{ij}$,当 $p_{\cdot j}>0$ 时,称 $P(X=x_i|Y=y_j)=\dfrac{p_{ij}}{p_{\cdot j}}$为已知 $Y=y_j$ 时 X 的条件分布列,同样,称 $P(Y=y_j|X=x_i)=\dfrac{p_{ij}}{p_{i\cdot}}$为已知 $X=x_i$ 时 Y 的条件分布列.

例 3.6 若在一段时间内进入某商场的顾客数 $X\sim P(\lambda)$,每位顾客买东西的概率为 p,且他们独立地作出购买与否的决定,Y 表示买东西的顾客数,求在已知 $Y=m$ 的条件下 X 的分布列.

解 对 $n<m$,显然有 $P(X=n,Y=m)=0$,对 $n\geqslant m$,有

$$P(X=n,Y=m)=P(X=n)P(Y=m|X=n)=\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\mathrm{C}_n^m p^m(1-p)^{n-m},$$

$$P(Y=m)=\sum_{n=m}^{+\infty}\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\mathrm{C}_n^m p^m(1-p)^{n-m}=\mathrm{e}^{-\lambda p}\frac{(\lambda p)^m}{m!},\quad m=0,1,2,\cdots,$$

所以,在已知 $Y=m$ 的条件下 X 的分布列为:对 $n\geqslant m$,有

$$P(X=n|Y=m)=\frac{\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}\mathrm{C}_n^m p^m(1-p)^{n-m}}{\mathrm{e}^{-\lambda p}\frac{(\lambda p)^m}{m!}}=\mathrm{e}^{-\lambda(1-p)}\frac{(\lambda(1-p))^{(n-m)}}{(n-m)!},$$

对 $n<m$,显然有 $P(X=n|Y=m)=0$.

3.2.3 连续情形

对于连续型随机变量,由于对任意 y 都有 $P(Y=y)=0$,所以我们无法直接用条件概率来

计算条件分布函数 $F(x|y)=P(X\leqslant x|Y=y)$，但我们可以很自然地将它看做 $\Delta y\to 0$ 时 $P(X\leqslant x|y<Y\leqslant y+\Delta y)$ 的极限.

$$F(x|y)=P(X\leqslant x|Y=y)=\lim_{\Delta y\to 0}P(X\leqslant x|y<Y\leqslant y+\Delta y)$$

$$=\lim_{\Delta y\to 0}\frac{P(X\leqslant x,y<Y\leqslant y+\Delta y)}{P(y<Y\leqslant y+\Delta y)}=\lim_{\Delta y\to 0}\frac{F(x,y+\Delta y)-F(x,y)}{F_Y(y+\Delta y)-F_Y(y)}=\frac{\frac{\partial}{\partial y}F(x,y)}{f_Y(y)}.$$

对 x 求导，有 $\frac{\partial}{\partial x}F(x|y)=\frac{\frac{\partial^2}{\partial x\partial y}F(x,y)}{f_Y(y)}=\frac{f(x,y)}{f_Y(y)}$，所以我们有如下定义.

定义 3.6　若(X,Y)为二维连续型随机变量，称 $f(x|y)=\frac{f(x,y)}{f_Y(y)}$ 为已知 $Y=y$ 时 X 的条件概率密度. 同理，称 $f(y|x)=\frac{f(x,y)}{f_X(x)}$ 为已知 $X=x$ 时 Y 的条件概率密度.

例 3.7　若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则有

$$f(x|y)=\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_1}\exp\left\{-\frac{\left[x-\left(\mu_1+\frac{\sigma_1}{\sigma_2}\rho(y-\mu_2)\right)\right]^2}{2(1-\rho^2)\sigma_1^2}\right\}.$$

所以在已知 $Y=y$ 的条件下，X 服从正态分布 $N\left(\mu_1+\frac{\sigma_1}{\sigma_2}\rho(y-\mu_2),(1-\rho^2)\sigma_1^2\right)$. 同理，在已知 $X=x$ 的条件下，Y 服从正态分布 $N\left(\mu_2+\frac{\sigma_2}{\sigma_1}\rho(x-\mu_1),(1-\rho^2)\sigma_2^2\right)$.

可见，多维正态分布的条件分布仍为正态分布.

由定义 3.6，有 $f(x,y)=f_X(x)f(y|x)$ 或 $f(x,y)=f_Y(y)f(x|y)$，进一步，有

$$f_Y(y)=\int_{-\infty}^{+\infty}f_X(x)f(y\mid x)\mathrm{d}x, \tag{3.12}$$

及

$$f_X(x)=\int_{-\infty}^{+\infty}f_Y(y)f(x\mid y)\mathrm{d}y, \tag{3.13}$$

此为连续情形下的全概率公式.

例 3.8　设 $X\sim N(0,1)$，在给定 $X=x$ 的条件下，$Y\sim N(\rho x,1-\rho^2)$，求在给定 $Y=y$ 的条件下 X 的分布.

解　(X,Y)的联合概率密度

$$f(x,y)=f_X(x)f(y|x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$

$$=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[x^2-2\rho xy+y^2]\right\},$$

所以$(X,Y)\sim N(0,0,1,1,\rho)$，由例 3.7 知，在给定 $Y=y$ 条件下 $X\sim N(\rho y,1-\rho^2)$.

3.3　随机变量的独立性

对于多维随机变量，各个分量的取值有时会相互影响(如例 3.3 中无放回情形)，而有时也可能没有影响(如例 3.3 中有放回情形). 当随机变量取值的统计规律相互之间没有影响时，就称它们是相互独立的.

定义 3.7 设$(X_1,X_2,\cdots,X_n)$为 n 维随机变量,如果对任意 $B_1,B_2,\cdots,B_n\subset\mathbf{R}$,且对 $i=1,2,\cdots,n$,$\{X_i\in B_i\}\in\mathscr{F}$,都有

$$P(X_1\in B_1,X_2\in B_2,\cdots,X_n\in B_n)=\prod_{i=1}^{n}P(X_i\in B_i),$$

或等价地有联合分布函数 $F(x_1,x_2,\cdots,x_n)=\prod\limits_{i=1}^{n}F_{X_i}(x_i)$,则称 $X_1,X_2,\cdots,X_n$ 相互独立.

更一般地,对于一族随机变量,若其中任意有限个相互独立,则称这一族随机变量相互独立.

下面的结果给出随机变量独立性的一些性质及判别方法.

定理 3.3 (1) 若 $X_1,X_2,\cdots,X_n$ 相互独立,则其中任意 $k(k=2,3,\cdots,n)$个随机变量相互独立;

(2) 若$(X_1,X_2,\cdots,X_n)$为离散型随机变量,则 $X_1,X_2,\cdots,X_n$ 相互独立当且仅当对于任意 n 个可能的取值 $x_1,x_2,\cdots,x_n$,有

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i),\tag{3.14}$$

即联合分布列等于边缘分布列的乘积;

(3) 若$(X_1,X_2,\cdots,X_n)$为连续型随机变量,则 $X_1,X_2,\cdots,X_n$ 相互独立当且仅当对于任意 $x_1,x_2,\cdots,x_n$,有

$$f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f_{X_i}(x_i),\tag{3.15}$$

即联合概率密度等于边缘概率密度的乘积;

(4) 若 $X_1,X_2,\cdots,X_n$ 相互独立,则对任意 n 个函数 $g_1,g_2,\cdots,g_n$,$g_1(X_1),g_2(X_2),\cdots,g_n(X_n)$相互独立.

例 3.9 对于二维正态分布$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,由定理 3.3 中的(3)及(3.10)式可见 X,Y 独立的充分必要条件是 $\rho=0$.

例 3.10 设三维随机变量(X,Y,Z)的联合概率密度

$$f(x,y,z)=\begin{cases}\dfrac{1}{8\pi^3}(1-\sin x\sin y\sin z), & 0\leqslant x,y,z\leqslant 2\pi,\\ 0, & \text{其他},\end{cases}$$

则(X,Y)的联合概率密度

$$f(x,y)=\int_{-\infty}^{+\infty}f(x,y,z)\mathrm{d}z=\begin{cases}\dfrac{1}{4\pi^2}, & 0\leqslant x,y\leqslant 2\pi,\\ 0, & \text{其他},\end{cases}$$

X 的密度

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\dfrac{1}{2\pi}, & 0\leqslant x\leqslant 2\pi,\\ 0, & \text{其他}.\end{cases}$$

可见,X,Y,Z 两两独立,但不相互独立.

3.4 多维随机变量函数的分布

在实际应用中,常常会遇到求多维随机变量函数的分布问题.例如,若已知平面上随机点

(X,Y)的分布,求其到原点的距离 $Z=\sqrt{X^2+Y^2}$的分布或更一般地求其极坐标(ρ,θ)的分布.

3.4.1　多维离散情形

对于多维离散型随机变量,其函数依然为离散型随机变量,因此只需计算其分布列.

例 3.11　若 X,Y 独立,且 $X\sim B(m,p),Y\sim B(n,p)$,求 $Z=X+Y$ 的分布.

解　显然 Z 取从 0 到 $m+n$ 的整数,对其中的任意 k,有

$$\begin{aligned}P(Z=k)&=\sum_{l=0}^{k}P(X=l,Y=k-l)=\sum_{l=0}^{k}P(X=l)P(Y=k-l)\\&=\sum_{l=0}^{k}\mathrm{C}_m^l p^l q^{m-l}\mathrm{C}_n^{k-l}p^{k-l}q^{n-(k-l)}=p^k q^{m+n-k}\sum_{l=0}^{k}\mathrm{C}_m^l\mathrm{C}_n^{k-l}\\&=\mathrm{C}_{m+n}^k p^k q^{m+n-k},\end{aligned}$$

所以 $X+Y\sim B(m+n,p)$. 这一点从二项分布的含义很容易看出来.

3.4.2　多维连续情形

对于多维连续型随机变量$(X_1,X_2,\cdots,X_n)$及一个 n 元函数 g,若 $Y=g(X_1,X_2,\cdots,X_n)$,我们一般先求 Y 的分布函数,若 Y 为一连续型随机变量,再求导得其概率密度.

例 3.12　若 X,Y 独立同分布于$E(\lambda)$,$Z=\dfrac{X}{X+Y}$,求 Z 的分布.

解　显然 $0\leqslant Z\leqslant 1$,所以当 $0\leqslant z\leqslant 1$ 时,Z 的分布函数

$$\begin{aligned}F_Z(z)&=P\left(\frac{X}{X+Y}\leqslant z\right)=\iint\limits_{\frac{x}{x+y}\leqslant z,x>0,y>0}\lambda^2\mathrm{e}^{-\lambda(x+y)}\,\mathrm{d}x\mathrm{d}y\\&=\int_0^{+\infty}\mathrm{d}x\int_{\frac{x(1-z)}{z}}^{+\infty}\lambda^2\mathrm{e}^{-\lambda(x+y)}\,\mathrm{d}y=z,\end{aligned}$$

所以 $Z\sim U(0,1)$.

下面我们来考察几个常用的特殊函数的情形.

1. 和的分布

设(X,Y)的联合密度为 $f(x,y)$,$Z=X+Y$,积分区域如图 3.7 所示,则 Z 的分布函数

$$\begin{aligned}F_Z(z)&=\iint\limits_{x+y\leqslant z}f(x,y)\,\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z-x}f(x,y)\,\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\mathrm{d}y\int_{-\infty}^{z-y}f(x,y)\,\mathrm{d}x,\end{aligned}$$

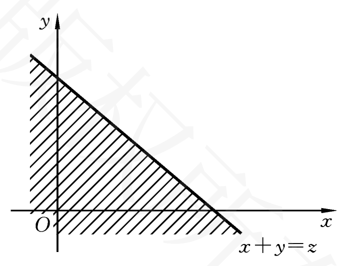

图 3.7

所以 Z 为连续型随机变量,其概率密度

$$f_Z(z)=F'(z)=\int_{-\infty}^{+\infty}f(x,z-x)\,\mathrm{d}x=\int_{-\infty}^{+\infty}f(z-y,y)\,\mathrm{d}y.\tag{3.16}$$

特别地,若 X,Y 独立,则

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\,\mathrm{d}x=\int_{-\infty}^{+\infty}f_X(z-y)f_Y(y)\,\mathrm{d}y,\tag{3.17}$$

可见,Z 的概率密度为 f_X 与 f_Y 的卷积.

例 3.13　若 X,Y 独立同分布于 $N(0,1)$,则 $Z=X+Y$ 的概率密度

$$f_Z(z)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left\{-\frac{x^2}{2}-\frac{(z-x)^2}{2}\right\}\mathrm{d}x=\frac{1}{2\pi}\exp\left[-\frac{z^2}{4}\right]\int_{-\infty}^{+\infty}\exp\left[-\left(x-\frac{z}{2}\right)^2\right]\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left[-\frac{z^2}{2\times 2}\right],$$

可见 $Z \sim N(0,2)$.

一般地,若 $X_1,\cdots,X_n$ 相互独立,且 $X_i \sim N(\mu_i,\sigma_i^2)$,则

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

例 3.14 若 X,Y 独立,且 $X \sim U(0,1)$,Y 的概率密度

$$f_Y(y) = \begin{cases} 2y, & 0 \leqslant y \leqslant 1, \\ 0, & \text{其他}, \end{cases}$$

求 $Z = X+Y$ 的概率密度.

解 由卷积公式(3.17)知,要使被积函数 $f_X(x)f_Y(z-x) \neq 0$,必须 $0 \leqslant x \leqslant 1, 0 \leqslant z-x \leqslant 1$,所以,

对于 $z<0$ 或 $z>2$,有 $\qquad f_Z(z) = 0$;

对于 $0 \leqslant z \leqslant 1$,有 $\qquad f_Z(z) = \int_0^z 2(z-x)\mathrm{d}x = z^2$;

对于 $1 \leqslant z \leqslant 2$,有 $\qquad f_Z(z) = \int_{z-1}^{1} 2(z-x)\mathrm{d}x = 2z - z^2$.

2. 商的分布

设(X,Y)的联合密度为 $f(x,y)$,$Z = \dfrac{X}{Y}$,积分区域如图 3.8 所示,则 Z 的分布函数为

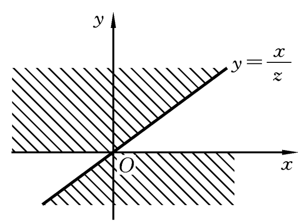

图 3.8

$$\begin{aligned} F_Z(z) &= \iint\limits_{\frac{x}{y} \leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \iint\limits_{y>0, x \leqslant yz} f(x,y)\mathrm{d}x\mathrm{d}y + \iint\limits_{y<0, x \geqslant yz} f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_0^{+\infty} \mathrm{d}y \int_{-\infty}^{yz} f(x,y)\mathrm{d}x + \int_{-\infty}^{0} \mathrm{d}y \int_{yz}^{+\infty} f(x,y)\mathrm{d}x, \end{aligned}$$

所以 Z 为连续型随机变量,其概率密度

$$\begin{aligned} f_Z(z) = F'(z) &= \int_0^{+\infty} y f(yz,y)\mathrm{d}y + \int_{-\infty}^{0} (-y) f(yz,y)\mathrm{d}y \\ &= \int_{-\infty}^{+\infty} |y| f(yz,y)\mathrm{d}y. \end{aligned} \tag{3.18}$$

例 3.15 若 X,Y 独立同分布,且 X 服从 $E(1)$,则 $Z = \dfrac{X}{Y}$ 的概率密度

$$f_Z(z) = \int_0^{+\infty} y\mathrm{e}^{-(yz+y)}\mathrm{d}y = \frac{1}{(1+z)^2},\ z > 0.$$

3. 最大最小值分布

若 $X_1,X_2,\cdots,X_n$ 独立同分布,X_1 的分布函数与概率密度分别为 $F(x)$,$f(x)$,令 $Y = \max\{X_1,X_2,\cdots,X_n\}$,$Z = \min\{X_1,X_2,\cdots,X_n\}$,则 Y,Z 的分布函数分别为

$$\begin{aligned} F_Y(y) &= P(\max\{X_1,X_2,\cdots,X_n\} \leqslant y) = P(X_1 \leqslant y, X_2 \leqslant y, \cdots, X_n \leqslant y) \\ &= \prod_{i=1}^{n} P(X_i \leqslant y) = (F(y))^n, \end{aligned}$$

$$\begin{aligned}F_Z(z) &= P(\min\{X_1,X_2,\cdots,X_n\}\leqslant z) = 1-P(\min\{X_1,X_2,\cdots,X_n\}>z)\\ &= 1-P(X_1>z,X_2>z,\cdots,X_n>z)\\ &= 1-\prod_{i=1}^{n}P(X_i>z) = 1-(1-F(z))^n.\end{aligned}$$

它们的概率密度分别为

$$f_Y(y)=n(F(y))^{n-1}f(y), \tag{3.19}$$

$$f_Z(z)=n(1-F(z))^{n-1}f(z). \tag{3.20}$$

例 3.16 系统装有 n 个独立工作的同样的电子元件，元件寿命服从 $E(\lambda)$，若这 n 个元件采用(1) 并联；(2) 串联的使用方法用于系统，求系统寿命的分布.

解 (1) 并联时，系统寿命为 n 个独立工作的同样的电子元件寿命的最大值，当 $x\geqslant 0$ 时，其密度 $f_{\max}(x)=n(1-\mathrm{e}^{-\lambda x})^{n-1}\lambda\mathrm{e}^{-\lambda x}=n\lambda(1-\mathrm{e}^{-\lambda x})^{n-1}\mathrm{e}^{-\lambda x}$.

(2) 串联时，系统寿命为 n 个独立工作的同样的电子元件寿命的最小值，当$x\geqslant 0$ 时，其密度 $f_{\min}(x)=n(\mathrm{e}^{-\lambda x})^{n-1}\lambda\mathrm{e}^{-\lambda x}=n\lambda\mathrm{e}^{-n\lambda x}$，服从参数为 $n\lambda$ 的指数分布.

3.4.3 一般情形

最后，我们通过两个例子来说明更一般的情形.

例 3.17 若 $X\sim N(\mu,\sigma^2)$，Y 的分布列为 $\begin{array}{c|cc} Y & -1 & 1 \\ \hline P & 1/3 & 2/3\end{array}$，且 X,Y 独立，求 $Z=XY$ 的分布.

解 由全概率公式知，Z 的分布函数

$$\begin{aligned}F_Z(z)&=P(XY\leqslant z)\\ &=P(Y=-1)P(XY\leqslant z|Y=-1)+P(Y=1)P(XY\leqslant z|Y=1)\\ &=\frac{1}{3}P(-X\leqslant z|Y=-1)+\frac{2}{3}P(X\leqslant z|Y=1)=\frac{1}{3}P(X\geqslant -z)+\frac{2}{3}P(X\leqslant z)\\ &=\frac{1}{3}\left(1-\Phi\left(\frac{-z-\mu}{\sigma}\right)\right)+\frac{2}{3}\Phi\left(\frac{z-\mu}{\sigma}\right)=\frac{1}{3}\Phi\left(\frac{z+\mu}{\sigma}\right)+\frac{2}{3}\Phi\left(\frac{z-\mu}{\sigma}\right)\end{aligned}$$

所以 Z 为连续型随机变量，其概率密度

$$f_Z(z)=F'(z)=\frac{1}{3\sqrt{2\pi}\sigma}\exp\left\{-\frac{(z+\mu)^2}{2\sigma^2}\right\}+\frac{2}{3\sqrt{2\pi}\sigma}\exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\}.$$

例 3.18 若 $X_1,X_2,\cdots,X_n$ 独立同分布，分布函数与概率密度分别为 $F(x)$，$f(x)$，令 $Y=\max\{X_1,X_2,\cdots,X_n\}$，$Z=\min\{X_1,X_2,\cdots,X_n\}$，求$(Y,Z)$的联合密度.

解 当 $y<z$ 时，显然有 $f(y,z)=0$；当 $y\geqslant z$ 时，(Y,Z)的联合分布函数

$$\begin{aligned}F(y,z)&=P(\max\{X_1,X_2,\cdots,X_n\}\leqslant y,\min\{X_1,X_2,\cdots,X_n\}\leqslant z)\\ &=P(\max\{X_1,X_2,\cdots,X_n\}\leqslant y)-P(\max\{X_1,X_2,\cdots,X_n\}\leqslant y,\min\{X_1,X_2,\cdots,X_n\}>z)\\ &=P(X_1\leqslant y,X_2\leqslant y,\cdots,X_n\leqslant y)-P(z<X_1\leqslant y,z<X_2\leqslant y,\cdots,z<X_n\leqslant y)\\ &=(F(y))^n-(F(y)-F(z))^n,\end{aligned}$$

所以，(Y,Z)的概率密度

$$f(y,z)=\frac{\partial^2F(y,z)}{\partial y\partial z}=n(n-1)(F(y)-F(z))^{n-2}f(y)f(z).$$

习　题　三

3.1　选择题.

(1) 设(X,Y)的联合密度为$f(x,y)$,则$P(X>1)=$(　　).

(A) $\int_{-\infty}^{1} f(x,y)\mathrm{d}x$　　(B) $\int_{1}^{+\infty} f(x,y)\mathrm{d}x$

(C) $\int_{-\infty}^{1}\mathrm{d}x\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$　　(D) $\int_{1}^{+\infty}\mathrm{d}x\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$

(2) 若随机变量X,Y独立,其分布列分别为

X	-1	1
P	$1/3$	$2/3$

,

Y	-1	1
P	$1/3$	$2/3$

,

则下列正确的是(　　).

(A) $P(X=Y)=\dfrac{2}{3}$　(B) $P(X=Y)=1$　(C) $P(X=Y)=\dfrac{1}{2}$　(D) $P(X=Y)=\dfrac{5}{9}$

(3) 设X,Y是两个独立的均匀分布,则下列随机变量中服从均匀分布的是(　　).

(A) $X+Y$　(B) $X-Y$　(C) XY　(D) (X,Y)

(4) 设二维随机变量(X,Y)的分布列为

X \ Y	0	1
0	0.4	a
1	b	0.1

已知事件$\{X=0\}$与$\{X+Y=1\}$相互独立,则(　　).

(A) $a=0.2,b=0.3$　(B) $a=0.4,b=0.1$　(C) $a=0.3,b=0.2$　(D) $a=0.1,b=0.4$

(5) 边缘分布为正态分布的二维随机变量的分布(　　).

(A) 必为二维正态分布　　(B) 必为均匀分布

(C) 不必为二维正态分布　　(D) 由这两个边缘分布决定

3.2　填空题.

(1) 设$(X,Y)\sim N(1,-1,2^2,3^2,0)$,则$(X,Y)$的概率密度为______,且$3X-2Y\sim$______.

(2) 某电子元件的使用寿命(单位:h)服从参数为$1/a$的指数分布,则两个元件一个坏了后,接着使用第二个,一共用了不足$2a$(h)的概率为______.

(3) 设X,Y为两个随机变量,且$P(X\geqslant 0)=P(Y\geqslant 0)=\dfrac{4}{7}$,$P(X\geqslant 0,Y\geqslant 0)=\dfrac{3}{7}$,则$P(\max(X,Y)\geqslant 0)=$______.

(4) 从1,2,3,4中任取一数X,再从$1,2,\cdots,X$任取一数Y,则$P(X=Y)=$______.

(5) 设平面区域D由曲线$y=\dfrac{1}{x}$及直线$y=0,x=1,x=\mathrm{e}^2$所围成,(X,Y)服从区域D上的均匀分布,则(X,Y)关于X的边缘分布函数在$x=2$处的值为______.

3.3　二元函数

$$F(x,y)=\begin{cases}0, & x+y<0,\\ 1, & x+y>0\end{cases}$$

是不是某二维随机变量的分布函数?若是,求边缘分布函数;若不是,说明理由.

3.4　将一硬币连掷3次,以X表示3次投掷中出现正面的次数,以Y表示3次投掷中正反面次数之差的绝对值,求(X,Y)的概率分布.

3.5　袋中有 5 个白球、3 个红球、2 个黑球，若从中任取 5 个球，设其中白球个数为 X，红球个数为 Y，求 (X,Y) 的概率分布.

3.6　设随机变量 (X,Y) 的分布函数为

$$F(x,y)=a\left(b+\arctan\frac{x}{2}\right)\left(c+\arctan\frac{y}{2}\right),$$

求常数 a,b,c 及 (X,Y) 的概率密度.

3.7　设 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}C(R-\sqrt{x^2+y^2}), & x^2+y^2<R^2,\\ 0, & x^2+y^2\geqslant R^2,\end{cases}$$

求：(1) 系数 C 的值；(2) (X,Y) 落在 $x^2+y^2<r^2\ (r<R)$ 内的概率.

3.8　设随机变量 $X\sim U(-2,2)$，定义

$$Y=\begin{cases}-1, & X\leqslant -1,\\ 1, & X>-1\end{cases}\quad 及\quad Z=\begin{cases}-1, & X\leqslant 1,\\ 1, & X>1,\end{cases}$$

求 (Y,Z) 的概率分布.

3.9　设随机变量 (X,Y) 具有下列概率密度，求边缘概率密度 $f_X(x)$，$f_Y(y)$ 与条件概率密度 $f(y|x)$，$f(x|y)$.

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

(2) $f(x,y)=\begin{cases}1, & 0\leqslant x\leqslant 2,\max\{0,x-1\}\leqslant y\leqslant \min\{1,x\},\\ 0, & 其他.\end{cases}$

3.10　设随机变量 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}3x, & 0<x<1,0<y<x,\\ 0, & 其他,\end{cases}$$

求 $P\left(Y<\dfrac{1}{8}\,\middle|\,X=\dfrac{1}{4}\right)$.

3.11　设随机变量 (X,Y) 的概率分布为

X \ Y	0	1	
-1	$\frac{1}{10}$	α	
0	$\frac{1}{15}$	$\frac{4}{15}$	
1	β	$\frac{2}{15}$	

问 α,β 为何值时，X 与 Y 才能相互独立？

3.12　若 X,Y 独立，且 $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$，(1) 证明 $X+Y\sim P(\lambda_1+\lambda_2)$；(2) 求在已知 $X+Y=m$ 的条件下，X 的分布.

3.13　设随机变量 $X\sim U(0,1)$，且在 $X=x$ 的条件下，随机变量 Y 服从区间 $(0,x)$ 内的均匀分布，求：(1) (X,Y) 的概率密度；(2) Y 的概率密度；(3) 概率 $P(X+Y>1)$.

3.14　若 X_1,X_2,X_3,X_4 独立同分布于 $B(1,p)$，求 $X=\begin{vmatrix}X_1 & X_2\\ X_3 & X_4\end{vmatrix}$ 的概率分布.

3.15　若 X,Y 独立，且 $X\sim E(\lambda_1)$，$Y\sim E(\lambda_2)$，求 $X+Y$ 的概率密度.

3.16　设随机变量 (X,Y) 服从矩形 $D=\{(x,y)|0\leqslant x\leqslant 2,0\leqslant y\leqslant 1\}$ 上的均匀分布，求边长为 X 和 Y 的矩形周长 L 及面积 S 的概率密度.

3.17 若 X,Y 独立同分布于 $N(0,1)$,求 $Z=\dfrac{X}{Y}$ 的概率密度.

3.18 若 $X\sim U(0,1)$,Y 的概率分布为 $\begin{array}{c|ccc} Y & -1 & 0 & 1 \\ \hline P & 1/2 & 1/3 & 1/6 \end{array}$,且 X,Y 独立,求 $Z=X+Y$ 的概率分布.

3.19 在线段(0,1)内任取两点,求两点间距离的分布函数与概率密度.

第四章　数 字 特 征

随机变量的数字特征是指与随机变量分布有关的某些数值.随机变量的分布函数是随机变量最重要的概率特征,它完全决定随机变量的概率性质和其他一切特征,其中包括数字特征.但是需要指出下列事实.

(1) 虽然概率分布能够完全描述随机变量的概率规律,但是往往不能明显而集中地表现随机变量的某些特点.例如,它取值的集中位置和分散程度等.

(2) 在实际应用中,人们往往并不知道随机变量的确切分布,但就应用而言,通常也没有必要知道它的一切概率特征,只需要知道它的某些数字特征.例如,水稻的平均亩产量、股票的平均收益率等.

(3) 有些重要分布(如泊松分布、正态分布等)的数学形式是已知的,但是它的分布由某些参数决定,而这些参数本身往往是随机变量重要数字特征的函数,并且有明显的概率意义.

总之,随机变量的数字特征在概率论与数理统计的理论研究和实际应用中占有重要地位.

4.1　随机变量的数学期望

“期望”这个术语首次出现在惠更斯的主要著作《论赌博中的计算》中,基于这个术语解决了一些当时感兴趣的博弈问题,他在这部著作中提出了 14 条命题,第一条命题是:如果某人在赌博中以概率$\frac{1}{2}$赢 a 元,以概率$\frac{1}{2}$输 b 元,则他收益的期望是$\frac{a-b}{2}$元.由此现实问题可导出随机变量最重要的数字特征,即随机变量的数学期望.

4.1.1　离散型随机变量的数学期望

定义 4.1　设离散型随机变量 X 的概率分布为 $P(X=x_i)=p_i(i=1,2,\cdots)$,若 $\sum_{i=1}^{\infty}|x_i|p_i<+\infty$,则称$\sum_{i=1}^{\infty}x_ip_i$ 为 X 的**数学期望**,记为 EX,简称为**期望**或**均值**,即

$$EX=\sum_{i=1}^{\infty}x_ip_i. \tag{4.1}$$

在数学期望的定义中,级数 $\sum_i x_ip_i$ 的和不应该依赖于各被加项的顺序,而只有当级数绝对收敛时(即$\sum_i |x_i|p_i<+\infty$),级数$\sum_i x_ip_i$ 的和才与项的顺序无关,这就是在期望的定义中要求级数(4.1)绝对收敛的原因.若级数$\sum_i |x_i|p_i$ 发散,则称此随机变量的数学期望不存在.

随机变量的数学期望是一个常数,它刻画了随机变量取值的平均值.直观地,期望可以看

做是随机变量的可能取值以概率为权重的加权平均.

例 4.1 设随机变量 $X\sim B(n,p)$,求 EX.

解 由定义 4.1 有

$$\begin{aligned}EX &= \sum_{k=0}^{n} x_k p_k = \sum_{k=0}^{n} k\mathrm{C}_n^k p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &= np\sum_{k=1}^{n}\mathrm{C}_{n-1}^{k-1}p^{k-1}(1-p)^{n-k} = np\sum_{k=0}^{n-1}\mathrm{C}_{n-1}^{k}p^k(1-p)^{n-1-k}\\ &= np[p+(1-p)]^{n-1} = np,\end{aligned}$$

即 $EX=np$.

例 4.2 设随机变量 $X\sim P(\lambda)$,求 EX.

解 由定义 4.1 有

$$EX = \sum_{k=0}^{\infty} x_k p_k = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda,$$

即 $EX=\lambda$.

例 4.3 某人的一串钥匙共有 n 把,其中只有一把能开家中房门,他随意地试用这些钥匙,分别按如下两种情形求试用次数的数学期望:(1) 把每次试用过的钥匙分开;(2) 把每次试用过的钥匙又混杂进去.

解 设 X 表示某人试用的次数.

(1) 经过简单计算,X 有如下的分布列

X	1	2	3	…	n
P	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	…	$\frac{1}{n}$

$$EX=\frac{1}{n}(1+2+\cdots+n)=\frac{n+1}{2}.$$

(2) 根据第二章的知识,可知 X 服从几何分布,即有如下的概率分布

$$P(X=k)=\left(1-\frac{1}{n}\right)^{k-1}\cdot\frac{1}{n},\ k=1,2,\cdots.$$

由定义 4.1 有 $$EX = \sum_{k=1}^{\infty}k\left(1-\frac{1}{n}\right)^{k-1}\cdot\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{\infty}k\left(1-\frac{1}{n}\right)^{k-1}.$$

令 $q=1-\frac{1}{n}$,则

$$\sum_{k=1}^{\infty}kq^{k-1} = \sum_{k=1}^{\infty}\frac{\mathrm{d}q^k}{\mathrm{d}q} = \frac{\mathrm{d}\sum_{k=1}^{\infty}q^k}{\mathrm{d}q} = \frac{\mathrm{d}\frac{q}{1-q}}{\mathrm{d}q} = \frac{1}{(1-q)^2},$$

所以 $$EX=\frac{1}{n}\times n^2=n.$$

例 4.4 设 A 是事件,I_A 是 A 的示性函数,则 I_A 服从两点分布,求 EI_A.

解 由 I_A 的定义可知其有如下分布列

I_A	0	1
P	$1-P(A)$	$P(A)$

所以 $$E(I_A)=1\times P(A)=P(A).$$

例 4.5　若随机变量 X 的概率分布由下式给定：

$$P\left(X=(-1)^{i+1}\cdot\frac{3^i}{i}\right)=\frac{2}{3^i},i=1,2,\cdots,$$

求 EX.

解　由于 $\sum\limits_{i=1}^{\infty}|x_i|p_i=\sum\limits_{i=1}^{\infty}\frac{2}{i}=\infty$，可见级数发散，故 EX 不存在.

4.1.2　连续型随机变量的数学期望

设 X 有概率密度 $f(x)$，我们将 X 取值的区间分为若干个小区间，其分点为 $x_0<x_1<\cdots<x_n$，n 越大分点越密. 由概率密度的定义可知，X 的取值落在小区间 (x_i,x_{i+1}) 的概率近似地等于 $f(x_i)\Delta x_i$，其中 $\Delta x_i=x_{i+1}-x_i$. 因此，X 与以概率 $f(x_i)\Delta x_i$ 取值 x_i 的离散型随机变量近似. 由此，X 的数学期望近似为

$$\sum_i x_i f(x_i)\Delta x_i,$$

它正是积分 $\int_{-\infty}^{+\infty}xf(x)$ 的渐近和式. 根据直观分析，我们有如下定义.

定义 4.2　设 X 为连续型随机变量，其概率密度为 $f(x)$，若积分 $\int_{-\infty}^{+\infty}|x|f(x)\mathrm{d}x<+\infty$，则称积分值 $\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$ 为 X 的**数学期望**，记为 EX，简称**期望**或**均值**，即

$$EX=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x. \tag{4.2}$$

例 4.6　设随机变量 $X\sim E(\lambda)$，求 EX.

解　由定义 4.2 有

$$EX=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\int_0^{+\infty}x\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x=-\int_0^{+\infty}x\mathrm{d}\mathrm{e}^{-\lambda x}=\int_0^{+\infty}\mathrm{e}^{-\lambda x}\mathrm{d}x=\frac{1}{\lambda},$$

即 $EX=\frac{1}{\lambda}$.

例 4.7　设随机变量 $X\sim N(\mu,\sigma^2)$，求 EX.

解　由定义 4.2 有

$$\begin{aligned}EX&=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}x\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma t+\mu)\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sigma t\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t+\frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\mu,\end{aligned}$$

即 $EX=\mu$.

例 4.8　若随机变量 X 的概率密度为

$$f(x)=\frac{1}{\pi(1+x^2)},\ -\infty<x<+\infty,$$

则称其为柯西分布，求 EX.

解　由于

$$\begin{aligned}\int_{-\infty}^{+\infty}|x|f(x)\mathrm{d}x&=\int_{-\infty}^{+\infty}|x|\frac{1}{\pi(1+x^2)}\mathrm{d}x=\frac{2}{\pi}\int_0^{+\infty}\frac{x}{1+x^2}\mathrm{d}x=\frac{1}{\pi}\int_0^{+\infty}\frac{\mathrm{d}(1+x^2)}{1+x^2}\\&=\frac{1}{\pi}\ln(1+x^2)\Big|_0^{+\infty}=\lim_{x\to+\infty}\frac{1}{\pi}\ln(1+x^2),\end{aligned}$$

可见积分为 $+\infty$,故 EX 不存在.

例 4.9　设 X 的数学期望存在,概率密度 $f(x)$ 关于 $x=\mu$ 对称,即 $f(\mu+x)=f(\mu-x)$,则 $EX=\mu$.

证明　由定义 4.2 有

$$
\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_{-\infty}^{+\infty}(x-\mu+\mu)f(x)\mathrm{d}x = \int_{-\infty}^{+\infty}(x-\mu)f(x)\mathrm{d}x + \int_{-\infty}^{+\infty}\mu f(x)\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} t f(t+\mu)\mathrm{d}t + \mu,
\end{aligned}
$$

由于 $tf(t+\mu)$ 是奇函数,故 $EX=\mu$.

4.1.3　随机变量函数的数学期望

在理论研究和实际应用中经常遇到要求随机变量函数的期望问题.例如,圆的半径是一个随机变量,圆的面积是半径的函数,要求圆的平均面积就是要求随机变量函数的期望.

例 4.10　在一场赌博中,某人同时掷 3 枚硬币,如果看见 3 枚硬币同时正面向上或同时正面向下,则他可以获得 5 元;如果看见 1 枚或 2 枚硬币正面向上,则他要付出 3 元,那么某人的平均收益是多大?

解　令 X 为 3 枚硬币中正面向上的个数,其分布列如下:

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

令 Y 为某人每次赌局所获得的收益,则有如下概率分布:

$$P(Y=5)=P(X=0\text{ 或 }X=3)=\frac{1}{8}+\frac{1}{8}=\frac{1}{4},$$

$$P(Y=-3)=P(X=1\text{ 或 }X=2)=\frac{3}{4}.$$

因此某人的平均收益为

$$EY=5\times\frac{1}{4}+(-3)\times\frac{3}{4}=-1.$$

在此例中,X 的概率分布是已知的,由 Y 与 X 的关系可以先求出 Y 的概率分布,然后根据期望的定义求出 EY.在很多问题中,已知 X 的概率分布,求 X 的函数即 Y 的分布是很困难的,因此我们将不加证明地给出解决这类问题的定理.

定理 4.1　设 $y=g(x)$ 是函数,而 $Y=g(X)$ 是随机变量 X 的函数.

(1) 若 X 是离散型随机变量,概率分布为

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

且级数 $\sum_{i=1}^{\infty}|g(x_i)|p_i$ 收敛,则

$$EY=E[g(X)]=\sum_{i=1}^{\infty}g(x_i)p_i. \tag{4.3}$$

(2) 若 X 是连续型随机变量,概率密度为 $f(x)$,且积分 $\int_{-\infty}^{+\infty}|g(x)|f(x)\mathrm{d}x<+\infty$,则

$$EY=E[g(X)]=\int_{-\infty}^{+\infty}g(x)f(x)\mathrm{d}x. \tag{4.4}$$

利用定理 4.1,不必先求出 Y 的概率分布,而是根据 X 的概率分布及函数 $g(x)$ 直接计算 EY.

例 4.11　假设 $X \sim P(\lambda)$,求 $E\left(\frac{1}{1+X}\right)$.

解　根据定理 4.1,有

$$E\left(\frac{1}{1+X}\right)=\sum_{k=0}^{\infty}\frac{1}{1+k}\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\frac{\mathrm{e}^{-\lambda}}{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!}=\frac{\mathrm{e}^{-\lambda}}{\lambda}(\mathrm{e}^{\lambda}-1)=\frac{1-\mathrm{e}^{-\lambda}}{\lambda}.$$

例 4.12　(混合型随机变量的数学期望)由例 2.11,求 EX.

解　根据定理 4.1,有

$$\begin{aligned}EX&=E\min\{V,V_0\}=\int_{-\infty}^{+\infty}\min\{x,V_0\}f_V(x)\mathrm{d}x=\int_{-\infty}^{V_0}xf_V(x)\mathrm{d}x+\int_{V_0}^{+\infty}V_0f_V(x)\mathrm{d}x\\&=\int_0^{V_0}x\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x+\int_{V_0}^{+\infty}V_0\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x=\frac{1}{\lambda}(1-\mathrm{e}^{-\lambda V_0}).\end{aligned}$$

当 Y 为多个随机变量 $X_1,X_2,\cdots,X_n(n\geqslant 2)$ 的函数时,有类似的结论.

定理 4.2　设 $z=g(x,y)$ 为函数,而 $Z=g(X,Y)$ 是随机变量 (X,Y) 的函数.

(1) 若 (X,Y) 是二维离散型随机变量,概率分布为

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

且级数 $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|g(x_i,y_j)|p_{ij}$ 收敛,则

$$EZ=E[g(X,Y)]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}g(x_i,y_j)p_{ij}. \tag{4.5}$$

(2) 若 (X,Y) 是二维连续型随机变量,概率密度为 $f(x,y)$,且积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|g(x,y)|f(x,y)\mathrm{d}x\mathrm{d}y<+\infty$,则

$$EZ=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y. \tag{4.6}$$

例 4.13　假设同时掷两个骰子,一个红色和一个白色.令 X 表示红色骰子出现的点数,Y 表示白色骰子出现的点数,求 $E(X+Y)$.

解　根据定理 4.2,有

$$E(X+Y)=\sum_{x=1}^{6}\sum_{y=1}^{6}(x+y)P(X=x,Y=y)=\sum_{x=1}^{6}\sum_{y=1}^{6}x\times\frac{1}{36}+\sum_{x=1}^{6}\sum_{y=1}^{6}y\times\frac{1}{36}=7.$$

例 4.14　设随机变量 X_1 和 X_2 相互独立,且都服从 $N(0,1)$,求 $E[\max\{X_1,X_2\}]$.

解　由(4.6)式及图 4.1 有

$$\begin{aligned}&E[\max\{X_1,X_2\}]\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\max\{x_1,x_2\}\cdot\exp\left[-\frac{x_1^2+x_2^2}{2}\right]\mathrm{d}x_1\mathrm{d}x_2\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left[-\frac{x_2^2}{2}\right]\mathrm{d}x_2\int_{x_2}^{+\infty}x_1\exp\left[-\frac{x_1^2}{2}\right]\mathrm{d}x_1\\&\quad+\frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left[-\frac{x_1^2}{2}\right]\mathrm{d}x_1\int_{x_1}^{+\infty}x_2\exp\left[-\frac{x_2^2}{2}\right]\mathrm{d}x_2\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-x_2^2}\mathrm{d}x_2+\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-x_1^2}\mathrm{d}x_1=\frac{1}{\pi}\sqrt{\pi}=\frac{1}{\sqrt{\pi}}.\end{aligned}$$

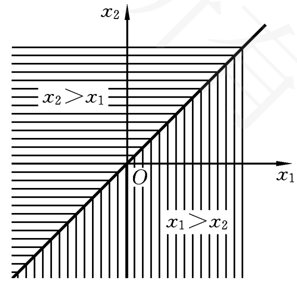

图 4.1

4.1.4　数学期望的性质

熟练掌握数学期望的性质,有助于计算随机变量的数学期望,在很多情况下,可大大简化计算.

定理 4.3　随机变量的数学期望具有如下性质:

(1) 设 C 为常数,则 $EC=C$;

(2) 对任意 $n\geqslant 1$ 及常数 $k_1,k_2,\cdots,k_n$,有

$$E\left(\sum_{i=1}^{n}k_iX_i\right)=\sum_{i=1}^{n}k_iEX_i;$$

(3) 当 $n\geqslant 2$ 时,若 $X_1,X_2,\cdots,X_n$ 相互独立,则

$$E\left(\prod_{i=1}^{n}X_i\right)=\prod_{i=1}^{n}EX_i;$$

(4) 若 $X\geqslant 0$,则 $EX\geqslant 0$;若 $X_1\geqslant X_2$,则

$$EX_1\geqslant EX_2;$$

(5) $|EX|\leqslant E|X|$;

(6) 柯西-施瓦茨(Cauchy-Schwarz)不等式:若 EX^2,EY^2 均存在,则 $E(XY)$ 存在,且 $[E(XY)]^2\leqslant EX^2EY^2$.

证明　(1) 由于 $P(X=C)=1$,从而 $EC=C$.

(2) 以连续型随机变量为例.设 X 的概率密度为 $f(x)$,则对任意常数 k,有

$$E(kX)=\int_{-\infty}^{+\infty}kxf(x)\mathrm{d}x=k\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=kEX.$$

当 $n=2$ 时,设(X_1,X_2)的概率密度为 $f(x_1,x_2)$,边缘概率密度为 $f_{X_1}(x_1),f_{X_2}(x_2)$,则对任意常数 k_1,k_2,有

$$\begin{aligned}E(k_1X_1+k_2X_2)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(k_1x_1+k_2x_2)f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\&=k_1\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x_1f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2+k_2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x_2f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\\&=k_1EX_1+k_2EX_2.\end{aligned}$$

对 $n\geqslant 3$ 的情形,可用数学归纳法证明结论成立.

(3) 以连续型随机变量为例.设 $n=2$,由于 X_1,X_2 相互独立,沿用性质(2)中记号,应用 $f(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)$,故

$$\begin{aligned}E(X_1X_2)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x_1x_2f(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x_1x_2f_{X_1}(x_1)f_{X_2}(x_2)\mathrm{d}x_1\mathrm{d}x_2\\&=\int_{-\infty}^{+\infty}x_1f_{X_1}(x_1)\mathrm{d}x_1\int_{-\infty}^{+\infty}x_2f_{X_2}(x_2)\mathrm{d}x_2=EX_1EX_2.\end{aligned}$$

对 $n\geqslant 3$ 的情形,可用数学归纳法证明.

(4) 当 $X\geqslant 0$ 时,由 EX 的定义即可推知 $EX\geqslant 0$.若 $X_1\geqslant X_2$,则 $X_1-X_2\geqslant 0$,从而 $E(X_1-X_2)\geqslant 0$,即 $EX_1-EX_2\geqslant 0$,故 $EX_1\geqslant EX_2$.

(5) 由于 $-|X|\leqslant X\leqslant |X|$,由性质(4)即知 $-E|X|\leqslant EX\leqslant E|X|$,从而 $|EX|\leqslant E|X|$.

(6) 考虑变量 t 的二次函数

$$g(t)=E(tX-Y)^2=t^2EX^2-2tE(XY)+EY^2,$$

对于一切 t,均有 $g(t)\geqslant 0$.令 $g(t)=0$,此方程或无实根,或只有一个二重根.故由一元二次方

程有实根的判定条件,有$[E(XY)]^2-EX^2EY^2\leqslant 0$,此即$[E(XY)]^2\leqslant EX^2EY^2$.

例 4.15　设随机变量 $X\sim B(n,p)$,求 EX.

解　令 X_i 表示第 i 次试验的结果,并令

$$X_i=\begin{cases}1, & \text{第 } i \text{ 次试验成功},\\ 0, & \text{第 } i \text{ 次试验不成功},\end{cases}$$

则 $EX_i=p$,$X=X_1+X_2+\cdots+X_n$ 是 n 次试验中成功的次数,则

$$EX=E(X_1+X_2+\cdots+X_n)=np.$$

例 4.16　设对某目标进行射击,每次发射一枚子弹,直到击中 n 次为止,设每次射击时击中目标的概率为 p $(0<p<1)$,且各次射击相互独立,试求子弹消耗量的数学期望.

解　令 X_i 为第 $i-1$ 次击中到第 i 次击中目标之间所消耗的子弹数,其分布列如下:

X_i	1	2	3	…	n	…
P	p	pq	q^2p	…	$q^{n-1}p$	…

其中 $q=1-p$,可见 X_i 服从几何分布. 令 $X=X_1+X_2+\cdots+X_n$ 表示击中 n 次目标所消耗的子弹数,则

$$EX=E(X_1+X_2+\cdots+X_n)=\frac{n}{p}.$$

上式说明,平均消耗子弹多少与击中概率 p 成反比,而这有很直观的意义.

4.2　随机变量的方差

在前一节中,我们引入了随机变量数学期望的概念,但在刻画随机变量的性质时,仅有数学期望是不够的. 例如,考察 A、B 两台测量仪器的测量误差,设随机变量 $X_A\sim U(-1,1)$,$X_B\sim U(-2,2)$,显然 $EX_A=EX_B$,两台仪器的平均测量误差是相同的,但 X_B 的取值不确定性要大于 X_A 的,为了反映出这种差异,我们引入随机变量方差的概念.

定义 4.3　设 X 为一随机变量,若 $E(X-EX)^2$ 存在,则称它为随机变量 X 的方差,记为 DX,即

$$DX=E(X-EX)^2. \tag{4.7}$$

方差的算术平方根 $\sqrt{DX}$ 称为 X 的**标准差**.

如果随机变量 X 是有度量单位的量,例如,米,那么方差的单位是平方米,而标准差具有与 X 相同的量纲. 方差刻画了随机变量分布的集中(或分散)程度,一个随机变量的方差越大,表示它取值越分散.

方差的定义式(4.7)可以看成是随机变量函数的期望,其函数取特殊形式 $g(x)=(x-EX)^2$,因此由定理 4.1,有如下公式:

若 X 是离散型随机变量,概率分布为 $P(X=x_i)=p_i$,$i=1,2,\cdots$,则有

$$DX=E(X-EX)^2=\sum_{i=1}^{\infty}(x_i-EX)^2p_i. \tag{4.8}$$

若 X 是连续型随机变量,概率密度为 $f(x)$,则有

$$DX=E(X-EX)^2=\int_{-\infty}^{+\infty}(x-EX)^2f(x)\,\mathrm{d}x. \tag{4.9}$$

经过简单推导,我们还可得到如下计算方差的常用公式:

$$DX = EX^2 - (EX)^2. \tag{4.10}$$

例 4.17 设 $X \sim B(n,p)$,求 DX.

解 由于 $EX = np$,所以

$$\begin{aligned} EX^2 &= \sum_{k=0}^{n} x_k^2 p_k = \sum_{k=0}^{n} k^2 C_n^k p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} (k-1) \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np[(n-1)p+1], \end{aligned}$$

记 $1-p=q$,由 $DX = EX^2 - (EX)^2$ 可得

$$DX = npq.$$

若 X 服从两点分布,则 $X \sim B(1,p)$,因此由此例可知 $DX = pq$.

例 4.18 设随机变量 $X \sim P(\lambda)$,求 DX.

解 $EX^2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{k=0}^{\infty} (k+1) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda(\lambda+1)$,

根据(4.10)式,有

$$DX = EX^2 - (EX)^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.$$

例 4.19 设随机变量 X 服从几何分布,求 DX.

解
$$EX^2 = \sum_{k=1}^{\infty} k^2 p q^{k-1} = p \sum_{k=1}^{\infty} k^2 q^{k-1},$$

而
$$\sum_{k=1}^{\infty} k^2 q^{k-1} = \frac{d}{dq}\Big(\sum_{k=1}^{\infty} k q^k\Big) = \frac{d}{dq}[q(1-q)^{-2}] = \frac{1+q}{(1-q)^3},$$

故 $EX^2 = \frac{1+q}{p^2}$. 根据(4.10)式,有

$$DX = EX^2 - (EX)^2 = \frac{1+q}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}.$$

例 4.20 设随机变量 $X \sim U(a,b)$,求 DX.

解 $EX^2 = \int_{-\infty}^{+\infty} g(x) f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$,

根据(4.10)式,有

$$DX = EX^2 - (EX)^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

例 4.21 设随机变量 $X \sim E(\lambda)$,求 DX.

解
$$\begin{aligned} EX^2 &= \int_{-\infty}^{+\infty} g(x) f(x) dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = -\int_0^{+\infty} x^2 de^{-\lambda x} \\ &= -x^2 e^{-\lambda x} \Big|_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x} dx = 2\int_0^{+\infty} x e^{-\lambda x} dx = \frac{2}{\lambda}\int_0^{+\infty} e^{-\lambda x} dx = \frac{2}{\lambda^2}, \end{aligned}$$

根据(4.10)式,有

$$DX = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

例 4.22 设随机变量 $X \sim N(\mu, \sigma^2)$,求 DX.

解 根据(4.9)式,有

$$DX = \int_{-\infty}^{+\infty} (x-\mu)^2 \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \sigma^2 \sqrt{2\pi} = \sigma^2.$$

由例 4.7 及例 4.22 可知，当随机变量 $X \sim N(\mu,\sigma^2)$ 时，其中的参数 μ,σ^2 都有明确的概率含义.

在给出方差的性质之前，我们先介绍一个重要的不等式.

定理 4.4（切比雪夫不等式）　设随机变量 X 的期望和方差均存在，则对任意 $\varepsilon>0$，有

$$P(|X-EX|\geqslant\varepsilon)\leqslant\frac{DX}{\varepsilon^2}, \tag{4.11}$$

等价形式为

$$P(|X-EX|<\varepsilon)\geqslant 1-\frac{DX}{\varepsilon^2}. \tag{4.12}$$

证明　令

$$Y=\begin{cases}1, & \omega\in\{|X-EX|\geqslant\varepsilon\},\\ 0, & \text{其他},\end{cases}$$

则 $Y\leqslant\frac{(X-EX)^2}{\varepsilon^2}$，根据期望的性质，有

$$P(|X-EX|\geqslant\varepsilon)=EY\leqslant E\left[\frac{(X-EX)^2}{\varepsilon^2}\right]=\frac{DX}{\varepsilon^2}.$$

(4.11)式说明：若 X 的方差较小，则事件 $\{|X-EX|\geqslant\varepsilon\}$ 发生的概率就小，事件 $\{|X-EX|<\varepsilon\}$ 发生的概率就大，随机变量 X 取值基本集中于 EX 附近.

定理 4.5　随机变量的方差具有如下性质：

(1) 对任意随机变量 X，有 $DX\geqslant 0$，且 $DX=0$ 的充分必要条件为 $P(X=C)=1$ ($C=EX$ 为常数)；

(2) 对任意常数 a,b，有 $D(a+bX)=b^2DX$；

(3) 若 $X_1,X_2,\cdots,X_n(n\geqslant 2)$ 相互独立，则 $D(X_1+X_2+\cdots+X_n)=DX_1+DX_2+\cdots+DX_n$；

(4) 对一切实数 C，有 $DX=E(X-EX)^2\leqslant E(X-C)^2$. 这个不等式的统计含义是，若用一个实数去集中代表一个随机变量，则其数学期望使均方误差最小，而此最小均方误差是该随机变量的方差.

证明　(1) 由于 $(X-EX)^2\geqslant 0$，从而 $DX=E(X-EX)^2\geqslant 0$，当 $DX=0$ 时，由切比雪夫不等式，对任意的 $n\geqslant 1$，取 $\varepsilon=\frac{1}{n}$，有

$$0\leqslant P\left(|X-EX|\geqslant\frac{1}{n}\right)\leqslant n^2DX=0,$$

由于 $\{|X-EX|>0\}=\bigcup\limits_n\left\{|X-EX|\geqslant\frac{1}{n}\right\}$，根据概率的下连续性，有

$$P(X\neq EX)=P(|X-EX|>0)=P\left(\bigcup_n\left\{|X-EX|\geqslant\frac{1}{n}\right\}\right)=\lim_{n\to+\infty}P\left(|X-EX|\geqslant\frac{1}{n}\right)=0.$$

即 $P(X=EX)=1$，必要性得证.

再证充分性. 当 $P(X=C)=1$ 时，$EX=C$，从而

$$DX=E(C-C)^2=0.$$

(2) 根据方差的定义，得

$$D(a+bX)=E[(a+bX)-E(a+bX)]^2=E[b(X-EX)]^2=b^2E(X-EX)^2=b^2DX.$$

(3) 由于 $X_1,X_2,\cdots,X_n$ 相互独立，所以由方差定义，有

$$\begin{aligned}D\left(\sum_{i=1}^{n}X_i\right)&=E\left[\sum_{i=1}^{n}X_i-E\left(\sum_{i=1}^{n}X_i\right)\right]^2\\&=E[(X_1-EX_1)+(X_2-EX_2)+\cdots+(X_n-EX_n)]^2\\&=E\left[\sum_{i=1}^{n}(X_i-EX_i)^2+2\sum_{i<j}(X_i-EX_i)(X_j-EX_j)\right]\\&=\sum_{i=1}^{n}E(X_i-EX_i)^2+2\sum_{i<j}E[(X_i-EX_i)(X_j-EX_j)]\\&=DX_1+DX_2+\cdots+DX_n.\end{aligned}$$

(4) 由期望的性质(4)知,对任意常数 C,有

$$\begin{aligned}E(X-C)^2&=E(X-EX+EX-C)^2\\&=E[(X-EX)^2+2(X-EX)(EX-C)+(EX-C)^2]\\&=E[(X-EX)^2+(EX-C)^2]\geqslant E(X-EX)^2=DX.\end{aligned}$$

例 4.23　设有随机变量 X,$EX=\mu$,$DX=\sigma^2$,称 $Y=\dfrac{X-\mu}{\sigma}$为 X 的标准化,证明 $EY=0$,$DY=1$.

证明　根据期望与方差的性质,有

$$EY=E\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma}E(X-\mu)=\frac{1}{\sigma}(EX-\mu)=\frac{1}{\sigma}(\mu-\mu)=0,$$

$$DY=D\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma^2}D(X-\mu)=\frac{1}{\sigma^2}DX=\frac{\sigma^2}{\sigma^2}=1.$$

例 4.24　设 $X_1,X_2,\cdots,X_n$ 相互独立,$EX_i=\mu$,$DX_i=\sigma^2(i=1,2,\cdots,n)$,令 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$,求 $E\overline{X}$,$D\overline{X}$.

解　由期望和方差的性质知

$$E\overline{X}=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}EX_i=\mu,$$

$$D\overline{X}=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}D\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{\sigma^2}{n}.$$

例 4.24 的结论在实际中是非常有用的.例如,在进行精密测量时,为了减少测量误差,往往重复测量多次,然后再取其算术平均值.例 4.24 的结果为这种做法给出了一个合理的解释.设被测物的真值为 μ,n 次重复测量可以认为是互不影响的.因而各次测量结果 $X_1,X_2,\cdots,X_n$ 可认为是相互独立同分布的随机变量,且可设 $EX_i=\mu$,$DX_i=\sigma^2(i=1,2,\cdots,n)$,即每一次测量的结果都在真值 μ 的周围波动.由例 4.24 的结果可知,n 次测量的算术平均值 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 应仍在真值 μ 的周围取值,即 $E\overline{X}=\mu$,但 $D\overline{X}=\dfrac{\sigma^2}{n}$,是原来 σ^2 的$\dfrac{1}{n}$倍,因此 $\overline{X}$ 更有可能接近真值,这就是随机方法在精密测量中的应用.但问题并没有到此终结,若在给出一定测量精度的要求和一定的概率保证的前提下,问至少要测量多少次呢?在第五章中,读者将得到一个满意的答案.

例 4.25　设随机变量 X 与随机变量 Y 相互独立,都服从 $N\left(0,\dfrac{1}{2}\right)$分布,求 $E|X-Y|$,$D|X-Y|$.

解　令 $Z=X-Y$，由于 $X\sim N\left(0,\frac{1}{2}\right)$，$Y\sim N\left(0,\frac{1}{2}\right)$，且相互独立，所以 $Z\sim N(0,1)$. 又

$$D(|X-Y|)=D(|Z|)=E(|Z|^2)-[E(|Z|)]^2=E(Z^2)-(E|Z|)^2,$$

而 $E(Z^2)=DZ=1$，且

$$E\mid Z\mid=\int_{-\infty}^{+\infty}\mid z\mid\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z=\frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty}z\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z=\sqrt{\frac{2}{\pi}},$$

所以

$$D(|X-Y|)=1-\frac{2}{\pi}.$$

4.3　随机变量的矩

数学期望和方差都是更广泛的一类数字特征——矩的特殊情形. 在描绘物体质量的分布时矩的概念广泛应用于力学中. 在概率论中，描绘随机变量的概率分布形状，以及第七章的矩估计，也都要使用矩的概念.

定义 4.4　设 X 为随机变量，若 $E|X|^k<\infty$，记 $\upsilon_k=EX^k$，$\alpha_k=E|X|^k$，则称 υ_k 为 X 的 k **阶原点矩**，α_k 为 X 的 k **阶原点绝对矩**.

又若 EX 存在，且 $E(|X-EX|^k)<\infty$，记 $\mu_k=E[X-EX]^k$，则称 μ_k 为 X 的 k **阶中心矩**，称 $\beta_k=E|X-EX|^k$ 为 X 的 k **阶中心绝对矩**.

若随机变量的概率密度是关于期望对称的，则它的一切奇数阶中心矩恒为零，故任何不为零的奇数阶中心矩都可用来衡量分布的偏移（拖尾、不对称）程度. 任何随机变量的一阶中心矩如果存在的话，必恒为零.

在实际应用中，我们经常将原点矩和中心矩进行换算，它们之间的这种关系式在数理统计的参数估计中有重要应用.

显然 $\upsilon_0=1$，$\upsilon_1=EX$，数学期望为一阶原点矩；而 $\mu_0=1$，$\mu_1=0$，$\mu_2=DX$，方差为二阶中心矩.

原点矩和中心矩有如下关系：

$$\mu_k=E(X-EX)^k=\sum_{n=0}^{k}\mathrm{C}_k^n EX^n(-1)^{k-n}(EX)^{k-n}=\sum_{n=0}^{k}(-1)^{k-n}\mathrm{C}_k^n\upsilon_1^{k-n}\upsilon_n,$$

于是有

$$\begin{aligned}
&\mu_0=1,\\
&\mu_1=0,\\
&\mu_2=\upsilon_2-\upsilon_1^2,\\
&\mu_3=\upsilon_3-3\upsilon_2\upsilon_1+2\upsilon_1^2,\\
&\mu_4=\upsilon_4-4\upsilon_3\upsilon_1+6\upsilon_2\upsilon_1^2-3\upsilon_1^4,\\
&\quad\vdots
\end{aligned}$$

上式表明，中心矩可以通过原点矩来表达，反之，原点矩也可以通过中心矩来表达.

事实上，有

$$\upsilon_k=EX^k=E[(X-\upsilon_1)+\upsilon_1]^k=\sum_{n=0}^{k}\mathrm{C}_k^n\upsilon_1^n E(X-\upsilon_1)^{k-n}=\sum_{n=0}^{k}\mathrm{C}_k^n\upsilon_1^n\mu_{k-n},$$

于是有

$$\begin{aligned}
\upsilon_0 &= 1,\\
\upsilon_1 &= EX,\\
\upsilon_2 &= \mu_2+\upsilon_1^2,\\
\upsilon_3 &= \mu_3+3\upsilon_1\mu_2+\upsilon_1^3,\\
\upsilon_4 &= \mu_4+4\upsilon_1\mu_3+6\upsilon_1^2\mu_2+\upsilon_1^4,\\
&\ \ \vdots
\end{aligned}$$

例 4.26　设随机变量 $X\sim N(\mu,\sigma^2)$,求 $E(X-EX)^k(k=1,2,\cdots)$.

解　对任意 $k\geqslant 1$,有

$$\begin{aligned}
\mu_k &= E(X-\mu)^k = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^k\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]\mathrm{d}x\\
&= \frac{\sigma^k}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^k\exp\left[-\frac{t^2}{2}\right]\mathrm{d}t,\quad t=\frac{x-\mu}{\sigma}.
\end{aligned}$$

当 k 为奇数时,被积函数是奇函数,故 $\mu_k=0$.下面设 $k=2n,n=1,2,\cdots,t^2=2u$,则有

$$\begin{aligned}
\mu_{2n} &= \frac{2\sigma^{2n}}{\sqrt{2\pi}}\int_0^{+\infty}t^{2n}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t = \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\int_0^{+\infty}u^{n-\frac{1}{2}}\mathrm{e}^{-u}\mathrm{d}u = \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\Gamma\left(n+\frac{1}{2}\right)\\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\cdot\frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\\
&= \sigma^{2n}(2n-1)!!,
\end{aligned}$$

于是

$$\mu_k=\begin{cases}\sigma^k(k-1)!!, & k\text{ 为偶数},\\ 0, & k\text{ 为奇数}.\end{cases}$$

矩的概念不仅对概率论本身具有重要的应用价值,且在工程技术、生物、医学等领域都有十分重要的应用.

例 4.27(偏度 Skewness)　假设随机变量 X 有二阶和三阶中心矩:$\mu_2=DX$,$\mu_3=E(X-EX)^3$,那么称

$$S(X)=\frac{\mu_3}{\mu_2^{3/2}}$$

为随机变量 X 的**偏度**.

如果 $X\sim N(\mu,\sigma^2)$分布,由例 4.26 可知 $S(X)=0$.实际上只要随机变量的概率密度关于期望对称,其偏度均为零.可见,偏度$S(X)$是描绘随机变量关于其均值不对称程度的数字特征.图 4.2 是具有不同偏度的概率密度示意图,其中 $\mu_i(i=1,2,3)$表示相应分布的数学期望.

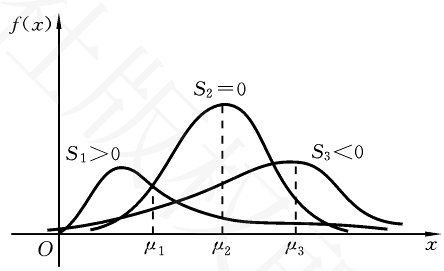

图 4.2

例 4.28(峰度 Kurtosis)　对于任意随机变量 X,称

$$K(X)=\frac{\mu_4}{\mu_2^2}-3$$

为随机变量 X 的**峰度**.

如果 $X\sim N(\mu,\sigma^2)$分布,由例 4.26 可知 $K(X)=0$.事实上,峰度的定义式正是为了便于以正态曲线为标准进行比较.通常峰度越大,密度曲线的顶部越"尖";峰度越小,密度曲线的顶部越"平".图 4.3 是具有不同峰度的概率密度示意图.

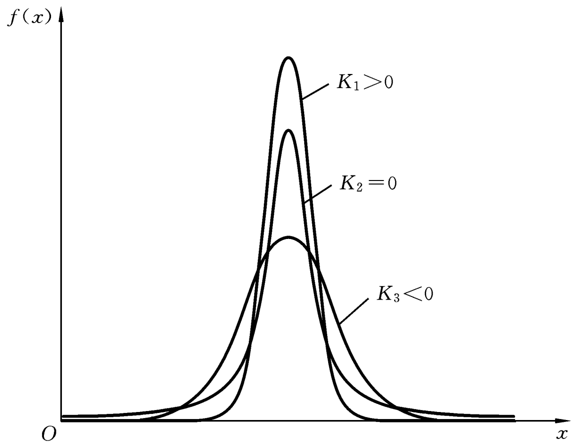

图 4.3

4.4　协方差和相关系数

在研究二维随机变量(X,Y)时,EX、DX、EY、DY等数字特征仅反映了X和Y各自取值的集中位置及其取值相对集中位置的偏离程度,此外我们还想了解X与Y之间在某种意义下的相互关系.由前面研究可知,独立性可用来描述X与Y之间的关系.实际表明,随机变量相互独立的关系很强,且经过观察和研究发现,在随机变量之间还存在一种比相互独立弱的另一种关系,由此我们引入协方差的概念.

4.4.1　随机变量的协方差

定义 4.5　设(X,Y)为二维随机变量,若

$$E|(X-EX)(Y-EY)|$$

存在,则称$E[(X-EX)(Y-EY)]$为随机变量X与Y的**协方差**,记为$\mathrm{Cov}(X,Y)$,即

$$\mathrm{Cov}(X,Y)=E[(X-EX)(Y-EY)]. \tag{4.13}$$

经过简单推导,有

$$\mathrm{Cov}(X,Y)=E(XY)-EXEY. \tag{4.14}$$

例 4.29　设(X,Y)服从二维正态分布$N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,求$\mathrm{Cov}(X,Y)$.

解　根据定义 4.5 知

$$\begin{aligned}\mathrm{Cov}(X,Y)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-EX)(y-EY)f(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\\&\quad\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}\mathrm{d}x\mathrm{d}y,\end{aligned}$$

令$u=\dfrac{x-\mu_1}{\sigma_1}$,$v=\dfrac{y-\mu_2}{\sigma_2}$,得

$$\mathrm{Cov}(X,Y)=\frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}uv\exp\left\{\frac{-1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)\right\}\mathrm{d}u\mathrm{d}v$$

$$
\begin{aligned}
&= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} v\mathrm{d}v \,\frac{1}{\sqrt{2\pi(1-\rho^2)}}\int_{-\infty}^{+\infty} u\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[(u-\rho v)^2+(1-\rho^2)v^2\right]\right\}\mathrm{d}u \\
&= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} v\mathrm{e}^{-\frac{v^2}{2}}\mathrm{d}v\left\{\frac{1}{\sqrt{2\pi(-\rho^2)}}\int_{-\infty}^{+\infty} u\cdot\exp\left[-\frac{1}{2(1-\rho^2)}(u-\rho v)^2\right]\mathrm{d}u\right\} \\
&= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\rho v^2\mathrm{e}^{-\frac{v^2}{2}}\mathrm{d}v = \rho\sigma_1\sigma_2.
\end{aligned}
$$

协方差具有下列性质.

定理 4.6　设以下涉及的协方差均存在,则

(1) 若 X 与 Y 相互独立,则 $\mathrm{Cov}(X,Y)=0$;

(2) 对称性:$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$;

(3) 对任意常数 a,b,有 $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$;

(4) 可加性:$\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.

证明　由协方差的定义可直接推出.

定理 4.7　设 $k_0,k_1,\cdots,k_n$ 为任意常数,$X_1,X_2,\cdots,X_n$ 为随机变量,则

$$
D(k_0+k_1X_1+\cdots+k_nX_n) = \sum_{i=1}^{n}k_i^2DX_i + 2\sum_{\substack{i=1\\ i<j}}^{n}\sum_{j=1}^{n}k_ik_j\mathrm{Cov}(X_i,X_j).
$$

证明　可参见定理 4.5 中(3)的证明过程.

例 4.30　设 N 件产品中有 M 件次品,无放回地从中依次取 n 件,用 X 表示取得的次品数(称 X 服从超几何分布),求 DX.

解　定义随机变量

$$
X_i=\begin{cases}1, & \text{第 } i \text{ 次抽取取得次品},\\ 0, & \text{第 } i \text{ 次抽取取得正品},\end{cases}
$$

则 $X_1,X_2,\cdots,X_n$ 有如下相同的分布列

X_i	0	1
P	$1-\frac{M}{N}$	$\frac{M}{N}$

经过简单计算,有

$$
EX_i=\frac{M}{N},\quad EX_i^2=\frac{M}{N},\quad DX_i=\frac{NM-M^2}{N^2},
$$

令 $X=X_1+X_2+\cdots+X_n$ 表示取得的次品数,则

$$
P(X_i=1,X_j=1)=P(X_i=1)P(X_j=1\mid X_i=1)=\frac{M}{N}\cdot\frac{M-1}{N-1},
$$

因此

$$
\mathrm{Cov}(X_i,X_j)=E(X_iX_j)-EX_iEX_j=\frac{M(M-1)}{N(N-1)}-\frac{M^2}{N^2}.
$$

由定理 4.7 有

$$
\begin{aligned}
DX &= nDX_i + 2\sum_{\substack{i=1\\ i<j}}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i,X_j) \\
&= n\frac{NM-M^2}{N^2} + (n^2-n)\frac{M(M-N)}{N^2(N-1)} = \frac{N-n}{N-1}n\frac{M}{N}\left(1-\frac{M}{N}\right).
\end{aligned}
$$

随机变量 X 与 Y 的协方差虽然反映了 X 与 Y 之间的联系,但它受到 X 和 Y 本身大小及数值尺度的影响.例如,让 X 和 Y 分别增大 k 倍,即 $X_1=kX,Y_1=kY$,这时 X_1 与 Y_1 的联系和

X 与 Y 的联系是一样的，但反映这种联系的协方差却增大到 k^2 倍，即

$$\mathrm{Cov}(X_1,Y_1)=k^2\mathrm{Cov}(X,Y).$$

为克服这一缺陷，将 X 与 Y 标准化：$\dfrac{X-EX}{\sqrt{DX}},\dfrac{Y-EY}{\sqrt{DY}}$，可得到相关系数的概念.

4.4.2　相关系数

定义 4.6　若(X,Y)是二维随机变量，称

$$\rho_{XY}=\mathrm{Cov}\left(\frac{X-EX}{\sqrt{DY}},\frac{Y-EY}{\sqrt{DY}}\right)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\sqrt{DY}} \tag{4.15}$$

为随机变量 X 与 Y 的**相关系数**，简记为 ρ.

例 4.31　设随机变量$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求 ρ_{XY}.

解　已知 $\mathrm{Cov}(X,Y)=\rho\sigma_1\sigma_2$，由定义 4.6 得

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}=\frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2}=\rho.$$

定理 4.8　设 ρ 为随机变量 X 与 Y 的相关系数，则

(1) $|\rho|\leqslant 1$；

(2) $|\rho|=1$ 的充分必要条件为 $P(Y=aX+b)=1$，其中 a,b 为常数.

证明　(1) 令 $X_1=X-EX,Y_1=Y-EY$，对 X_1,Y_1 应用柯西-施瓦茨不等式，有

$$\rho^2=\frac{[E(X-EX)(Y-EY)]^2}{DXDY}=\frac{[E(X_1Y_1)]^2}{EX_1^2EY_1^2}\leqslant 1,$$

即$|\rho|\leqslant 1$.

(2) 若$|\rho|=1$，由(1)知$[E(X_1Y_1)]^2-EX_1^2EY_1^2=0$，从而 $g(t)=E(tX_1-Y_1)^2=0$ 有二重根，记为 t_0，故

$$E(t_0X_1-Y_1)^2=0.$$

又由于

$$E(t_0X_1-Y_1)=E[1\cdot(t_0X_1-Y_1)],$$

以及柯西-施瓦茨不等式和 $E(t_0X_1-Y_1)^2=0$ 的结论，有 $E(t_0X_1-Y_1)=0$，所以

$$D(t_0X_1-Y_1)=0.$$

由方差的性质知，上式成立的充要条件为

$$P[t_0X_1-Y_1=E(t_0X_1-Y_1)]=P(t_0X_1-Y_1=0)=1,$$

将 $X_1=X-EX,Y_1=Y-EY$ 代入，得

$$P(Y=aX+b)=1$$

其中 $a=t_0,b=EY-t_0EX$ 均为常数，定理得证.

我们将 $\rho=0$，称为 X 与 Y 不相关，即 X 与 Y 不相关等价于 $\mathrm{Cov}(X,Y)=0$；$|\rho|=1$，称 X 与 Y 完全相关；$\rho>0$，称 X 与 Y 正相关；$\rho<0$，称 X 与 Y 负相关. 相关系数的符号与随机变量间取值的关系如图 4.4 所示.

相关系数有着广泛的应用. 例如，亲缘关系是生物学中一个十分重要的概念. 数量分类学引进了比亲缘关系更广泛的概念，即相似性的概念. 相似程度的数值表示称为相似性系数，相似性系数的出现是生物分类由定性朝定量方向发展的重要标志. 描述相似性系数有 5 个重要方法：距离系数、相关系数、联合系数、信息系数和模糊系数. 其中将概率统计中的相关系数的

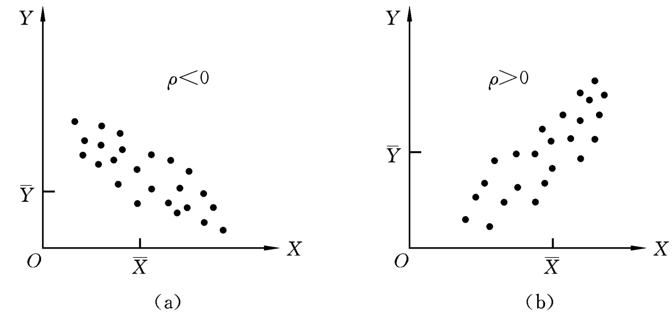

图 4.4

概念移植到数量分类学中,可起到描述生物之间的相似性关系的作用.两个物种相关系数的值越大,两个物种之间的亲缘关系越近;反之值越小,亲缘关系越疏远.另外在经济、金融领域,在证券投资市场上,投资者为了将投资风险分散化,需购进一定种类的股票及配置适当的份额进行投资,此称为证券组合.在最优证券组合的研究中,涉及股票的筛选及份额的配置.其中一个很重要的方法就是尽量避免重复选择两个回报率的相关系数较大的股票作为投资对象,若重复选择,将起不到风险分散化的作用.

例 4.32　设(X,Y)是二维离散型随机变量,在平面上取四个点:$(1,0)$,$(0,1)$,$(-1,0)$,$(0,-1)$,且取每一个点的概率相同,求 ρ_{XY}.

解　根据题意,(X,Y)有如下的联合分布列

X \ Y	-1	0	1	$p_{i\cdot}$
-1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
1	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$p_{\cdot j}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	

由上表可得到 $EX=-\frac{1}{4}+\frac{1}{4}=0$,$EY=0$,$E(XY)=0$,$\mathrm{Cov}(X,Y)=0$,$\rho_{XY}=0$,因此 X 与 Y 不相关,但

$$P(X=-1)P(Y=-1)=\frac{1}{4}\times\frac{1}{4}\neq 0=P(X=-1,Y=-1),$$

说明 X 与 Y 不相互独立.

例 4.33　设(X,Y)的概率密度为

$$f(x,y)=\begin{cases}\frac{1}{\pi}, & x^2+y^2\leqslant 1,\\ 0, & x^2+y^2>1,\end{cases}$$

问:(1) X 与 Y 不相关吗?(2) X 与 Y 是否相互独立?

解　(1) 由于

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\dfrac{1}{\pi}\displaystyle\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\mathrm{d}y=\dfrac{2}{\pi}\sqrt{1-x^2}, & -1\leqslant x\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

同理
$$f_Y(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2}, & -1\leqslant y\leqslant 1,\\ 0, & \text{其他}.\end{cases}$$

$f_X(x),f_Y(y)$ 均为偶函数，故 $EX=EY=0$，而
$$E(XY)=\iint\limits_{x^2+y^2\leqslant 1}\frac{1}{\pi}xy\,\mathrm{d}x\mathrm{d}y=0.$$

故
$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\ \sqrt{DY}}=\frac{E(XY)-EXEY}{\sqrt{DX}\ \sqrt{DY}}=0.$$

所以 X 与 Y 不相关.

(2) 由于 $f(x,y)\neq f_X(x)f_Y(y)$，从而 X,Y 不相互独立.

由例 4.32 和例 4.33 可以看出，虽然 X 与 Y 不相关，但 X 与 Y 并不相互独立，显然当 X 与 Y 独立时一定不相关，反之不成立. 当然，对于二维正态分布，不相关和独立两者是一致的，因为 X 与 Y 相互独立的充要条件是 $\rho=0$. 同时也可看出随机变量独立性条件强于不相关性条件，因此定理 4.3 中(3)和定理 4.5 中(3)的前提条件可减弱为不相关.

为了进一步探明相关系数的概率含义，我们考虑这样一个问题：利用随机变量 X 的信息，构造一个线性函数 $a+bX$，使它尽可能地近似随机变量 Y（也就是使均方误差最小），那么 a,b 该如何取值？这个均方误差又与两随机变量间的相关系数有何关系？为此我们有如下定理.

定理 4.9　设 X 和 Y 为两个随机变量，则
$$E[Y-(\hat{a}+\hat{b}X)]^2=\min_{a,b}E[Y-(a+bX)]^2=DY(1-\rho_{XY}^2),$$
其中 $\hat{a},\hat{b}$ 由方程
$$\hat{b}=\frac{\mathrm{Cov}(X,Y)}{DX}=\rho_{XY}\frac{\sqrt{DY}}{\sqrt{DX}},\quad EY=\hat{b}EX+\hat{a}$$
所确定.

证明　根据问题的含义要计算如下的均方误差
$$e(a,b)=E[Y-(a+bX)]^2=EY^2+b^2EX^2+a^2-2bE(XY)+2abEX-2aEY,$$
要使它达到最小，分别对 a,b 求偏导数，则
$$\begin{cases}\dfrac{\partial e(a,b)}{\partial a}=2a+2bEX-2EY=0,\\[2mm] \dfrac{\partial e(a,b)}{\partial b}=2bEX^2-2E(XY)+2aEX=0,\end{cases}$$
解方程组得到 $\hat{b}=\dfrac{\mathrm{Cov}(X,Y)}{DX}=\rho_{XY}\dfrac{\sqrt{DY}}{\sqrt{DX}}$，$EY=\hat{b}EX+\hat{a}$，则
$$\begin{aligned}\min_{a,b}e(a,b)&=e(\hat{a},\hat{b})=E[Y-(\hat{a}+\hat{b}X)]^2=D[Y-(\hat{a}+\hat{b}X)]+\{E[Y-(\hat{a}+\hat{b}X)]\}^2\\&=D(Y-\hat{b}X)=DY+\hat{b}^2DX-2\hat{b}\mathrm{Cov}(X,Y)=DY(1-\rho_{XY}^2).\end{aligned}$$

定理 4.9 说明，$|\rho_{XY}|$ 越接近 1，用 $a+bX$ 近似 Y 的效果越好，X 与 Y 的线性关系越强；$|\rho_{XY}|$ 越接近于 0，用 $a+bX$ 近似 Y 的效果越差，X 与 Y 的线性关系越弱；如果 $\rho_{XY}=0$，则 $\hat{b}=0$，此时用 X 的线性函数去近似 Y 的最佳情形是常数，说明在统计意义下 X 与 Y 没有线性关

系.所以相关系数只是刻画 X 与 Y 的线性相关程度的数字特征.

综上,我们得到如下关系:

(1) 若 X,Y 相互独立,且 X,Y 的二阶矩非零有限,则 X 与 Y 一定不相关;

(2) 若 X,Y 不相关,则 X 与 Y 不一定独立;

(3) 若 X,Y 相关,则 X 与 Y 一定不独立.

例 4.34　设随机变量 X,Y 的方差都存在,令

$$U=aX+b,\quad V=cY+d,$$

其中 a,b,c,d 为常数,则 U,V 的相关系数

$$\rho_{UV}=\frac{ac}{|ac|}\rho_{XY}=\pm\rho_{XY}.$$

证明　根据协方差的性质,有

$$\mathrm{Cov}(U,V)=\mathrm{Cov}(aX+b,cY+d)=ac\mathrm{Cov}(X,Y),$$
$$DU=D(aX+b)=a^2DX,\quad DV=D(cY+d)=c^2DY.$$

因此有
$$\rho_{UV}=\frac{\mathrm{Cov}(U,V)}{\sqrt{DU}\sqrt{DV}}=\frac{ac\mathrm{Cov}(X,Y)}{|ac|\sqrt{DX}\sqrt{DY}}=\frac{ac}{|ac|}\rho_{XY}.$$

此例题表明,随机变量经线性变换不改变两随机变量间线性关系的密切程度,只是有可能改变正负相关的性质.说明相关系数不依赖于原点和单位的选取.

4.4.3　协方差矩阵

定义 4.7　设 (X,Y) 为二维随机变量,若

$$E(X^kY^l),\quad k,l=1,2,\cdots$$

存在,则称它为 X 和 Y 的 (k,l) 阶**联合原点矩**;若

$$E[(X-EX)^k(Y-EY)^l],\quad k,l=1,2,\cdots$$

存在,则称它为 X 和 Y 的 (k,l) 阶**联合中心矩**.

注　定义 4.7 要求

$$E[|X^kY^l|]<+\infty,\quad k,l=1,2,\cdots,$$
$$E[|(X-EX)^k(Y-EY)^l|]<+\infty,\quad k,l=1,2,\cdots.$$

显然,协方差为 $(1,1)$ 阶联合中心矩.

对于二维随机变量 (X_1,X_2),记

$$C_{11}=DX_1=E(X_1-EX_1)^2,$$
$$C_{22}=DX_2=E(X_2-EX_2)^2,$$
$$C_{12}=\mathrm{Cov}(X_1,X_2)=E(X_1-EX_1)(X_2-EX_2),$$
$$C_{21}=\mathrm{Cov}(X_2,X_1)=E(X_2-EX_2)(X_1-EX_1),$$

其中 $C_{12}=C_{21}$.它们排成的矩阵

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

称为 (X_1,X_2) 的**协方差矩阵**.

例如,设 $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,则 (X_1,X_2) 的协方差矩阵为

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

一般地，对于 n 维随机变量$(X_1,X_2,\cdots,X_n)$，其协方差矩阵为

$$\begin{pmatrix} \mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \cdots & \mathrm{Cov}(X_2,X_n) \\ \vdots & \vdots & & \vdots \\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Cov}(X_n,X_n) \end{pmatrix}$$

简记为 $\boldsymbol{\Sigma}$.

例 4.35　假设随机变量 X 和 Y 独立同服从参数为 λ 的泊松分布，令 $U=2X+Y,V=2X-Y$，求 ρ_{UV} 及协方差矩阵 $\boldsymbol{\Sigma}$.

解　由于 $EX=EY=\lambda,DX=DY=\lambda$，所以

$$EX^2=EY^2=DX+(EX)^2=\lambda+\lambda^2,$$

$$EU=2EX+EY=3\lambda,\quad EV=2EX-EY=\lambda,$$

$$DU=DV=4DX+DY=4\lambda+\lambda=5\lambda.$$

$$E(UV)=E(4X^2-Y^2)=4EX^2-EY^2=3\lambda+3\lambda^2,$$

$$\mathrm{Cov}(U,V)=EUV-EUEV=3\lambda+3\lambda^2-3\lambda^2=3\lambda.$$

从而

$$\rho_{UV}=\frac{\mathrm{Cov}(U,V)}{\sqrt{DU}\sqrt{DV}}=\frac{3\lambda}{5\lambda}=\frac{3}{5},\quad \boldsymbol{\Sigma}=\begin{pmatrix} 5\lambda & 3\lambda \\ 3\lambda & 5\lambda \end{pmatrix}.$$

该题的求解过程并不需要求 U 和 V 的联合分布及边缘分布，而是完全利用有关性质进行运算，这是一种常用方法.

4.5　条件数学期望

在第三章我们曾经引进了条件分布的概念，依照数学期望的定义，在此我们引入相应的条件数学期望的定义，平行地引出有关性质、全期望公式，并说明它们的应用.

4.5.1　条件期望的定义

设(X,Y)是二维离散型随机变量，有联合概率分布

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

则 Y 有边缘分布

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

已知 $Y=y_j$ 时 X 有条件分布

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

若

$$E(|X||Y=y_j)=\sum_{i=1}^{\infty}|x_i|\frac{p_{ij}}{p_{\cdot j}}<+\infty,$$

就可以定义在 $Y=y_j$ 条件下 X 的数学期望，即

$$m(y_j)=E(X|Y=y_j)=\sum_{i=1}^{\infty}x_i\frac{p_{ij}}{p_{\cdot j}}. \tag{4.16}$$

设(X,Y)是二维连续型随机变量，有概率密度 $f(x,y)$，由 3.2 节可知 Y 有边缘概率密度

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\,,$$

对满足 $f_Y(y)>0$ 的 y,已知 $Y=y$ 时 X 有条件概率密度

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}.$$

若 $$E(|X||Y=y)=\int_{-\infty}^{+\infty}|x|f_{X|Y}(x|y)\mathrm{d}x<+\infty,$$

就可以定义在 $Y=y$ 条件下,X 的数学期望如下:

$$m(y)=E(X|Y=y)=\int_{-\infty}^{+\infty}xf_{X|Y}(x|y)\mathrm{d}x. \tag{4.17}$$

定义 4.8 设(X,Y)是二维随机变量,$E|X|<\infty$,如果 $m(y)$是在条件 $Y=y$ 下 X 的数学期望,即

$$m(y)=E(X|Y=y),$$

就称 $m(Y)$为 X 关于 Y 的**条件数学期望**,简称为**条件期望**,记为 $E(X|Y)$. 同样,也可定义 Y 关于 X 的条件期望 $E(Y|X)$.

由定义看出,$E(X|Y=y)$依赖于 y,当 y 固定时,$E(X|Y=y)$是一个常数,但 y 是随机变量 Y 的一个可能取值,因此条件期望 $E(X|Y)$是一个随机变量,为随机变量 Y 的一个函数.

$E(X|Y=y)$直观地可以理解为:在已知信息 $Y=y$ 时,对于随机变量 X 的最合理估值(因为它是“已知信息 $Y=y$ 下的平均值”).

例 4.36 假设随机变量 X 和 Y 独立,且分别服从参数为 λ 和 μ 的泊松分布,求 $P(X=k|X+Y=m)$和 $E(X|X+Y=m)$,其中 $k\leqslant m$.

解 $X\sim P(\lambda)$,$Y\sim P(\mu)$,根据前面的结论有 $X+Y\sim P(\lambda+\mu)$,因此

$$\begin{aligned}P(X=k|X+Y=m)&=\frac{P(X=k,X+Y=m)}{P(X+Y=m)}=\frac{P(X=k,Y=m-k)}{P(X+Y=m)}\\&=\frac{\lambda^k\mathrm{e}^{-\lambda}}{k!}\times\frac{\mu^{m-k}\mathrm{e}^{-\mu}}{(m-k)!}\Big/\frac{(\lambda+\mu)^m\mathrm{e}^{-(\lambda+\mu)}}{m!}=\mathrm{C}_m^k\frac{\lambda^k\mu^{m-k}}{(\lambda+\mu)^m}.\end{aligned}$$

由条件期望的定义,有

$$\begin{aligned}E(X|X+Y=m)&=\sum_{k=0}^{m}kP(X=k|X+Y=m)=\sum_{k=0}^{m}k\mathrm{C}_m^k\frac{\lambda^k\mu^{m-k}}{(\lambda+\mu)^m}\\&=\frac{m\lambda}{(\lambda+\mu)^m}\sum_{k=1}^{m}\frac{(m-1)!}{(k-1)!(m-k)!}\lambda^{k-1}\mu^{m-k}\\&=\frac{m\lambda}{(\lambda+\mu)^m}\times(\lambda+\mu)^{m-1}=\frac{m\lambda}{\lambda+\mu}.\end{aligned}$$

例 4.37 若$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,求条件期望 $E(X|Y)$和 $E(Y|X)$.

解 条件密度

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{1-\rho^2}\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y-\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2},$$

它是正态分布 $N\left(\mu_2+\rho\dfrac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\right)$的概率密度,于是由

$$E(Y|X=x)=\int_{-\infty}^{+\infty}yf_{Y|X}(y|x)\mathrm{d}y=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),$$

得到 $$E(Y|X)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1),$$

同理,得到
$$E(X|Y)=\mu_1+\rho\frac{\sigma_1}{\sigma_2}(Y-\mu_2).$$

4.5.2 条件期望的性质

定理 4.10 设 $X,Y,X_1,X_2,\cdots,X_n$ 是随机变量,$g(x)$,$h(y)$是函数,又设所有随机变量的数学期望均存在,则

(1) $E\left(C_0+\sum_{i=1}^{n}C_iX_i\mid Y\right)=C_0+\sum_{i=1}^{n}C_iE(X_i\mid Y)$,其中 $C_0,C_1,\cdots,C_n$ 是常数;

(2) $E[h(Y)g(X)|Y]=h(Y)E[g(X)|Y]$;

(3) 当 X,Y 独立时,$E[g(X)|Y]=E[g(X)]$;

(4) $E[E(g(X)|Y)]=E[g(X)]$.

证明 (1) 根据条件期望的定义可直接得到.

(2) 由于 $E[h(Y)g(X)|Y=y]=E[h(y)g(X)|Y=y]=h(y)E[g(X)|Y=y]$,

可得到
$$E[h(Y)g(X)|Y]=h(Y)E[g(X)|Y].$$

(3) 设(X,Y)的联合概率密度为 $f(x,y)$,由于 X 与 Y 独立,满足 $f(x,y)=f_X(x)\cdot f_Y(y)$,由条件期望的定义,有
$$\begin{aligned}E[g(X)\mid Y=y]&=\int_{-\infty}^{+\infty}g(x)f_{X|Y}(x\mid y)\mathrm{d}x=\int_{-\infty}^{+\infty}g(x)\frac{f(x,y)}{f_Y(y)}\mathrm{d}x\\&=\int_{-\infty}^{+\infty}g(x)f_X(x)\mathrm{d}x=E[g(X)].\end{aligned}$$

离散情形可类似证明.

(4) 由期望的定义,有
$$\begin{aligned}E[E(g(X)\mid Y)]&=\int_{-\infty}^{+\infty}E[g(X)\mid Y=y]f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x)f_{X|Y}(x\mid y)\mathrm{d}xf_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x)f_{X|Y}(x\mid y)f_Y(y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x)f(x,y)\mathrm{d}x\mathrm{d}y=E[g(X)].\end{aligned}$$

离散情形可类似证明.

例 4.38(最佳预测问题) 设 X,Y 为两个随机变量,且 $EX^2<+\infty$,记 $m(Y)=E(X|Y)$,则对任何函数 $g(y)$,有
$$E[X-m(Y)]^2\leqslant E[X-g(Y)]^2.$$

证明
$$\begin{aligned}E[X-g(Y)]^2&=E[X-m(Y)+m(Y)-g(Y)]^2\\&=E[(X-m(Y))^2+2(X-m(Y))(m(Y)-g(Y))+(m(Y)-g(Y))^2]\\&=E[X-m(Y)]^2+E[m(Y)-g(Y)]^2+2E(X-m(Y))(m(Y)-g(Y)),\end{aligned}$$

令 $h(Y)=m(Y)-g(Y)$,由条件期望的性质,有
$$\begin{aligned}E((X-m(Y))(m(Y)-g(Y)))&=E(X-m(Y))h(Y)=E[E[h(Y)(X-m(Y))|Y]]\\&=E[h(Y)[E(X-m(Y))|Y]]=0.\end{aligned}$$

因此 $E[X-g(Y)]^2=E[X-m(Y)]^2+E[m(Y)-g(Y)]^2\geqslant E[X-m(Y)]^2$.

例 4.38 可直观地理解为:我们希望找到 X 与 Y 之间的函数关系 $g(y)$,目的是要通过 Y

的观测值来预测 X,由证明可知最佳的函数是 X 在 Y 下的条件期望 $E(X|Y)$.

由例 4.37 可知 $E(X|Y)$ 是 Y 的线性函数,所以对正态分布而言,最佳预测等于最佳线性预测.

推论 4.1(全期望公式) 设 X,Y 为两个随机变量,若 EX 存在,则

$$E[E(X|Y)]=EX. \tag{4.18}$$

若 (X,Y) 为二维离散型随机变量,则

$$EX=\sum_{j=1}^{\infty}E(X\mid Y=y_j)P(Y=y_j), \tag{4.19}$$

若 (X,Y) 为二维连续型随机变量,则

$$EX=\int_{-\infty}^{+\infty}E(X\mid Y=y)f_Y(y)\mathrm{d}y. \tag{4.20}$$

证明 在定理 4.10 的(4)中,令 $g(x)=x$,即可得证.

例 4.39 设某超级市场周日的顾客总数 N 服从泊松分布 $P(\lambda)$,又设顾客之间的消费额相互独立同分布,并且和 N 独立,用 S 表示该日的全天营业额,当顾客的平均消费是 μ 时,求 $E(S|N)$ 和 ES.

解 令 X_i 表示第 i 个顾客的消费额,于是由

$$\begin{aligned}E(S|N=n)&=E(X_1+X_2+\cdots+X_n|N=n)\\&=E(X_1|N=n)+E(X_2|N=n)+\cdots+E(X_n|N=n)\\&=EX_1+EX_2+\cdots+EX_n=n\mu,\end{aligned}$$

得到

$$E(S|N)=N\mu.$$

$$ES=E[E(S|N)]=E[N\mu]=\mu EN=\mu\lambda.$$

例 4.40 假设随机变量 X_1 在 $[0,1]$ 上有均匀分布,X_i 在 $[X_{i-1},X_{i-1}+1]$ 上有均匀分布,其中 $i=1,2,\cdots,n$,求 EX_n.

解 由均匀分布的期望可知

$$E(X_n|X_{n-1}=x)=\frac{x+x+1}{2}=x+\frac{1}{2},$$

因此 $E(X_n|X_{n-1})=X_{n-1}+\dfrac{1}{2}$. 根据全期望公式,由递推法得到

$$EX_n=E[E(X_n|X_{n-1})]=E\left(X_{n-1}+\frac{1}{2}\right)=EX_{n-1}+\frac{1}{2}=EX_1+\frac{n-1}{2}=\frac{1}{2}+\frac{n-1}{2}=\frac{n}{2}.$$

习 题 四

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

4.2 如果随机变量 X、Y 满足 $D(X+Y)=D(X-Y)$,则必有( ).

(A) X 与 Y 独立 (B) X 与 Y 不相关 (C) $DY=0$ (D) $DX=0$

4.3 设随机变量 X 和 Y 都服从正态分布,且它们不相关,则( ).

(A) X 和 Y 一定独立 (B) (X,Y) 服从二维正态分布

(C) X 和 Y 未必独立 (D) $X+Y$ 服从一维正态分布

4.4 设随机变量 X 服从正态分布 $N(\mu_1,\sigma_1^2)$,随机变量 Y 服从正态分布 $N(\mu_2,\sigma_2^2)$,且 $P(|X-\mu_1|<1)>$

$P(|Y-\mu_2|<1)$，则有(　　).

(A) $\sigma_1<\sigma_2$　　(B) $\sigma_1>\sigma_2$　　(C) $\mu_1<\mu_2$　　(D) $\mu_1>\mu_2$

4.5　一整数等可能地在 1 到 10 中取值，以 X 记除得尽这一整数的正整数的个数，求 EX.

4.6　设有离散型随机变量 X，其可能取值为 $1,2,\cdots$，如果 $P(X=k)$ 对 $k=1,2,\cdots$ 是不增的，试证 $P(X=k)\leqslant 2EX/k^2$.

4.7　某地区流行某种疾病，患者约占 10%，为开展防治工作，要对 n 位居民验血，现将 k 人合并为一组，混合化验，如果合格，则 k 人只要化验一次，若发现有问题，再对这一组 k 人逐个化验，问 k 取何值时最优？

4.8　设有独立试验序列，每次试验只有两个结果：成功或失败，且成功的概率为 p，又设第 r 次成功恰好出现在第 X 次，求 EX.

4.9　逐批检查产品，对于一批产品，若抽查到第 n_0 件仍未发现废品，则立即停止检查认为质量合格，设产品的废品率为 p，且每批产品的数量都很大，因此可以认为每次查到废品的概率都是 p，问平均每批产品要检查多少件？

4.10　试证：对取非负整数值的随机变量 X，若 EX 存在，则 $EX=\sum\limits_{k=1}^{\infty}P(X\geqslant k)$.

4.11　设一次试验成功的概率为 p，进行 100 次独立重复试验，当 p 为何值时，成功次数的标准差最大？其最大值等于什么？

4.12　设 X 表示 10 次独立重复射击命中目标的次数，每次命中目标的概率为 0.4，求 EX^2.

4.13　设随机变量 X 的概率密度为

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

求 DX.

4.14　设随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{2}{\pi}\cos^2 x, & |x|\leqslant\dfrac{\pi}{2},\\ 0, & |x|>\dfrac{\pi}{2},\end{cases}$$

求 EX,DX.

4.15　设随机变量 X 的概率密度为

$$f(x)=\frac{1}{2}e^{-|x|},\quad -\infty<x<+\infty,$$

(1) 求 X 的数学期望 EX 和方差 DX；

(2) 求 X 与 $|X|$ 的协方差，并问 X 与 $|X|$ 是否不相关？

(3) 问 X 与 $|X|$ 是否相互独立？为什么？

4.16　设连续型随机变量 X 的分布函数为

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

试确定常数 a,b，并求 EX 及 DX.

4.17　在长为 l 的线段上任取两点，试求两点间距离的期望及方差.

4.18　设随机变量 (X,Y) 在区域 $D(D=\{(x,y):0<x<1,|y|<x\})$ 内服从均匀分布，求随机变量 $Z=2X+1$ 的方差 DZ.

4.19　设随机变量 X_1 与 X_2 相互独立，其概率密度分别为

$$f_1(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}\qquad f_2(x)=\begin{cases}e^{-(x-5)}, & x>5,\\ 0, & \text{其他}.\end{cases}$$

求 $E(X_1X_2)$.

4.20 设 A,B 为随机事件,且 $P(A)=\frac{1}{4},P(B|A)=\frac{1}{3},P(A|B)=\frac{1}{2}$,令

$$X=\begin{cases}1, & A\text{ 发生},\\ 0, & A\text{ 不发生},\end{cases}\quad Y=\begin{cases}1, & B\text{ 发生},\\ 0, & B\text{ 不发生}.\end{cases}$$

求:(1) 二维随机变量 (X,Y) 的概率分布;(2) X 与 Y 的相关系数 ρ_{XY}.

4.21 将 n 个球放入 M 个盒子中,设每个球放入各个盒子是等可能的,求有球的盒子数 X 的期望.

4.22 设随机变量 X 和 Y 同分布,X 的概率密度为

$$f(x)=\begin{cases}\frac{3}{8}x^2, & 0<x<2,\\ 0, & \text{其他}.\end{cases}$$

(1) 已知事件 $A=\{X>a\}$ 和 $B=\{Y>a\}$ 独立,且 $P(A\cup B)=\frac{3}{4}$,求常数 a.

(2) 求 $\frac{1}{X^2}$ 的数学期望.

4.23 证明:若对随机变量 X 有 $E\mathrm{e}^{aX}<+\infty$,其中 $a>0$ 为常数,则

$$P(X\geqslant\varepsilon)\leqslant E\mathrm{e}^{aX}/\mathrm{e}^{a\varepsilon}.$$

4.24 设由自动生产线加工的某种零件的内径 X(单位:mm)$\sim N(\mu,1)$,内径小于 10 或大于 12 的为不合格品,余下的为合格品,销售每件合格品获利,处理销售不合格的产品亏损,已知销售利润 T(单位:元)与销售零件内径 X 有如下关系:

$$T=\begin{cases}-1, & X<10,\\ 20, & 10\leqslant X\leqslant 12,\\ -5, & X>12,\end{cases}$$

问平均内径 μ 取何值时,销售一个零件的平均利润最大?

4.25 设二维随机变量 (X,Y) 的概率密度为

$$f(x,y)=\frac{1}{2}[\varphi_1(x,y)+\varphi_2(x,y)],$$

其中 $\varphi_1(x,y)$ 和 $\varphi_2(x,y)$ 都是二维正态概率密度,且它们对应的二维随机变量的相关系数分别为 $\frac{1}{3}$ 和 $-\frac{1}{3}$,它们的边缘密度函数所对应的随机变量的数学期望都是零,方差都是 1.

(1) 求随机变量是 X 和 Y 的概率密度 $f_1(x)$ 和 $f_2(y)$,以及 X 和 Y 的相关系数 ρ.

(2) 问 X 和 Y 是否独立? 为什么?

4.26 设随机变量 (X,Y) 的联合分布列为

X \ Y	0	1	2
0	0.1	0	0.2
1	0	0.1	0.2
2	0.2	0	0.2

求 $EX,EY,\mathrm{Cov}(X,Y)$.

4.27 设 X 与 Y 相互独立,证明

$$D(XY)=DXDY+(EX)^2DY+(EY)^2DX.$$

4.28 已知随机变量 $X\sim N(1,3^2),Y\sim N(0,4^2)$,且 $\rho_{XY}=-\frac{1}{2}$,设 $Z=\frac{X}{3}+\frac{Y}{2}$,求:(1) EZ,DZ;(2) ρ_{XZ}.

4.29 设 X,Y 相互独立同分布,已知 X 的概率分布为 $P(X=i)=\frac{1}{3},i=1,2,3$. 设 $Z=\max(X,Y),W=$

$\min(X,Y)$. 求:(1) (Z,W)的概率分布;(2) EZ.

4.30　$X_1, X_2, \cdots, X_n$ 是独立同分布且方差有限的随机变量,设 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$,证明:$X_i-\overline{X}$ 与 $X_j-\overline{X}$ $(i \neq j)$ 的相关系数为 $-\frac{1}{n-1}$.

4.31　设随机变量 X 的概率分布为 $P(X=1)=P(X=2)=\frac{1}{2}$,在给定 $X=i$ 的条件下,随机变量 Y 服从均匀分布 $U(0,i)(i=1,2)$. 求:(1) Y 的分布函数 $F_Y(y)$;(2) EY.

4.32　设随机变量 X 的概率密度为

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

对 X 进行独立重复的观测,直到第 2 个大于 3 的观测值出现时停止,记 Y 为观测次数. 求:(1) Y 的概率分布;(2) EY.

第五章　大数定律和中心极限定理

大数定律和中心极限定理是概率论中两类重要的极限定理的统称.一类极限定理阐明大量随机现象平均值的稳定性,这类极限定理的命题统称为大数定律;另一类极限定理说明了在一定条件下,大量随机变量和的极限分布为正态分布,这类极限定理的命题统称为中心极限定理.它们是概率论的基本理论之一,也是数理统计学的理论基础,因此占有极其重要的地位.下面分别对大数定律和中心极限定理作简单介绍.

5.1　大数定律

概率公理化定义的客观基础是频率的稳定性.所谓事件发生的频率具有稳定性,是指随着试验次数的增加,事件发生的频率逐渐稳定于某个常数.在实践中人们还发现,大量观测值的算术平均值也具有稳定性.大数定律就是要对上述客观事实给出严格的数学论证.

首先,引入随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立、互不相关两个概念.若对于任意 $n>1$,$X_1,X_2,\cdots,X_n$ 都独立,则称随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立.若 $X_1,X_2,\cdots,X_n,\cdots$ 两两不相关,则称随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ 互不相关.

其次,给出大数定律和依概率收敛的定义.

定义 5.1　设 $X_1,X_2,\cdots,X_n,\cdots$ 是一列随机变量,记 $\overline{X}_n=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i,n=1,2,\cdots$. 若存在常数序列 $a_1,a_2,\cdots,a_n,\cdots$,使得对任意给定的 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\{|\overline{X}_n-a_n|<\varepsilon\}=1,\tag{5.1}$$

则称序列 $X_1,X_2,\cdots,X_n,\cdots$ 服从**大数定律**.

该定义的直观含义是:服从大数定律的随机变量序列 $\{X_n\}$,其前 n 个随机变量的算术平均值 $\{\overline{X}_n\}$,也是一个随机变量序列,在概率意义下以常数序列 $\{a_n\}$ 为极限,即当 n 充分大时,二者的误差小于任意正数 ε 这一事件的概率将无限接近于 1.通俗地说,n 个随机变量的算术平均这一随机序列当 n 无限增加时几乎必然趋于一个常数序列.

一般地,常数序列 $a_1,a_2,\cdots,a_n,\cdots$ 取为 $E\overline{X}_1,E\overline{X}_2,\cdots,E\overline{X}_n,\cdots$.

定义 5.2　设 $Y_1,Y_2,\cdots,Y_n,\cdots$ 是一列随机变量,a 是常数,若对任意给定的 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\{|Y_n-a|<\varepsilon\}=1,\tag{5.2}$$

则称序列 $Y_1,Y_2,\cdots,Y_n,\cdots$ 依概率收敛于 a.记为

$$Y_n\xrightarrow{P}a.$$

该定义的直观含义是:依概率收敛于 a 的随机变量序列 $\{Y_n\}$ 在概率意义下以常数 a 为极限.换言之,当 n 无限增加时,随机变量序列 $\{Y_n\}$ 将几乎必然趋于常数 a,二者的误差小于任给正数 ε 这一事件的概率将无限接近于 1.其数学描述跟微积分中常数序列的极限定义非常类似.

依概率收敛的序列有以下常用性质.

设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，又设函数 $g(x,y)$ 在点 (a,b) 处连续，则

$$g(X_n, Y_n) \xrightarrow{P} g(a,b).$$

此性质证明如下.

由 $g(x,y)$ 在点 (a,b) 处的连续性知，对任给 $\varepsilon>0$，必存在 $\delta>0$，使得当 $|x-a|+|y-b|<\delta$ 时有 $|g(x,y)-g(a,b)|<\varepsilon$，于是

$$\{|g(X_n,Y_n)-g(a,b)|\geqslant\varepsilon\}\subset\{|X_n-a|+|Y_n-b|\geqslant\delta\}\subset\{|X_n-a|\geqslant\delta/2\}\cup\{|Y_n-b|\geqslant\delta/2\}.$$

因此　$P\{|g(X_n,Y_n)-g(a,b)|\geqslant\varepsilon\}\leqslant P\{|X_n-a|\geqslant\delta/2\}+P\{|Y_n-b|\geqslant\delta/2\}\xrightarrow{n\to\infty}0$，

亦即
$$\lim_{n\to\infty}P\{|g(X_n,Y_n)-g(a,b)|<\varepsilon\}=1.$$

下面介绍三个常用的大数定律，它们分别反映了算术平均值及频率的稳定性.

定理 5.1（契比雪夫大数定律的特殊情况）　设 $X_1, X_2, \cdots, X_n, \cdots$ 是互不相关的随机变量序列，且具有相同的期望和方差：$EX_i=\mu, DX_i=\sigma^2 (i=1,2,\cdots)$，则随机变量序列 $\{X_n\}$ 服从大数定律，即对于任意 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|\overline{X}_n-\mu|<\varepsilon\}=1, \tag{5.3}$$

或等价地有

$$\lim_{n\to\infty}P\{|\overline{X}_n-\mu|\geqslant\varepsilon\}=0. \tag{5.4}$$

换言之，前 n 个随机变量的算术平均值序列 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ 依概率收敛于 μ.

证明　用契比雪夫不等式证(5.3)式. 先求 $\overline{X}_n$ 的期望、方差：

$$E\overline{X}_n=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\left(\sum_{i=1}^{n}EX_i\right)=\frac{1}{n}n\mu=\mu,$$

$$D\overline{X}_n=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\left(\sum_{i=1}^{n}DX_i\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}.$$

由契比雪夫不等式，可得下式中的右侧不等式成立：

$$1\geqslant P\{|\overline{X}_n-\mu|<\varepsilon\}\geqslant 1-\frac{\sigma^2/n}{\varepsilon^2}.$$

当 $n\to\infty$ 时，上式左、右两边的常数序列极限皆为 1，由极限的夹逼定理可知，不等式中间的常数序列的极限也存在且为 1，即

$$\lim_{n\to\infty}P\{|\overline{X}_n-\mu|<\varepsilon\}=1.$$

从而(5.3)式得证.

由对立事件的概率公式立得(5.4)式.

特别地，若将定理中的条件“互不相关”改为“相互独立”，即对一个相互独立且具有相同的期望和方差的随机变量序列，结论也成立.

契比雪夫大数定律说明互不相关（含相互独立）且具有相同期望和方差的随机变量序列的算术平均 $\overline{X}_n$，也是一个随机变量序列，依概率收敛于每个随机变量的期望（常数）μ.

定理 5.1 要求随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 的方差存在. 但在这些随机变量服从相同分布的场合，并不需要这一条件. 我们有以下的定理.

定理 5.2（辛钦大数定律）　设 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量序列，若期望

$EX_n=\mu$ 有限,则对任意正数 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1, \tag{5.5}$$

或

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geqslant\varepsilon\right\}=0.$$

即前 n 个随机变量的算术平均值序列 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ 依概率收敛于 μ.

定理 5.2 的证明超出了本书范围,从略.

辛钦大数定律说明独立同分布且期望有限的随机变量序列的算术平均值,也是一个随机变量序列,依概率收敛于每个随机变量的期望(常数)μ.

契比雪夫大数定律和辛钦大数定律都从理论上严格地证明了算术平均值的稳定性,与我们在实际中观察到的平均值具有稳定性的客观事实相吻合.

契比雪夫大数定律和辛钦大数定律的直观解释是:虽然每个随机变量的取值都相对于它们的期望 μ 有一定的正负偏差,但是取了算术平均值以后的随机变量序列,其正负偏差在很大程度上得到了抵消,从而稳定在每一个随机变量的期望 μ(常数)附近.

辛钦大数定律应用非常广泛.实际中随机变量 X 的期望 μ 通常是未知的,根据辛钦大数定律,利用实际推断原理,我们可以对随机变量 X 独立地观测多次,然后用它们的算术平均值估计 μ,就可以达到很好的效果.这是数理统计中参数估计的重要理论依据之一.

定理 5.3(伯努利大数定律)　设 n_A 是 n 次独立重复(伯努利)试验中事件 A 发生的次数,p 是每次试验中事件 A 发生的概率,则对于任给正数 $\varepsilon>0$,有

$$\lim_{n\to\infty}P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1, \tag{5.6}$$

或

$$\lim_{n\to\infty}P\left\{\left|\frac{n_A}{n}-p\right|\geqslant\varepsilon\right\}=0.$$

证明　引入随机变量

$$X_i=\begin{cases}0,\text{在第 } i \text{ 次试验中 } A \text{ 不发生},\\ 1,\text{在第 } i \text{ 次试验中 } A \text{ 发生},\end{cases}\quad i=1,2,\cdots.$$

显然

$$n_A=X_1+X_2+\cdots+X_n.$$

由于 X_i 只依赖于第 i 次试验,而各次试验是独立的,因此 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立;又因 X_i 服从同样的(0-1)分布即 $B(1,p)$ 分布,故有

$$EX_i=p,DX_i=p(1-p),\quad i=1,2,\cdots.$$

根据定理 5.1 或定理 5.2,得

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}(X_1+X_2+\cdots+X_n)-p\right|<\varepsilon\right\}=1,$$

即

$$\lim_{n\to\infty}P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1.$$

显然,伯努利大数定律是契比雪夫大数定律和辛钦大数定律的特殊情况.

伯努利大数定律说明在多次重复试验中,任一事件 A 发生的频率 $\frac{n_A}{n}$(随机变量序列)依概率收敛于事件 A 的概率 p(常数),即 n 很大时,事件发生的频率与概率有较大偏差的可能性很小.

伯努利大数定律从数学上严格地证明了事件频率的稳定性，与我们观察到的客观事实相一致.

伯努利大数定律应用非常广泛. 实际中事件 A 的概率 p 通常是未知的，根据伯努利大数定律，利用实际推断原理，我们可以对随机试验独立重复地观测多次，然后用事件 A 发生的频率去估计概率 p，就可以达到很好的效果. 在数理统计中我们将会看到，伯努利大数定律是估计随机变量分布函数的理论基础，因为分布函数就是事件的概率.

例 5.1　设$\{X_n\}$为相互独立的随机变量序列，且 $P\{X_n=\pm\sqrt{n}\}=\dfrac{1}{n}$，$P\{X_n=0\}=1-\dfrac{2}{n}$，$n=2,3,\cdots$. 证明$\{X_n\}$服从大数定律.

证明　由于

$$EX_n=0\cdot\left(1-\frac{2}{n}\right)+\sqrt{n}\cdot\frac{1}{n}+(-\sqrt{n})\cdot\frac{1}{n}=0,$$

$$DX_n=EX_n^2-[EX_n]^2=0^2\cdot\left(1-\frac{2}{n}\right)+n\cdot\frac{1}{n}+n\cdot\frac{1}{n}=2,$$

故$\{X_n\}$满足契比雪夫大数定律的条件，因此$\{X_n\}$服从大数定律，即

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=2}^{n+1}X_i-0\right|<\varepsilon\right\}=1.$$

或算术平均 $\dfrac{1}{n}\sum\limits_{i=2}^{n+1}X_i$ 依概率收敛到 0，亦即 $\dfrac{1}{n}\sum\limits_{i=2}^{n+1}X_i\xrightarrow{P}0$. 换言之，当 n 充分大时，随机变量 $\dfrac{1}{n}\sum\limits_{i=2}^{n+1}X_i$ 将几乎变为一个常数 0.

例 5.2　设$\{X_n\}$为独立同分布的随机变量序列，$E(X_i^k)=\alpha_k(i=1,2,\cdots)$存在且有限，其中 $k\geqslant 1$ 为整数. 证明 $\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k\xrightarrow[n\to\infty]{P}\alpha_k$.

证明　由于$\{X_n\}$为独立同分布的随机变量序列，因此 $X_1^k,X_2^k,\cdots,X_n^k,\cdots$也为独立同分布的随机变量序列，又期望 $EX_i^k=\alpha_k$ 存在且有限，故 $X_1^k,X_2^k,\cdots,X_n^k,\cdots$满足辛钦大数定律的条件.

记 $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$，根据辛钦大数定律，有 $A_k\xrightarrow[n\to\infty]{P}\alpha_k$，结论得证. 这说明当 n 充分大时，随机变量 $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$ 将几乎变为一个常数 α_k.

在数理统计中我们将会看到，用辛钦大数定律得到的这一结论是参数矩估计的理论基础.

例 5.3　设随机变量 $X_1,X_2,\cdots,X_n$ 是对随机变量 X 的 n 次(n 充分大)独立观测，若具体观察值为 $x_1,x_2,\cdots,x_n$，且这些观察值中有三分之一是小于或等于 0 的. 试估计概率 $P\{X\leqslant 0\}$.

解　因对随机变量 X 的 n 次观测是独立的，故 $X_1,X_2,\cdots,X_n$ 相互独立且跟 X 同分布. 记 $p=P\{X\leqslant 0\}$，根据伯努利大数定律，当 n 充分大时，事件 $A=\{X\leqslant 0\}$的概率 p 可用其频率来估计，即

$$P\{X\leqslant 0\}\approx\frac{1}{3}.$$

5.2　中心极限定理

本节将解决多个随机变量和的分布的极限问题. 在介绍正态分布时，我们提到在实际中有许多随机变量，它们受多种随机因素的影响，每种因素的影响所起的作用又很微小，这样的随机变量往往近似地服从正态分布. 这种现象就是中心极限定理的客观背景. 那么，我们能否严谨地从理论上对此加以证明呢？即在一定条件下，多个随机变量和的分布是否真的近似服从正态分布呢？本节介绍的三个常用的中心极限定理，将从理论上来回答这一问题.

定理 5.4(独立同分布中心极限定理)　设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立，服从同一分布，且具有数学期望和方差：$EX_i=\mu,DX_i=\sigma^2(\sigma^2>0),i=1,2,\cdots$，则随机变量之和 $\sum\limits_{i=1}^{n}X_i$ 的标准化随机变量

$$Y_n=\frac{\sum\limits_{i=1}^{n}X_i-E\left(\sum\limits_{i=1}^{n}X_i\right)}{\sqrt{D\left(\sum\limits_{i=1}^{n}X_i\right)}}=\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}$$

的分布函数处处收敛于标准正态分布的分布函数，即对任意 x，有

$$\lim_{n\to\infty}P\left\{\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}\leqslant x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt=\Phi(x). \tag{5.7}$$

证明从略.

这个定理通常也称为**列维-林德伯格**(Lévy-Lindeberg)**中心极限定理**.

这个定理说明，对于均值为 μ，方差为 σ^2 的独立同分布的随机变量的和 $\sum\limits_{i=1}^{n}X_i$ 的标准化随机变量，当 n 充分大时，有

$$Y_n=\frac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\overset{\text{近似}}{\sim}N(0,1).$$

或等价地有，当 n 充分大时，和 $\sum\limits_{i=1}^{n}X_i$ 近似服从正态分布，即

$$\sum_{i=1}^{n}X_i=\sqrt{n}\sigma\cdot Y_n+n\mu\overset{\text{近似}}{\sim}N(n\mu,n\sigma^2). \tag{5.8}$$

等价性由正态分布的线性性易得.

同理，可得更一般的结论：

$$a\sum_{i=1}^{n}X_i+b\overset{\text{近似}}{\sim}N(a\cdot n\mu+b,a^2\cdot n\sigma^2). \tag{5.9}$$

易知，跟正态分布一样，本节所有近似正态分布的参数恰是对应随机变量的期望和方差，由此我们可以很轻易地求得正态分布的参数.

在实际中，我们可以运用(5.8)式来求解多个随机变量和的分布函数及其相关概率问题. 记 $F_n(x)$ 为 $\sum\limits_{i=1}^{n}X_i$ 的分布函数，则 n 很大时(实际应用中，一般 $n\geqslant20$ 即可)，有如下的近似计

算分布函数和概率的公式：

$$P\left(\sum_{i=1}^{n} X_i \leqslant x\right) = F_n(x) \approx \Phi\left(\frac{x-n\mu}{\sqrt{n}\sigma}\right),$$

$$P\left(a < \sum_{i=1}^{n} X_i \leqslant b\right) = F_n(b) - F_n(a) \approx \Phi\left(\frac{b-n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right).$$

这时，只要查标准正态分布表就可以求出上述分布函数和概率相当精确的近似值，非常简便地解决了我们难以解决的多个随机变量和的确切分布的问题.

由(5.9)式还可推出

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right),$$

或

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0,1).$$

也就是说，均值为 μ，方差为 σ^2 的独立同分布的随机变量和的算术平均值 $\frac{1}{n}\sum\limits_{i=1}^{n} X_i$ 当 n 充分大时近似地服从均值为 μ，方差为 $\frac{\sigma^2}{n}$ 的正态分布. 这一结论在数理统计的大样本统计推断中很有用.

下面介绍一个虽是定理 5.4 的特例却应用非常广泛的中心极限定理.

定理 5.5(德莫佛-拉普拉斯(De Moivre-Laplace)中心极限定理)　设在 n 重伯努利试验中，事件 A 在每次试验中出现的概率为 $p(0<p<1)$，η_n 为 n 次试验中事件 A 出现的次数，且服从二项分布 $B(n,p)$，则 η_n 的标准化随机变量的分布函数处处收敛于标准正态分布的分布函数，即对 $\forall x$，有

$$\lim_{n\to\infty} P\left(\frac{\eta_n - np}{\sqrt{np(1-p)}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t. \tag{5.10}$$

证明　由定理 5.3 的证明知，η_n 可以表示为 n 个独立的服从 $B(1,p)$ 分布的随机变量 X_i $(i=1,2,\cdots,n)$之和，即

$$\eta_n = X_1 + X_2 + \cdots + X_n,$$

其中 $X_i(i=1,2,\cdots,n)$的分布列为

$$P(X_i = k) = \mathrm{C}_1^k p^k (1-p)^{1-k} = p^k(1-p)^{1-k}, \quad k=0,1.$$

又

$$E(X_i) = p, \quad D(X_i) = p(1-p) > 0, \quad i=1,2,\cdots,n.$$

由独立同分布的中心极限定理，有

$$\lim_{n\to\infty} P\left(\frac{\eta_n - np}{\sqrt{np(1-p)}} \leqslant x\right) = \lim_{n\to\infty} P\left(\frac{\sum\limits_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

此定理表明，二项分布的极限分布是正态分布，即若 $\eta_n \sim B(n,p)$，则当 n 很大时，有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1),$$

或等价地有

$$\eta_n \overset{\text{近似}}{\sim} N(np, np(1-p)). \tag{5.11}$$

等价性是由正态分布的线性性，以及正态分布的两个参数分别是随机变量的期望和方差得到

的，因 $E(\eta_n)=np, D(\eta_n)=np(1-p)$.

由(5.11)式，我们可以得到 n 很大时二项分布的分布函数及其相关概率的近似计算公式：

$$P(\eta_n\leqslant x)=F_n(x)\approx\Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right),$$

$$P(a<\eta_n\leqslant b)=F_n(b)-F_n(a)\approx\Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right).$$

其中 $F_n(x)$ 代表 η_n 的二项分布 $B(n,p)$ 的分布函数. 虽然由二项分布的分布律

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

可以精确地求得相关事件的概率，但是利用中心极限定理计算要简便得多.

下面我们分析一下德莫佛-拉普拉斯中心极限定理和伯努利大数定律的关系. 二者都是对二项分布极限的一种描述. 伯努利大数定律指出，当 n 趋于无穷时，频率 $\frac{\eta_n}{n}$ 依概率收敛于概率 p，即对于任给 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{\eta_n}{n}-p\right|<\varepsilon\right\}=1.$$

但伯努利大数定律却没能给出依概率收敛的速度. 而德莫佛-拉普拉斯中心极限定理却解决了这一问题，给出了 n 很大时上述概率的近似值，并可由其证得伯努利大数定律成立：

$$\begin{aligned}P\left(\left|\frac{\eta_n}{n}-p\right|<\varepsilon\right)&=P(n(p-\varepsilon)<\eta_n<n(p+\varepsilon))=F_n(n(p+\varepsilon))-F_n(n(p-\varepsilon))\\&\approx\Phi\left(\frac{n(p+\varepsilon)-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{n(p-\varepsilon)-np}{\sqrt{np(1-p)}}\right)\\&=\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)-\Phi\left(-\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)=2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)-1\xrightarrow{n\to\infty}1.\end{aligned}$$

由此还可得

$$P\left(\left|\frac{\eta_n}{n}-p\right|>\varepsilon\right)\approx1-\left[2\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)-1\right]=2\left[1-\Phi\left(\varepsilon\sqrt{\frac{n}{p(1-p)}}\right)\right]\xrightarrow{n\to\infty}0.$$

可见，德莫佛-拉普拉斯中心极限定理比伯努利大数定律的结论更精确更强大，也更实用. 类似地，可以分析独立同分布中心极限定理与辛钦大数定律的关系.

定理 5.4、定理 5.5 在独立同分布的条件下解决了随机变量和的分布的极限问题，下面研究各个加项为独立非同分布的随机变量和的分布的极限问题.

定理 5.6(李雅普诺夫(Liapunov)中心极限定理)　设 $X_1,X_2,\cdots,X_n,\cdots$ 是相互独立的随机变量序列，它们的期望和方差存在：

$$EX_i=\mu_i,\quad DX_i=\sigma_i^2\neq0,\quad i=1,2,\cdots,$$

记

$$B_n^2=\sum_{i=1}^n\sigma_i^2.$$

若存在 $\delta>0$，使

$$\frac{1}{B_n^{2+\delta}}\sum_{i=1}^nE[|X_i-\mu_i|^{2+\delta}]\to0,\quad n\to\infty,$$

则随机变量之和 $\sum\limits_{i=1}^nX_i$ 的标准化随机变量

$$Z_n = \frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{D\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n}$$

的分布函数处处收敛于标准正态分布的分布函数，即对任意的 x，有

$$\lim_{n\to\infty} P\left\{\frac{\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)}{B_n} \leqslant x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

证明从略.

定理 5.6 表明，在定理的条件下，当 n 充分大时，随机变量之和 $\sum_{i=1}^{n} X_i$ 的标准化随机变量近似服从标准正态分布，即

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n} \overset{\text{近似}}{\sim} N(0,1),$$

或等价地有

$$\sum_{i=1}^{n} X_i = B_n \cdot Z_n + \sum_{i=1}^{n} \mu_i \overset{\text{近似}}{\sim} N\left(\sum_{i=1}^{n} \mu_i, B_n^2\right).$$

这就是说，无论各个随机变量 $X_i(i=1,2,\cdots,)$ 服从什么分布，只要满足定理的独立性和矩的条件，那么当 n 很大时，它们的和 $\sum_{i=1}^{n} X_i$ 就近似地服从正态分布. 这就是为什么正态分布在概率论中占有重要地位的根本原因之一.

该定理的结论对一般正态分布的客观应用背景进行了严格的数学论证. 在很多实际问题中，若所考虑的随机变量都可以表示成很多个独立的随机变量之和，则它们就可用正态分布来刻画. 例如，一个物理实验的测量误差是由许多观察不到的、可加的微小误差所合成的，则合成的总误差就近似地服从正态分布. 正态分布的其他应用实例都可与此类似地得到合理的解释.

在数理统计中我们将会看到，中心极限定理是大样本统计推断的理论基础.

下面举几个关于独立同分布中心极限定理和德莫佛－拉普拉斯中心极限定理的应用实例.

例 5.4　一加法器同时收到 20 个噪声电压 $V_i(i=1,2,\cdots,20)$，设它们是相互独立的随机变量，且都在区间 $(0,10)$ 内服从均匀分布. 记加法器受到的总噪声电压 $V=\sum_{i=1}^{20} V_i$，求 V 的分布函数及 $P(V>105)$ 的近似值.

解　从第三章随机变量和的分布计算可知，要求得 V 的分布函数及其相关概率是何等的困难. 现用独立同分布中心极限定理求解. 因 $EV_i=5, DV_i=100/12, i=1,2,\cdots,20$，故

$$V=\sum_{i=1}^{20} V_i \overset{\text{近似}}{\sim} N(EV, DV),$$

即

$$V=\sum_{i=1}^{20} V_i \overset{\text{近似}}{\sim} N(20\times 5, 20\times 100/12).$$

于是 V 的分布函数

$$F_V(v)\approx\Phi\left(\frac{v-20\times5}{\sqrt{20\times100/12}}\right)=\Phi\left(\frac{v-100}{\sqrt{500/3}}\right),$$

概率
$$P(V>105)=1-P(V\leqslant105)=1-F_V(105)$$
$$\approx1-\Phi\left(\frac{105-100}{\sqrt{500/3}}\right)\approx1-\Phi(0.39)=0.3483.$$

例 5.5　设电话总机共有 200 个电话分机，若每个分机都有 5%的时间要用外线，且是否使用外线相互独立，要保证每个用户有 95%以上的把握接通外线，问总机至少要设置多少条外线？

解　设 X 为某时刻使用外线的用户数(分机数). 将各个电话分机是否使用外线看做是一次试验，则各次试验是独立的，因此 $X\sim B(n,p)$，其中 $n=200,p=0.05$.

设应设置 N 条外线，依题意，要求最小的整数 N，使得
$$P(X\leqslant N)\geqslant0.95.$$
用二项分布计算非常复杂，现用德莫佛-拉普拉斯中心极限定理求解. 因 $X\overset{\text{近似}}{\sim}N(np,np(1-p))$，故
$$P(X\leqslant N)=F_X(N)\approx\Phi\left(\frac{N-np}{\sqrt{np(1-p)}}\right),$$
其中 $F_X(x)$ 为 X 的分布函数. 反查标准正态分布表得 $\Phi(1.645)=0.95$，故 N 应满足
$$\frac{N-200\times0.04}{\sqrt{200\times0.05\times0.95}}\geqslant1.645,$$
解出 $N\geqslant15.1$，故取 $N=16$. 即总机至少应备 16 条外线，才能有 95%以上的把握保证每个用户随时有 95%以上的把握接通外线.

例 5.6　设某天文学家对某星球与天文台的距离进行的 n 次独立观测为 $X_1,X_2,\cdots,X_n$(单位：光年)，期望 $EX_i=D$，方差 $DX_i=4,i=1,2,\cdots,n$. 天文学家用 $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ 作为 D 的估计，为使估计的误差在±0.25 光年之间的概率大于 0.98，问他至少要作多少次独立观测？

解　由题设知 $X_1,X_2,\cdots,X_n$ 独立同分布，又
$$E\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}EX_i=D,\quad D\overline{X}_n=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{4}{n},$$
由独立同分布中心极限定理立得
$$\overline{X}_n\overset{\text{近似}}{\sim}N\left(D,\frac{4}{n}\right).$$
记 $\overline{X}_n$ 的分布函数为 $F_n(x)$. 依题意，要求最小的正整数 n 使得
$$P(-0.25\leqslant\overline{X}_n-D\leqslant0.25)\geqslant0.98\Leftrightarrow P(D-0.25\leqslant\overline{X}_n\leqslant D+0.25)\geqslant0.98,$$
即
$$P(D-0.25\leqslant\overline{X}_n\leqslant D+0.25)=F_n(D+0.25)-F_n(D-0.25)$$
$$\approx\Phi\left(\frac{D+0.25-D}{\sqrt{4/n}}\right)-\Phi\left(\frac{D-0.25-D}{\sqrt{4/n}}\right)=2\Phi(\sqrt{n}/8)-1\geqslant0.98,$$
亦即
$$\Phi(\sqrt{n}/8)\geqslant0.99,$$
反查标准正态分布表得 $\Phi(2.33)=0.99$，故 n 应满足 $\sqrt{n}/8\geqslant2.33$，解出 $n\geqslant347.45$，即这位天文学家至少要作 348 次独立观测才能满足误差的要求.

例 5.7　从次品率为 0.05 的一大批产品中随机地取 200 件产品,分别用二项分布、泊松分布和德莫佛-拉普拉斯中心极限定理计算取出的产品中至少有 3 件次品的概率.

解　设 X 表示取出的 200 件产品中的次品数,由题意 $X \sim B(200, 0.05)$.

(1) 用二项分布计算.

$$\begin{aligned} P(X \geqslant 3) &= 1 - P(X < 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \sum_{k=0}^{2} \mathrm{C}_{200}^{k} 0.05^{k} 0.95^{200-k} \approx 0.9996. \end{aligned}$$

(2) 用泊松分布 $X \overset{\text{近似}}{\sim} P(200 \times 0.05)$近似计算.

$$P(X \geqslant 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - \sum_{k=0}^{2} \frac{10^k}{k!} \mathrm{e}^{-10} \approx 0.9992 .$$

(3) 用德莫佛-拉普拉斯中心极限定理近似计算.

$$P(X \geqslant 3) = 1 - P(X \leqslant 2) = 1 - F_X(2) \approx 1 - \Phi\left(\frac{2-10}{\sqrt{9.5}}\right) \approx \Phi(2.6) \approx 0.9953.$$

注　对 $X \sim B(n,p)$且 n 比较大的情形,用二项分布求解,可求得概率的精确值,但计算繁琐,尤其是 X 的取值范围较大时;用泊松分布求解,近似精度较高,但要求 p 很小,且 X 的取值范围较大时计算也繁琐;用中心极限定理求解,则计算简单精度也较高,只要 p 不是很小或很大的情形,其近似效果比泊松近似更好.三种方法比较而言,中心极限定理最简单实用.

例 5.8　设随机变量 $X_1, X_2, \cdots, X_n$ 独立同分布,已知 $EX_i^k = \alpha_k (k=1,2,3,4), i=1,2,\cdots,n$,且 $\alpha_4 > \alpha_2^2$.证明当 n 充分大时,随机变量

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

近似服从正态分布.

证明　因 $X_1, X_2, \cdots, X_n$ 独立同分布,故 $X_1^2, X_2^2, \cdots, X_n^2$ 也独立同分布.由题设有

$$EX_i^2 = \alpha_2, \quad DX_i^2 = EX_i^4 - [EX_i^2]^2 = \alpha_4 - \alpha_2^2 > 0, \quad i = 1,2,\cdots,n.$$

于是

$$EZ_n = \frac{1}{n} \sum_{i=1}^{n} EX_i^2 = \alpha_2, \quad DZ_n = \frac{1}{n^2} \sum_{i=1}^{n} DX_i^2 = \frac{1}{n}(\alpha_4 - \alpha_2^2).$$

根据独立同分布中心极限定理,当 n 充分大时有

$$Z_n \overset{\text{近似}}{\sim} N(EZ_n, DZ_n),\text{即 } Z_n \overset{\text{近似}}{\sim} N\left(\alpha_2, \frac{1}{n}(\alpha_4 - \alpha_2^2)\right).$$

小结　中心极限定理是本章的核心内容.在一般情况下,n 较大时很难求出 n 个随机变量之和的确切分布.中心极限定理表明:当随机变量序列满足一定的条件时,随着 n 无限增大,序列的前 n 个随机变量之和的标准化随机变量依概率收敛于标准正态分布,而标准正态分布又有表可查.这样多个随机变量之和的概率问题就迎刃而解了.本章重点要求掌握独立同分布中心极限定理和德莫佛-拉普拉斯中心极限定理的应用.

中心极限定理主要用来解决三类问题:① 已知多个随机变量和的数目,求其落在某个范围内的概率;② 已知多个随机变量之和落在某给定范围内的概率,求和的数目;③ 已知多个随机变量和的数目和概率,求其取值的范围.简言之,多个随机变量和的数目、取值范围及其落在相应范围内的概率三个量中,只要已知其中两个,就可用中心极限定理求得第三个.

习　题　五

5.1　填空题.

(1) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $EX_i=\mu, DX_i=\sigma^2\neq 0, i=1,2,\cdots,n$，则 $\lim\limits_{n\to\infty}P\left(\dfrac{\sum\limits_{i=1}^{n}X_i-n\mu}{\sqrt{n}\sigma}\leqslant x\right)=$______.

(2) 若种子的出苗率为 0.6，则 10000 粒种子中出苗数在 5900～6100 之间的概率为______.

5.2　选择题.

(1) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，$EX_i=0, DX_i=\sigma^2, i=1,2,\cdots$. 则对任意 $\varepsilon>0$，有(　　)成立.

(A) $\lim\limits_{n\to\infty}P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i-\sigma^2\right|<\varepsilon\right)=1$　　(B) $\lim\limits_{n\to\infty}P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i-\sigma^2\right|<\varepsilon\right)=0$

(C) $\lim\limits_{n\to\infty}P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\sigma^2\right|<\varepsilon\right)=1$　　(D) $\lim\limits_{n\to\infty}P\left(\left|\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\sigma^2\right|<\varepsilon\right)=0$

(2) 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，$S_n=\sum\limits_{i=1}^{n}X_i$，由列维-林德伯格中心极限定理知，当 n 充分大时，S_n 近似服从正态分布，只要(　　).

(A) 有相同的数学期望　　(B) 有相同的方差

(C) 服从同一指数分布　　(D) 服从同一离散型分布

(3) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且皆服从参数为 $\lambda(\lambda>0)$ 的泊松分布，则正确的是(　　).

(A) $\lim\limits_{n\to\infty}P\left(\dfrac{\sum\limits_{i=1}^{n}X_i-n\lambda}{\sqrt{n\lambda}}\leqslant x\right)=\Phi(x)$

(B) 当 n 充分大时，$\sum\limits_{i=1}^{n}X_i$ 近似服从 $N(n\lambda, n\lambda)$

(C) 当 n 充分大时，$\sum\limits_{i=1}^{n}X_i$ 近似服从标准正态分布

(D) $\lim\limits_{n\to\infty}P\left(\dfrac{\sum\limits_{i=1}^{n}X_i-\lambda}{\sqrt{n\lambda}}\leqslant x\right)=\Phi(x)$

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

(4) 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $X_i\sim B(1,p)$，$0<p<1$，$i=1,2,\cdots,100$，则正确的是(　　).

(A) $\dfrac{1}{100}\sum\limits_{i=1}^{100}X_i\approx p$

(B) $\sum\limits_{i=1}^{100}X_i\sim B(100,p)$

(C) $P(a<\sum\limits_{i=1}^{100}X_i<b)\approx\Phi(b)-\Phi(a)$

(D) $P(a<\sum\limits_{i=1}^{100}X_i<b)\approx\Phi\left(\dfrac{b-100p}{\sqrt{np(1-p)}}\right)-\Phi\left(\dfrac{a-100p}{\sqrt{np(1-p)}}\right)$

(5) 已知 X_i 的概率密度为 $f(x_i)$，$i=1,2,\cdots,100$，它们相互独立，则对任何实数 x，概率 $P\left(\sum_{i=1}^{100} X_i \leqslant x\right)$(　　).

(A) 必能用中心极限定理计算出近似值　　(B) 必不能用中心极限定理计算出近似值

(C) $\int\limits_{\sum_{i=1}^{100} x_i \leqslant x} \cdots \int \left[\prod_{i=1}^{100} f(x_i)\right] \mathrm{d}x_1 \cdots \mathrm{d}x_{100}$　　(D) 无法计算

5.3　螺钉的重量是随机变量，其均值是 50 g，标准差是 5 g，求一盒螺钉(100 个装)的重量超过 5100 g 的概率.

5.4　某粮仓内老鼠数目服从参数为 2 的泊松分布，求 200 间同类粮仓中老鼠总数超过 350 只的概率.

5.5　某大型商场每天接待顾客 10000 人，设每位顾客的日消费额(单位:元)服从[100,1000]上的均匀分布，且不同顾客的消费额是相互独立的. 试求该商场的日消费额(单位:元)在平均销售额上下浮动不超过 20000 元的概率.

5.6　在一家零售商店中，其结帐柜台替各顾客服务的时间(单位:min)是相互独立的随机变量，均值为 1.5，方差为 1.(1) 求对 100 位顾客服务的总时间不多于 2 h 的概率.(2) 若离下班还剩 1 h，问 95%以上的概率能确保柜台至多还可对几位顾客提供结账服务?

5.7　某工厂生产电阻，正常情况下废品率为 0.01，现取 500 个装成一盒，求每盒中废品数不超过 5 个的概率.

5.8　某种难度很大的心脏手术成功率为 0.9，对 100 个病人进行这种手术，以 X 记手术成功的人数. (1) 求 $P(84 \leqslant X \leqslant 96)$；(2) 求 $P(X \geqslant 90)$.

5.9　一个供电网内共有 10000 盏功率相同的灯，夜晚每盏灯开着的概率都是 0.7，假设各盏灯开、关彼此独立. 求夜晚同时开着的灯数在 6900 到 7100 之间的概率.

5.10　在一家保险公司有 1 万人参加寿险，每人每年需支付 120 元保险费. 设一年内一个人死亡的概率为 0.003，死亡时保险公司需赔付其家属 2 万元. 问一年内保险公司亏本的概率及盈利不少于 40 万元的概率各是多少?

5.11　某工厂有 200 台机器，各台机器工作与否相互独立，每台机器的开工率为 0.6，工作时各需 1 kW 电力. 问供电局至少需要供应多少电力才能以 99%的把握保证工厂不会因供电不足而影响生产?

5.12　某商店负责供应某地区 1000 个人所需的商品，某商品在一段时间内每人需买一件的概率为 0.6，假如在这段时间内，每人购买与否彼此无关，问商店应准备多少件这种商品，才能以 99.7%的概率保证不会脱销?(假定该商品在该时间段内每人最多购买一件)

5.13　计算器进行加法运算时，把每个加数取为最接近于它的整数来计算. 设所有的取整误差是独立的且在[−0.5,0.5]上服从均匀分布. (1) 若将 1500 个数相加，问误差总和超过 15 的概率是多少? (2) 最多有几个数相加可使得误差总和的绝对值小于 10 的概率不小于 0.90?

5.14　某产品包含 100 个部件，每个部件的长度是一个随机变量，它们独立同分布，且均值为 2 mm，标准差为 0.05 mm，若规定总长度为 200 mm±0.5 mm 时产品合格，试求产品的合格率.

5.15　某生产线生产的产品成箱包装，每箱的重量是随机的，假设每箱平均重量为 50 kg，标准差为 5 kg. 若用最大载重量为 5 t 的汽车承运，试用中心极限定理说明每辆车最多可以装多少箱，才能使不超载的概率大于 0.9972.($\Phi(2)=0.9972$，其中 $\Phi(x)$ 是标准正态分布函数)

5.16　(1) 一个复杂系统由 100 个相互独立的元件组成，在系统运行期间每个元件损坏的概率为 0.10. 为使系统正常运行，至少需有 85 个元件工作，求系统的可靠性(即正常运行的概率). (2) 上述系统假如由 n 个相互独立的部件组成，且必需至少有 80%以上的部件工作系统才能正常运行. 问 n 至少为多大时才能使系统的可靠性不低于 0.95.

5.17　某心理学家研究一群孩子的智商的均值 μ，他用 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$ 作为 μ 的估计，其中 $X_1, X_2, \cdots, X_n$

是对其中 n 个孩子智商测试的结果.若 $EX_i=\mu,DX_i=263.66,i=1,2,\cdots,n$. 为使 $\overline{X}$ 对 μ 的估计误差在 ± 3 之间的概率不小于 0.95,问他至少要测试多少孩子?

5.18 投掷一枚骰子,为了至少有 95%的把握确保出现点 2 的频率与概率之间的误差在 ± 0.01 的范围之内,问需要投掷多少次.

5.19 有一批种子,其中良种占 $\frac{1}{6}$,从中任取 6000 粒,求数 ε 使得 6000 粒中的良种比例与 $\frac{1}{6}$ 之间的绝对误差不超过 ε 的概率为 0.99,并依此确定能以 0.99 的概率保证的良种数的范围.

5.20 设 μ_n 服从二项分布 $B(n,p)$,参数 n 充分大,$k(0\leqslant k\leqslant n)$ 是非负整数,且和 $\alpha(0<\alpha<1)$ 满足 $P(\mu_n\geqslant k)=\alpha$.(1) 对于给定的 n,k,求 α;(2) 对于给定的 n,α,求 k.

5.21 设随机变量 $X_1,X_2,\cdots,X_n$ 独立同分布,且 X_i 服从区间 $(-1,1)$ 内的均匀分布,$i=1,2,\cdots,n$,试证当 n 充分大时,随机变量 $Z_n=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ 近似服从正态分布,并指出其分布参数.

第六章　数理统计的基本概念

前面五章我们讨论了概率论的基本内容.我们知道概率空间及其上的随机变量全面地描述了随机现象的统计规律性.在概率论的问题中,概率分布通常是已知的,我们所进行的计算推理都是基于这个已知的分布.但在实际的应用问题中,我们往往并不确切地知道分布.先来看下面几个具体的例子.

(1) 某灯泡厂生产了一批灯泡,作为质量检测部门如何来确定这批灯泡是否符合产品质量标准?

(2) 某保险公司决定新开展一项自然灾害保险业务,由于精算的需要,要知道灾害发生的概率及造成损失的概率分布,等等,如何得到这些内容?

(3) 某医药公司开发了一种新的治疗心血管疾病的药物,并进行了临床试验,药监部门如何判断该药物的疗效?

(4) 中央电视台为确定节目的广告投放量及收费标准,需要进行收视调查,该如何进行调查才能保证调查结果相对客观准确,进一步又该如何分析得到的调查结果?

以上问题的解决需要用到本书中另外一个重要内容——数理统计.数理统计是一门应用广泛、内容丰富的学科.在科技飞速发展的今天,它是应用于社会经济、工农业生产和科学试验等诸多领域中必不可少的工具之一.

数理统计以概率论为理论基础,根据试验或观测得到的数据来研究随机现象,对研究对象的统计规律性作出种种合理的估计和推断.数理统计的内容非常丰富,由于篇幅的限制,本书只介绍参数估计、假设检验、方差分析和回归分析等部分内容.

本章我们介绍总体、随机样本及统计量等基本概念,并着重讨论几个常用统计量及抽样分布.

6.1　总体与样本

6.1.1　总体与个体

我们把研究对象的全体所组成的集合称为**总体**(或母体),组成总体的每个元素称为**个体**.例如,在本章前面的例子中,我们最想了解灯泡的寿命,灯泡厂所生产的整批灯泡的寿命就是总体,而每个灯泡的寿命就是该总体中的个体.

在数理统计中,我们常常关心并研究的是总体中各个个体的某一项(或某几项)数量指标及该指标在总体中的分布情况.用 X 表示数量指标,则指标值是一个随个体不同而变化的量.由于在统计问题中,从总体中抽取个体是随机抽取的,因此,我们可以从数学上将 X 视为一个随机变量,总体可以为随机变量 X 所有可能取值的全体,个体就是其中的一个具体值,而随机变量 X 的分布就完全描述了总体中所研究的数量指标的分布情况.今后,我们把总体与数量

指标 X 可能取值的全体所组成的集合等同起来,用随机变量 X 表示,总体的分布就是指数量指标 X 的分布.

总体作为一个随机变量有一维与多维、连续型与离散型之分;而作为一个集合,则还有有限总体与无限总体之分.

6.1.2 简单随机样本

从总体中抽取部分个体的手续(过程)叫做抽样,抽样的目的是为了研究总体的性质.为了使抽取的部分能够较好地反映总体的特性,从理论上来说,抽样方法必须满足以下基本要求:

(1) 随机性,即对于每一次抽样,总体中每个个体都有同等的机会被抽取;

(2) 独立性,即每次抽取的结果互相之间没有影响.

满足以上两个条件的抽样称为**简单随机抽样**,简称为**随机抽样**或**抽样**.对于有限总体,采用有放回抽样就能做到这一点.当然,在实际应用中,由于操作起来不方便或者条件的限制,我们一般很难做到简单随机抽样.当个体总数 N 比抽取次数 n 大得多时$\left(\text{一般当}\dfrac{N}{n}\geqslant 10\text{ 时}\right)$,我们可以将无放回抽样近似地作为有放回抽样来处理.有关抽样的技术是数理统计中一个非常重要的内容,有兴趣的读者可以参阅有关文献.为讨论方便起见,本书中所提到的抽样都是指简单随机抽样.

一次抽样试验的结果成为总体 X 的一个观测值,随机抽样 n 次就得到总体 X 的 n 个观测值 $x_1,x_2,\cdots,x_n$,其中 x_i 是第 i 次抽样观测的结果.显然,由于抽样的随机性和独立性,如果再抽 n 次,则会得到另外一组观测;如果不断地重复这一做法,则会得到许多组不同的观测.可见,就一次抽样观测而言,$x_1,x_2,\cdots,x_n$ 是一组确定的数,但它又随着每一次的抽样观测而变化.因而从数学上我们可以认为 n 次抽样就与 n 个随机变量 $X_1,X_2,\cdots,X_n$ 对应,n 次抽样所得到的结果实际上就是这个 n 维随机变量的观测值.

综上所述,我们有如下定义.

定义 6.1　设总体 X 的分布函数为 $F(x)$,若 $X_1,X_2,\cdots,X_n$ 是相互独立且与总体 X 同分布的随机变量,则称$(X_1,X_2,\cdots,X_n)$是总体 X(或 $F(x)$)的容量为 n 的**简单随机样本**,简称**样本**.当$(X_1,X_2,\cdots,X_n)$取定某组常数值$(x_1,x_2,\cdots,x_n)$(其中 X_i 取值 x_i)时,称这组常数值$(x_1,x_2,\cdots,x_n)$为样本$(X_1,X_2,\cdots,X_n)$的一组样本观测值(或样本实现).

6.1.3 理论分布与经验分布函数

我们把总体 X 的分布称为**理论分布**,而把总体 X 的分布函数称为**理论分布函数**.

若总体 X 的分布函数为 $F(x)$,$(X_1,X_2,\cdots,X_n)$是总体 X 的容量为 n 的样本,则由样本的独立同分布性质知道$(X_1,X_2,\cdots,X_n)$的分布函数为

$$F(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F(x_i). \tag{6.1}$$

若总体 X 是离散型,其概率分布为 $P(x)=P(X=x)$,则样本$(X_1,X_2,\cdots,X_n)$的概率分布为

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(x_i). \tag{6.2}$$

若总体 X 是连续型，其概率密度为 $f(x)$，则 $(X_1, X_2, \cdots, X_n)$ 的概率密度为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i). \tag{6.3}$$

例 6.1　(1) 设总体 X 服从两点分布 $B(1, p)$，其分布列为

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

则样本 $(X_1, X_2, \cdots, X_n)$ 的概率分布列为

$$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)= p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n-\sum\limits_{i=1}^{n} x_i}, \quad x_i=0,1, \quad i=1,\cdots,n.$$

(2) 设总体 X 服从正态分布 $N(\mu, \sigma^2)$，则样本 $(X_1, X_2, \cdots, X_n)$ 的概率密度为

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right\}.$$

总之，若总体 X 的理论分布已知，则其样本的联合分布可以确定. 数理统计中的问题恰好反过来，如何由样本来推断总体的分布？为什么能够由样本来推断总体的分布？为此我们引入经验分布函数的概念.

定义 6.2　设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，将其观测值 $(x_1, x_2, \cdots, x_n)$ 按由小到大的顺序排列为 $x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_n^*$，令

$$F_n(x)=\begin{cases} 0, & x<x_1^*, \\ \dfrac{k}{n}, & x_k^* \leqslant x<x_{k+1}^* \ (k=1,2,\cdots,n-1), \\ 1, & x \geqslant x_n^*, \end{cases} \tag{6.4}$$

称 $F_n(x)$ 为 X 的**经验分布函数**.

应当注意，经验分布函数是依赖于样本观测值的，由于样本的随机性，因而 $F_n(x)$ 具有随机性.

根据经验分布函数的构造，易见对给定的一组观测值 $F_n(x)$ 满足分布函数的所有性质. 实际上，$F_n(x)$ 是在该组观测值中随机事件 $\{X \leqslant x\}$ 发生的频率. 根据大数定律我们知道，事件发生的频率依概率收敛于这个事件发生的概率. 因此我们可以用经验分布函数 $F_n(x)$ 来估计总体 X 的理论分布函数 $F(x)$. Glivenko 于 1933 年从理论上严格证明了以下的结论.

定理 6.1(Glivenko 定理)　设总体 X 的理论分布函数为 $F(x)$，经验分布函数为 $F_n(x)$，则当 $n \to \infty$ 时，$F_n(x)$ 以概率 1 关于 x 均匀地收敛于 $F(x)$，即

$$P\{\lim_{n\to\infty} \sup_{-\infty<x<+\infty} |F_n(x)-F(x)|=0\}=1.$$

根据该定理，当 n 充分大时，经验分布函数 $F_n(x)$ 是总体 X 的理论分布函数 $F(x)$ 的一个很好的近似，这就是我们之所以能用样本来推断总体的依据.

图 6.1 和图 6.2 中的曲线是我们根据实际数据绘制的前面例子中灯泡寿命的分布函数 $F(x)$ 与经验分布函数 $F_{20}(x)$ 和 $F_{100}(x)$ 的图形. 可以看到，对于不同的样本，得到的经验分布函数不同，但它们都是总体理论分布函数的近似，而且样本容量愈大，近似程度愈好.

6.1.4　统计量和样本矩

样本取自于总体，含有总体的信息，是统计推断的基础. 为了对总体特征进行种种推断，需要对样本进行数学上的“加工处理”，使样本所含的信息更加集中，这个过程往往是通过构造一个合适的依赖于样本的函数——统计量来实现的.

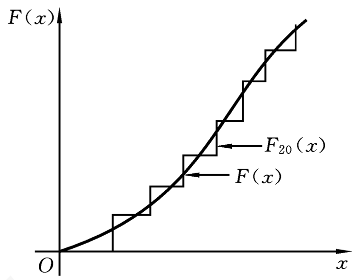

图 6.1

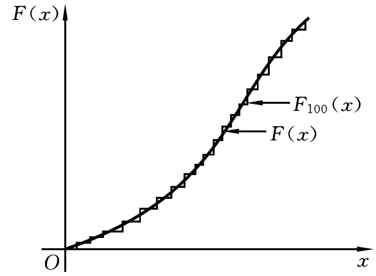

图 6.2

定义 6.3　设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的样本,$g(x_1,x_2,\cdots,x_n)$为一连续函数,且 g 中不含任何未知参数,则称 $g(X_1,X_2,\cdots,X_n)$为**统计量**.

由定义知,统计量是随机变量.若$(x_1,x_2,\cdots,x_n)$是样本$(X_1,X_2,\cdots,X_n)$的观测值,则 $g(x_1,x_2,\cdots,x_n)$是统计量 $g(X_1,X_2,\cdots,X_n)$的观测值.

下面我们来介绍一些数理统计中常用的统计量.

(1) 样本均值:$\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$.

(2) 样本方差:$S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$;

样本标准差:$S=\sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$.

与总体均值 EX 和方差 DX 一样,样本均值刻画了样本观测值的集中位置,样本方差刻画了样本观测值对样本均值的散离程度.

(3) 样本 k 阶原点矩:$A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$,　$k=1,2,\cdots$.

(4) 样本 k 阶中心矩:$B_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k$,　$k=1,2,\cdots$.

(5) 顺序统计量:设$(x_1,x_2,\cdots,x_n)$是样本$(X_1,X_2,\cdots,X_n)$的一组观测值,将它们按由小到大的顺序排列为 $x_1^*\leqslant x_2^*\leqslant\cdots\leqslant x_n^*$,引入 n 个函数 $g_i(x_1,x_2,\cdots,x_n)=x_i^*$,$i=1,2,\cdots,n$.记 $X_i^*=g_i(X_1,X_2,\cdots,X_n)$,$i=1,2,\cdots,n$,则我们得到 n 个新的统计量 $X_1^*,X_2^*,\cdots,X_n^*$,称 $X_1^*,X_2^*,\cdots,X_n^*$ 为总体 X 的一组顺序统计量,$X_i^*(i=1,2,\cdots,n)$称为第 i 位顺序统计量.当$(X_1,X_2,\cdots,X_n)$取值$(x_1,x_2,\cdots,x_n)$时,X_i^* 取值 x_i^*,$i=1,2,\cdots,n$.

由定义知,$X_1^*\leqslant X_2^*\leqslant\cdots\leqslant X_n^*$,且 $X_1^*=\min\{X_1,X_2,\cdots,X_n\}$,$X_n^*=\max\{X_1,X_2,\cdots,X_n\}$,即 X_1^* 的观测值是样本观测$(x_1,x_2,\cdots,x_n)$中最小的那一个,而 X_n^* 的观测值是样本观测$(x_1,x_2,\cdots,x_n)$中最大的那一个.

(6) 样本中位数:$\widetilde{X}=\begin{cases}X_{m+1}^*, & n=2m+1,\\ \dfrac{1}{2}(X_m^*+X_{m+1}^*), & n=2m.\end{cases}$

(7) 样本极差:$R=X_n^*-X_1^*$.

若$(x_1,x_2,\cdots,x_n)$是样本$(X_1,X_2,\cdots,X_n)$的观测值,则以上这些统计量的观测值我们相应地分别记为 $\bar{x}$,s^2,…等等,还是把它们叫做样本均值,样本方差,…等等.在实际应用中,我们常常用带有统计功能的计算器或用统计软件如 R,SAS,Excel 等来计算这些观测值.

6.2 抽样分布

统计量是随机变量，统计量的分布称为**抽样分布**. 理论上，只要知道总体的分布就可以求出统计量的分布；但在一般情况下，想求出统计量的分布是相当困难的，本书仅就总体为正态分布时，给出有关抽样分布的结果. 下面介绍数理统计中常见的抽样分布.

6.2.1 χ^2 分布

定义 6.4 设$(X_1,X_2,\cdots,X_n)$是正态总体 $N(0,1)$的样本，称统计量

$$\chi^2=X_1^2+X_2^2+\cdots+X_n^2$$

所服从的分布为自由度是 n 的 χ^2 **分布**，记为 $\chi^2\sim\chi^2(n)$.

χ^2 分布中的自由度可以理解为平方和中独立随机变量的个数. 可以算得 $\chi^2(n)$分布的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x\geqslant 0,\\ 0, & x<0.\end{cases}\tag{6.5}$$

图 6.3 给出了 $n=1,4,10,20$ 时，$\chi^2(n)$的概率密度函数曲线.

χ^2 分布具有以下性质.

(1) 若 $\chi^2\sim\chi^2(n)$，则 $E\chi^2=n, D\chi^2=2n$. 根据期望的线性性及方差的性质，我们有

$$E\chi^2=E\left(\sum_{i=1}^n X_i^2\right)=\sum_{i=1}^n EX_i^2=\sum_{i=1}^n DX_i=\sum_{i=1}^n 1=n,$$

$$D\chi^2=D\left(\sum_{i=1}^n X_i^2\right)=\sum_{i=1}^n DX_i^2=\sum_{i=1}^n 2=2n,$$

其中，$DX_i^2=EX_i^4-(EX_i^2)^2=3-1=2$.

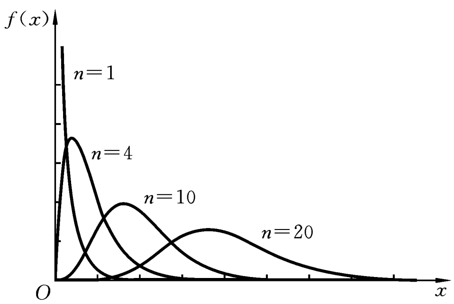

图 6.3

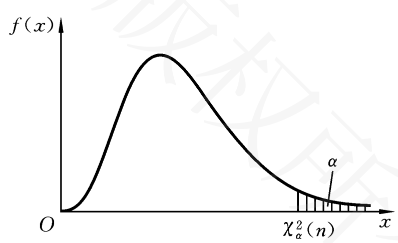

图 6.4

(2) χ^2 分布具有可加性. 设 $\chi_1^2\sim\chi^2(n_1)$，$\chi_2^2\sim\chi^2(n_2)$，且相互独立，则 $\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$.

对于给定 $\alpha(0<\alpha<1)$，称满足

$$P(\chi^2>\chi_\alpha^2(n))=\alpha$$

的点 $\chi_\alpha^2(n)$为 $\chi^2(n)$分布的**上侧 α 分位点**(见图 6.4)，其值可查附表 5.

可以证明，当 n 充分大时，$\sqrt{2\chi^2}$ 近似地服从正态分布 $N(\sqrt{2n-1},1)$. 一般地，当 $n>45$

时，$\chi_\alpha^2(n)$可由近似公式

$$\chi_\alpha^2(n)\approx\frac{1}{2}(u_\alpha+\sqrt{2n-1})^2$$

给出，其中 u_α 是在第二章中提到的标准正态分布的上侧 α 分位点.

6.2.2 t 分布

定义 6.5 设 $X\sim N(0,1)$，$Y\sim\chi^2(n)$，且相互独立，则称随机变量 $T=\dfrac{X}{\sqrt{Y/n}}$所服从的分布为自由度是 n 的 t(student)**分布**，记为 $T\sim t(n)$.

$t(n)$分布的概率密度为

$$f(x)=\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2},\quad -\infty<x<+\infty. \tag{6.6}$$

其概率密度函数曲线如图 6.5 所示.

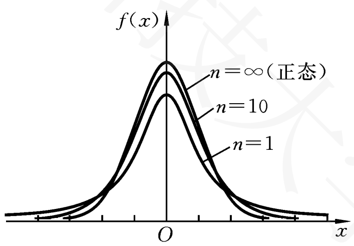

图 6.5

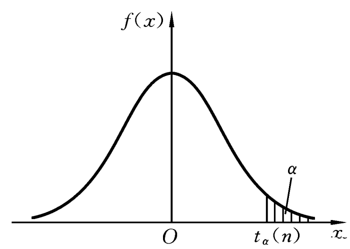

图 6.6

由于 $f(x)$为偶函数，其概率密度函数曲线关于纵轴对称，且

$$\lim_{|x|\to+\infty}f(x)=0,\quad \lim_{n\to+\infty}f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}},$$

即当自由度 $n\to+\infty$时，$t(n)$分布趋近于标准正态分布. 但当 n 较小时，$t(n)$分布与标准正态分布之间有较大的差异，且对于 $T\sim t(n)$，$X\sim N(0,1)$及较大的正数 t_0，有

$$P(T>t_0)\geqslant P(X>t_0),$$

即 $t(n)$的尾部比标准正态分布的尾部有较大的概率.

对于给定 $\alpha(0<\alpha<1)$，称满足

$$P(T>t_\alpha(n))=\alpha$$

的点 $t_\alpha(n)$为 $t(n)$分布的上侧 α 分位点(见图 6.6)，其值当 $n\leqslant45$ 时，可查附表 4，当 $n>45$ 时，$t_\alpha(n)\approx u_\alpha$.

如我们查表可得，$u_{0.025}=1.96$，$t_{0.025}(15)=2.1315$.

6.2.3 F 分布

定义 6.6 设 $X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，且相互独立，则称随机变量 $F=\dfrac{X/n_1}{Y/n_2}$所服从的分布为自由度是(n_1,n_2)的 F **分布**，记为 $F\sim F(n_1,n_2)$.

$F(n_1,n_2)$分布的概率密度为

$$f(x)=\begin{cases}\dfrac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)}\left(\dfrac{n_1}{n_2}\right)\left(\dfrac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x\geqslant 0,\\ 0, & x<0,\end{cases}\tag{6.7}$$

其概率密度函数曲线如图 6.7 所示.

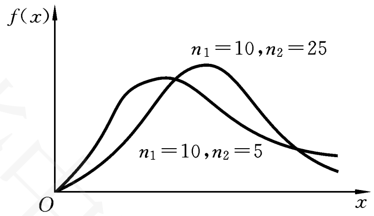

图 6.7

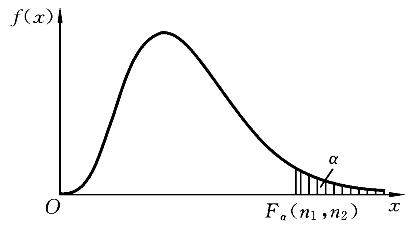

图 6.8

对于给定 $\alpha(0<\alpha<1)$,称满足

$$P(F>F_\alpha(n_1,n_2))=\alpha$$

的点 $F_\alpha(n_1,n_2)$为 $F(n_1,n_2)$分布的上侧 α 分位点(见图 6.8),其值可查附表 6.

由 F 分布的定义可得到它的一个重要性质:若 $F\sim F(n_1,n_2)$,则$\dfrac{1}{F}\sim F(n_2,n_1)$,由此我们有

$$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}.$$

如查附表 6,有 $F_{0.025}(10,15)=3.06$,所以 $F_{0.975}(15,10)=1/3.06=0.33$.

6.2.4　正态总体的样本均值与样本方差的分布

对于正态总体的样本均值与样本方差的分布,我们有以下的重要定理.

定理 6.2　设$(X_1,X_2,\cdots,X_n)$是取自正态总体 $N(\mu,\sigma^2)$的样本,$\overline{X}$ 和 S^2 分别为样本均值和样本方差,则

(1) $\overline{X}$ 与 S^2 独立;

(2) $\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$;

(3) $(n-1)\dfrac{S^2}{\sigma^2}\sim\chi^2(n-1)$.

定理中的(2)通过第三章和第四章的知识我们已经知道了,(1)和(3)的证明这里略去.

下面的几个推论给出了我们以后常常会用到的统计量的分布.

推论 6.1　$T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

证明　由定理 6.2 中(2)有$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$.再由定理 6.2 中(1)、(3)及定义6.5,有

$$T=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Bigg/\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}\sim t(n-1).$$

推论 6.2　设$(X_1,X_2,\cdots,X_{n_1})$和$(Y_1,Y_2,\cdots,Y_{n_2})$分别是取自两个相互独立的正态总体 $N(\mu_1,\sigma_1^2)$和 $N(\mu_2,\sigma_2^2)$的两个样本,其样本方差分别记为 S_1^2 和 S_2^2,则

(1) $F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$;

(2) $\dfrac{1}{n_1}\sum\limits_{i=1}^{n_1}\left(\dfrac{X_i-\mu_1}{\sigma_1}\right)^2\bigg/\dfrac{1}{n_2}\sum\limits_{i=1}^{n_2}\left(\dfrac{Y_i-\mu_2}{\sigma_2}\right)^2\sim F(n_1,n_2)$.

证明 (1) 由定理 6.2 中(3),有

$$(n_1-1)S_1^2/\sigma_1^2\sim\chi^2(n_1-1),\quad (n_2-1)S_2^2/\sigma_2^2\sim\chi^2(n_2-1),$$

又因为$(X_1,X_2,\cdots,X_{n_1})$和$(Y_1,Y_2,\cdots,Y_{n_2})$相互独立,所以$(n_1-1)S_1^2/\sigma_1^2$与$(n_2-1)S_2^2/\sigma_2^2$相互独立,再由定义 6.6,有

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\bigg/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\bigg/(n_2-1)}\sim F(n_1-1,n_2-1).$$

(2) 由 χ^2 分布的定义,有

$$\sum_{i=1}^{n_1}\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2\sim\chi^2(n_1),\quad \sum_{i=1}^{n_2}\left(\frac{Y_i-\mu_2}{\sigma_2}\right)^2\sim\chi^2(n_2),$$

由条件知上述两个随机变量相互独立,由定义 6.6 得到结论.

推论 6.3 设$(X_1,X_2,\cdots,X_{n_1})$和$(Y_1,Y_2,\cdots,Y_{n_2})$分别是取自两个相互独立的正态总体$N(\mu_1,\sigma^2)$和$N(\mu_2,\sigma^2)$的两个样本,其样本均值、样本方差分别记为$\overline{X},\overline{Y},S_1^2$和$S_2^2$,则

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w}\sqrt{\frac{n_1n_2}{n_1+n_2}}\sim t(n_1+n_2-2),$$

其中

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

证明 由于 $\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\dfrac{\sigma^2}{n_1}+\dfrac{\sigma^2}{n_2}\right)$,所以

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1),$$

又因为

$$(n_1-1)S_1^2/\sigma^2\sim\chi^2(n_1-1),\quad (n_2-1)S_2^2/\sigma^2\sim\chi^2(n_2-1),$$

且它们相互独立,由 χ^2 分布的可加性知,

$$V=(n_1-1)S_1^2/\sigma^2+(n_2-1)S_2^2/\sigma^2\sim\chi^2(n_1+n_2-2),$$

由定义 6.6,知

$$T=\frac{U}{\sqrt{V/(n_1+n_2-2)}}\sim t(n_1+n_2-2).$$

下面我们来看几个例子.

例 6.2 设$(X_1,X_2,\cdots,X_{10})$是取自正态总体$N(0,0.3^2)$的样本,求$P\left(\sum\limits_{i=1}^{10}X_i^2>1.44\right)$.

解 由于$\dfrac{X_i}{0.3}\sim N(0,1)$,$i=1,2,\cdots,10$,且相互独立,所以$\sum\limits_{i=1}^{10}\left(\dfrac{X_i}{0.3}\right)^2\sim\chi^2(10)$,因此,

$$P\left(\sum_{i=1}^{10}X_i^2>1.44\right)=P\left(\sum_{i=1}^{10}\left(\frac{X_i}{0.3}\right)^2>\frac{1.44}{0.3^2}\right)=P(\chi^2(10)>16)=0.1.$$

例 6.3　设$(X_1,X_2,\cdots,X_n,X_{n+1})$是取自正态总体$N(\mu,\sigma^2)$的样本,令

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad \widetilde{S}_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

求统计量$T=\dfrac{X_{n+1}-\overline{X}_n}{\widetilde{S}_n}\sqrt{\dfrac{n-1}{n+1}}$的分布.

解　因为$X_{n+1}-\overline{X}_n\sim N\left(0,\dfrac{n+1}{n}\sigma^2\right)$,所以$\dfrac{X_{n+1}-\overline{X}_n}{\sigma\sqrt{\dfrac{n+1}{n}}}\sim N(0,1)$,又因为$\dfrac{n\widetilde{S}_n^2}{\sigma^2}\sim\chi^2(n-1)$,且$\widetilde{S}_n^2$与$X_{n+1}-\overline{X}_n$独立,所以

$$T=\frac{X_{n+1}-\overline{X}_n}{\sqrt{\dfrac{n+1}{n}}\sigma}\Bigg/\sqrt{\frac{n\widetilde{S}_n^2}{\sigma^2}\Big/(n-1)}\sim t(n-1).$$

本题实际上是推论 6.3 的一个特殊情形$(n_2=1)$.

例 6.4　分别从方差为 20 和 35 的两个独立的正态总体中抽取容量为 8 和 10 的两个样本,估计第一个样本方差S_1^2不小于第二个样本方差S_2^2两倍的概率.

解　由推论 6.2 知,$\dfrac{S_1^2/20}{S_2^2/35}\sim F(7,9)$,所以

$$P(S_1^2\geqslant 2S_2^2)=P\left(\frac{S_1^2/20}{S_2^2/35}\geqslant 2\times\frac{35}{20}\right)=P(F(7,9)\geqslant 3.5),$$

查表有,$F_{0.05}(7,9)=3.29$,$F_{0.025}(7,9)=4.20$,所以

$$0.025\leqslant P(S_1^2\geqslant 2S_2^2)\leqslant 0.05.$$

6.2.5　顺序统计量的分布

在第三章中,我们给出了最大、最小值的分布,应用到数理统计中,实际上就是最大、最小顺序统计量的分布.下面我们来看更一般的顺序统计量$X_i^*(i=2,3,\cdots,n-1)$的分布.

设总体X的分布函数为$F(x)$,概率密度为$f(x)$,则顺序统计量$X_i^*(i=2,3,\cdots,n-1)$的分布函数为

$$\begin{aligned}F_{X_i^*}(x)&=P(X_i^*\leqslant x)=P(X_1,X_2,\cdots,X_n\text{ 中至少有 }i\text{ 个}\leqslant x)\\&=\sum_{k=i}^{n}P(X_1,X_2,\cdots,X_n\text{ 中恰好有 }k\text{ 个}\leqslant x)=\sum_{k=i}^{n}C_n^k(F(x))^k(1-F(x))^{n-k},\end{aligned}$$

求导得到其概率密度为

$$\begin{aligned}f_{X_i^*}(x)&=\sum_{k=i}^{n}C_n^k\left[k(F(x))^{k-1}(1-F(x))^{n-k}-(n-k)(F(x))^k(1-F(x))^{n-k-1}\right]f(x)\\&=\frac{n!}{(i-1)!\,(n-i)!}(F(x))^{i-1}(1-F(x))^{n-i}f(x).\end{aligned}$$

上式对任意$i=1,2,\cdots,n$都成立.

习　题　六

6.1　选择题.

(1) 设$(X_1,X_2,\cdots,X_n)(n\geqslant 2)$是取自正态总体$N(0,1)$的样本,$\overline{X}$和$S^2$分别为样本均值和样本方差,则(　　).

(A) $n\overline{X}\sim N(0,1)$ (B) $nS^2\sim\chi^2(n)$

(C) $\dfrac{(n-1)\overline{X}}{S}\sim t(n-1)$ (D) $\dfrac{(n-1)X_1^2}{\sum\limits_{i=2}^{n}X_i^2}\sim F(1,n-1)$

(2) 设总体 $X\sim N(\mu,\sigma^2)$,μ 已知,σ 未知,(X_1,X_2,X_3)是取自 X 的样本,则下列样本函数中非统计量的是( ).

(A) X_1,X_2 (B) $X_1+X_2+X_3-\mu$ (C) $(X_1-\mu)/\sigma$ (D) $\min\{X_1,X_2,X_3\}$

(3) 设 $X\sim t(n)$,则当 n 充分大时,$P(|X|<1)\approx$( ).

(A) 0 (B) 1 (C) 1/2 (D) 0.6826

6.2 填空题.

(1) 设 $X\sim t(n)$,则$\dfrac{1}{X^2}\sim$______.

(2) 设$(X_1,X_2,\cdots,X_n)$为来自总体 $\chi^2(10)$的样本,则统计量 $\sum\limits_{i=1}^{n}X_i\sim$ ______.

(3) 设$(X_1,X_2,\cdots,X_n)$是取自正态总体 $N(0,1)$的样本,$\overline{X}$ 和 S^2 分别为样本均值和样本方差,则当 $a=$______,$b=$______时,$a\overline{X}^2+bS^2$ 服从 χ^2 分布,且自由度为______,$DS^2=$______.

(4) 设$(X_1,X_2,\cdots,X_n)$是取自总体 $X\sim B(1,p)$的样本,$\overline{X}$ 为其样本均值,则对 $k=0,1,2,\cdots,n$,有 $P(n\overline{X}=k)=$______.

(5) 设$(X_1,X_2,\cdots,X_n)$是取自总体 $X\sim E(\lambda)$的样本,则联合概率密度 $f(x_1,x_2,\cdots,x_n)=$______.

6.3 设从总体 X 中取得一个容量为 5 的样本,观测值为 0,1,1,0,2.(1) 求此样本的经验分布函数,并作其图形;(2) 计算常用统计量的观测值.

6.4 设$(X_1,X_2,\cdots,X_n)$是取自总体 $X\sim P(\lambda)$的样本,求样本均值 $\overline{X}$ 的期望与方差.

6.5 设总体 $X\sim N(20,3)$,从 X 中分别抽取容量为 10,15 的两个相互独立的样本,求两样本均值之差的绝对值大于 0.3 的概率.

6.6 设$(X_1,X_2,\cdots,X_5)$是取自总体 $X\sim N(0,2^2)$的样本,求常数 a,b,使得 $a(X_1+X_2)^2+b(X_3+X_4-X_5)^2$服从 χ^2 分布.

6.7 设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的样本,S^2 为样本方差,证明 $ES^2=DX$.

6.8 设 $\overline{X}_n$ 和 S_n^2 分别为样本$(X_1,X_2,\cdots,X_n)$的样本均值和样本方差,现又独立地进行第$n+1$次观测 X_{n+1},记 $\overline{X}_{n+1}$和 S_{n+1}^2分别为样本$(X_1,X_2,\cdots,X_{n+1})$的样本均值和样本方差,证明:

(1) $\overline{X}_{n+1}=\overline{X}_n+\dfrac{1}{n+1}(X_{n+1}-\overline{X}_n)$;

(2) $S_{n+1}^2=\dfrac{n}{n+1}\left[S_n^2+\dfrac{1}{n+1}(X_{n+1}-\overline{X}_n)^2\right]$.

6.9 证明 $F_{1-\alpha}(n_1,n_2)=\dfrac{1}{F_\alpha(n_2,n_1)}$.

6.10 设 X,Y 独立同分布于 $N(0,3^2)$,$(X_1,X_2,\cdots,X_9)$与$(Y_1,Y_2,\cdots,Y_9)$为分别取自 X,Y 的样本,求 $U=\sum\limits_{i=1}^{9}X_i\Big/\sqrt{\sum\limits_{i=1}^{9}Y_i^2}$ 的分布.

6.11 设$(X_1,X_2,\cdots,X_9)$是取自正态总体 X 的样本,令 $Y_1=\dfrac{\sum\limits_{i=1}^{6}X_i}{6}$,$Y_2=\dfrac{\sum\limits_{i=7}^{9}X_i}{3}$,$S^2=\dfrac{1}{2}\sum\limits_{i=7}^{9}(X_i-Y_2)^2$,$Z=\dfrac{Y_1-Y_2}{S}$,求 $P(Z>1.33)$.

6.12 设$(X_1,X_2,\cdots,X_n)$是取自总体 $X\sim U(0,2)$的样本.

(1) 求顺序统计量 X_2^* 的概率密度;

(2) 求 $P(X_1^*\leqslant 1)$及 $P(X_n^*\geqslant 1)$.

第七章　参数估计

数理统计的核心内容是统计推断．所谓统计推断就是由样本信息来推断关于总体的种种结论．统计推断中的两个基本问题是：统计估计与统计假设检验．参数估计是统计估计的重要而基本的组成部分．本章讨论参数估计的常用方法及优良性评价问题．

7.1　参数估计概念

总体作为随机变量，其统计规律完全由分布函数刻画，因此，**统计推断**的一个重要任务是由样本推断总体的分布．在实际中我们常常可以根据专业的或经验的知识判断出所研究总体的分布形式，但分布中含有若干未知参数；或者，在已知总体分布形式的情形下会关心其未知参数的某函数为何值；又或者，虽然总体分布形式未知，但我们只关心其某些数字特征．例如，可认为某一行业从业者的年收入服从正态分布$N(\mu,\sigma^2)$，但参数μ,σ^2未知，需要估计μ,σ^2的值；另一方面，上一问题中，年收入超过定值a的概率$P(X>a)=1-\Phi\left(\frac{a-\mu}{\sigma}\right)$也是令人感兴趣的问题，为此，需要估计未知参数$\mu$和$\sigma$的函数$\Phi\left(\frac{a-\mu}{\sigma}\right)$；又例如，某一年龄段的人群中对某种疾病的发病率$X$的分布函数的类型未知，但我们只关心该疾病在该人群的平均发病率和发病率的波动情况，即需要估计其数学期值$E(X)$和方差$D(X)$．通常将总体参数的函数和总体数字特征也统称为总体参数．参数估计问题就是利用样本对上述各类总体未知参数作出估计．

参数估计的形式有两种：点估计和区间估计．

所谓点估计是用一个数值来估计某个未知参数．一般，设总体X的参数θ（θ可为向量）未知，记θ的所有可能取值所成的集合为Θ，称Θ为**参数空间**．用总体样本$(X_1,X_2,\cdots,X_n)$构造统计量$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$来估计$\theta$，称$\hat{\theta}$为$\theta$的**估计量**，对于样本的一组观察值$(x_1,x_2,\cdots,x_n)$，相应的估计量的观察值$\hat{\theta}(x_1,x_2,\cdots,x_n)$称为$\theta$的**估计值**，也简记为$\hat{\theta}$．若$\theta$为$m$维向量，则$\theta$为$m$维欧氏空间的点，用$\hat{\theta}$估计$\theta$亦即用一个$m$维点$\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m)$去估计$m$维点$\theta=(\theta_1,\theta_2,\cdots,\theta_m)$，因此称这样的估计为**点估计**．参数的点估计方法很多，本章介绍两种常用的方法：矩估计法和极大似然估计法．

对总体的待估参数，有时需要获得其值所在的一个范围（区间），以一给定的概率相信该范围包含了该参数．具体到一维参数θ，即要由总体样本$(X_1,X_2,\cdots,X_n)$构造一个随机区间$(\hat{\theta}_1,\hat{\theta}_2)$，其中$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$，使得区间$(\hat{\theta}_1,\hat{\theta}_2)$以给定的概率覆盖了参数$\theta$，称$(\hat{\theta}_1,\hat{\theta}_2)$为$\theta$的区间估计．当待估参数为多维向量时，则这个范围就是相应的多维空间的一个区域．

7.2　矩估计法和极大似然估计法

本节介绍点估计的两种常用方法:矩估计法和极大似然估计法,在实际应用中还有最小二乘法(见第九章)、判决函数法、自适应法、稳健估计法及贝叶斯法等.

7.2.1　矩估计法

矩估计法由英国统计学家皮尔逊(K. Pearson)于 1894 年提出,是一种简单而直观的传统估计方法.由辛钦大数定律知,样本 k 阶原点矩依概率收敛于对应的总体 k 阶原点矩.这就保证了样本 k 阶原点矩作为随机变量的取值将随样本容量的增大而越来越逼近对应总体矩的真值.以样本原点矩作为对应的总体原点矩的估计量,以样本原点矩的连续函数作为对应的总体原点矩的连续函数的估计量,进而得到未知参数的点估计量,这样的方法称为**矩估计法**.矩估计法的具体做法表述如下.

设总体 X 是概率密度为 $f(x;\theta_1,\cdots,\theta_k)$的连续型随机变量,或总体 X 是分布列为 $P(X=x)\triangleq f(x;\theta_1,\cdots,\theta_k)$的离散型随机变量,其中 $\theta_1,\cdots,\theta_k$ 为待估参数.假定总体 X 的 k 阶原点矩存在,则其 j 阶原点矩

$$\alpha_j=E(X^j)=\int_{-\infty}^{+\infty}x^j f(x;\theta_1,\cdots,\theta_k)\mathrm{d}x\quad(j=1,2,\cdots,k),$$

或

$$\alpha_j=E(X^j)=\sum_x x^j f(x;\theta_1,\cdots,\theta_k)\quad(j=1,2,\cdots,k)$$

是 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数:

$$\alpha_j=\alpha_j(\theta_1,\theta_2,\cdots,\theta_k)\quad(j=1,2,\cdots,k).$$

设$(X_1,X_2,\cdots,X_n)$为取自总体 X 的样本,相应的样本 j 阶原点矩为

$$A_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j\quad(j=1,2,\cdots,k).$$

令 X 的 j 阶总体矩等于对应的样本矩,得 k 个方程

$$\alpha_j(\theta_1,\theta_2,\cdots,\theta_k)=A_j\quad(j=1,2,\cdots,k).\tag{7.1}$$

求解上述方程组得到的一组解

$$\hat{\theta}_j=\hat{\theta}_j(X_1,X_2,\cdots,X_n)\quad(j=1,2,\cdots,k)$$

作为未知参数 $\theta_1,\theta_2,\cdots,\theta_k$ 的估计量,称为**矩估计量**(moment estimator).

例 7.1　设总体 X 服从参数为 λ 的指数分布,$(X_1,X_2,\cdots,X_n)$为取自 X 的样本,求未知参数 λ 的矩估计.

解　指数分布的概率密度为

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

由于只有一个待估参数,故据(7.1)式,只需考虑一阶矩方程.因

$$\alpha_1(\lambda)=E(X)=\int_{-\infty}^{+\infty}xf(x;\lambda)\mathrm{d}x=\frac{1}{\lambda},$$

令

$$\frac{1}{\lambda}=A_1=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X},$$

故解得 λ 的矩估计为 $\hat{\lambda}=\dfrac{1}{\overline{X}}$.

例 7.2　设总体 X 的均值 μ 和方差 σ^2 均未知.试求 μ,σ^2 的矩估计.

解　设$(X_1,X_2,\cdots,X_n)$是来自总体 X 的样本,由于有两个未知参数,故据(7.1)式,令

$$\begin{cases}\alpha_1(\mu,\sigma^2)=E(X)=A_1,\\ \alpha_2(\mu,\sigma^2)=E(X^2)=A_2,\end{cases}$$

即

$$\begin{cases}\mu=\overline{X},\\ \sigma^2+\mu^2=A_2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^2,\end{cases}$$

解得

$$\begin{cases}\hat{\mu}=\overline{X},\\ \hat{\sigma}^2=A_2-A_1^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^2-\overline{X}^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2=\widetilde{S}^2.\end{cases}$$

由上例知,对任何总体而言,只要其期望、方差存在,则总体期望的矩估计为样本均值 $\overline{X}$,总体方差的矩估计为样本二阶中心矩 $\widetilde{S}^2$.另外,将上例的结果用于例 7.1,由于指数分布总体的方差 $D(X)=1/\lambda^2$,于是 $\lambda=1/\sqrt{D(X)}$,故由矩估计法,λ 的矩估计也可为

$$\hat{\lambda}=\frac{1}{\widetilde{S}}.$$

这说明一个参数的矩估计量不一定唯一.这是矩估计法的一个缺点.通常应该尽量采用低阶矩来给出未知参数的矩估计.

事实上,例 7.2 中总体二阶中心矩(方差)的矩估计为样本二阶中心矩的结论可推广到一般,即总体的 k 阶中心矩(若存在)的矩估计是对应的样本 k 阶中心矩.这是由于总体 k 阶中心矩 $\mu_k=E(X-EX)^k$ 总可以展开为阶数不超过 k 的总体原点矩的函数,而样本 k 阶中心矩 $B_k=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^k$ 也可化为阶数不超过 k 的样本原点矩的相同函数.因此,这一结果使得我们可将矩估计法推广为:用样本矩作出对应的总体矩的估计.

例 7.3　设总体 X 在区间$[a,b]$上服从均匀分布,a,b 为待估参数,$(X_1,X_2,\cdots,X_n)$为来自 X 的样本.求 a,b 的矩估计.

解　均匀分布的概率密度为

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

因

$$\alpha_1(a,b)=E(X)=\int_a^b\frac{x}{b-a}\mathrm{d}x=\frac{1}{2}(a+b),$$

$$\alpha_2(a,b)=E(X^2)=\int_a^b\frac{x^2}{b-a}\mathrm{d}x=\frac{1}{3}(a^2+ab+b^2).$$

令

$$\begin{cases}\dfrac{1}{2}(a+b)=A_1=\overline{X},\\ \dfrac{1}{3}(a^2+ab+b^2)=A_2.\end{cases}$$

解得

$$\begin{cases}\hat{a}=A_1-\sqrt{3(A_2-A_1^2)},\\ \hat{b}=A_1+\sqrt{3(A_2-A_1^2)}.\end{cases}$$

由于例 7.2 已推出 $A_2-A_1^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_1^2-\overline{X}^2=\widetilde{S}^2$,故 a,b 的矩估计为

$$\begin{cases}\hat{a}=\overline{X}-\sqrt{3\widetilde{S}^2},\\ \hat{b}=\overline{X}+\sqrt{3\widetilde{S}^2}.\end{cases}$$

例 7.4 求事件 A 发生的概率的矩估计.

解 记事件 A 发生的概率 $P(A)=p$. 考虑以 p 为参数的两点分布总体

$$X=\begin{cases}1, \text{在一次试验中 } A \text{ 发生},\\ 0, \text{在一次试验中 } A \text{ 不发生}.\end{cases}$$

设$(X_1,X_2,\cdots,X_n)$是来自 X 的样本,则

$$\alpha_1(p)=E(X)=p.$$

令

$$p=\overline{X},$$

即得事件 A 发生的概率的矩估计

$$\hat{p}=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

注意 $\sum\limits_{i=1}^{n}X_i$ 是事件 A 在 n 次试验中发生的次数,从而 $\overline{X}$ 实际上是事件 A 发生的频率,即有结论:事件发生的概率的矩估计是事件发生的频率.

7.2.2 极大似然估计法

由高斯和费希尔(R. A. Figher)先后提出的极大似然估计法是被使用最广泛的一种参数估计方法. 该方法建立的依据是直观的极大似然原理:一个试验有若干个可能结果 $A_1,A_2,\cdots,A_n$,若一次试验的结果是 A_i 发生,则自然认为 A_i 在所有可能结果中发生的概率最大. 当总体 X 的未知参数 θ 待估时,应用这一原理,对 X 的样本$(X_1,X_2,\cdots,X_n)$作一次观测试验,得样本观察值$(x_1,x_2,\cdots,x_n)$为此次试验结果,那么参数 θ 的估计值应取为使这一结果发生的概率为最大才合理. 这就是极大似然估计法的基本思想. 下面用一个简单的例子加以说明.

例 7.5 设一袋中装有多个白球和黑球,已知两种球的数目之比为 1∶2,但不知黑球多还是白球多. 现从中有放回地依次取出 3 球,结果是(白,黑,白),试由此估计白球所占比例 p 究竟为$\frac{1}{3}$还是$\frac{2}{3}$.

解 白球所占比例 p 亦即从袋中任取一球得白球的概率. 因而 p 可视为两点分布总体 X 的参数,X 的分布律为

$$X=\begin{cases}1, & \text{取到一白球},\\ 0, & \text{取到一黑球}.\end{cases}$$

于是,有放回地取出 3 只球是对总体 X 的样本(X_1,X_2,X_3)的一次观察试验. 试验的结果(白,黑,白)即样本观察值为(1,0,1). 要由这一结果判定是 $p=\frac{1}{3}$还是 $p=\frac{2}{3}$. 依极大似然原理,此次试验结果发生的概率应该最大,故分别在 $p=\frac{1}{3}$和 $p=\frac{2}{3}$下计算此概率,得

$$\begin{aligned}P((X_1,X_2,X_3)=(1,0,1))&=P(X_1=1)P(X_2=0)P(X_3=1)\\&=\begin{cases}\left(\frac{1}{3}\right)^2\frac{2}{3}=\frac{2}{27}, \text{相应于 } p=\frac{1}{3},\\ \left(\frac{2}{3}\right)^2\frac{1}{3}=\frac{4}{27}, \text{相应于 } p=\frac{2}{3}.\end{cases}\end{aligned}$$

比较知，对应于 $p=\frac{2}{3}$，此概率有较大的值，因此，选取 $\frac{2}{3}$ 为 p 的估计值更合理. 类似地，可计算此次试验的任一可能结果，亦即样本 (X_1,X_2,X_3) 的任一组可能取值情形分别在 $p=\frac{1}{3}$ 和 $\frac{2}{3}$ 下发生的概率，如表 7.1 所示.

表 7.1　例 7.1 的概率值 $p=(X_1=x_1,X_2=x_2,X_3=x_3)$

(x_1,x_2,x_3)	$(0,0,0)$	$(1,0,0)$ $(0,1,0)$ $(0,0,1)$	$(1,1,0)$ $(0,1,1)$ $(1,0,1)$	$(1,1,1)$
$p=\frac{1}{3}$	$\frac{8}{27}$	$\frac{4}{27}$	$\frac{2}{27}$	$\frac{1}{27}$
$p=\frac{2}{3}$	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{8}{27}$

表 7.1 的第 2～5 列给出当 $p=\frac{1}{3}$ 或 $p=\frac{2}{3}$ 时试验各种可能结果的概率值，试验的任一结果位于某列的第 1 行，该列的第 2，3 行是这一结果分别在 $p=\frac{1}{3}$ 或 $\frac{2}{3}$ 下的概率值.

一般，设总体 X 的未知参数 θ 待估，$(x_1,x_2,\cdots,x_n)$ 为 X 的一组样本观察值，则：

（1）当 X 为离数型时，该组观察值发生的概率

$$L(\theta)=P(X_1=x_1,\cdots,X_n=x_n;\theta)=\prod_{i=1}^{n}P(X_i=x_i;\theta)$$

应最大，从而求 θ 的极大似然估计值即求使上式的 $L(\theta)$ 达最大的估计值 $\hat{\theta}(x_1,x_2,\cdots,x_n)$.

（2）当 X 为连续型时，观察值 $(x_1,x_2,\cdots,x_n)$ 发生的概率为 0，但由于样本 $(X_1,X_2,\cdots,X_n)$ 的概率密度在一点处的函数值大小反映和决定样本在该点附近取值的概率大小，因此，观察值 $(x_1,x_2,\cdots,x_n)$ 发生的概率最大可用样本的概率密度在该观察值点的函数值

$$f(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta) \tag{7.2}$$

最大来反映和表示（其中 $f(x;\theta)$ 为总体概率密度）. 求 θ 的极大似然估计就是求使以上函数值达最大的估计值 $\hat{\theta}(x_1,x_2,\cdots,x_n)$.

综合上述两种情形，我们给出如下定义.

定义 7.1　设总体 X 的概率密度为 $f(x;\theta)$（当 X 为离散型时，$f(x;\theta)=P(X=x;\theta)$），$\theta\in\Theta$，其中 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 为待估的未知参数，Θ 为参数 θ 的可能取值所成的参数空间，$(x_1,x_2,\cdots,x_n)$ 是总体样本 $(X_1,X_2,\cdots,X_n)$ 的一组观察值. 称

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta) \tag{7.3}$$

为样本的**似然函数**；若存在 $\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)$，使得

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta)$$

成立，则称

$$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)=(\hat{\theta}_1(x_1,x_2,\cdots,x_n),\cdots,\hat{\theta}_k(x_1,x_2,\cdots,x_n))$$

为参数 θ 的**极大似然估计值**，而称

$$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)=(\hat{\theta}_1(X_1,X_2,\cdots,X_n),\cdots,\hat{\theta}_k(X_1,X_2,\cdots,X_n))$$

为 θ 的**极大似然估计量**(maximum likelihood estimator).

由于 $f(x;\theta)$往往是关于 θ 可微的,因此一般$\hat{\theta}$可由方程组

$$\frac{\partial L(\theta)}{\partial \theta_i}=0 \quad (i=1,2,\cdots,k) \tag{7.4}$$

解得. 又由于 $\ln L(\theta)$与 $L(\theta)$同时取得最大值,故等价地可由方程组

$$\frac{\partial \ln L(\theta)}{\partial \theta_i}=0 \quad (i=1,2,\cdots,k) \tag{7.5}$$

求得$\hat{\theta}$. 但要注意,当 $f(x;\theta)$关于 θ 不可微或上述方程组无解时,就必须根据极大似然估计的定义和 θ 的取值范围 Θ 求$\hat{\theta}$.

(7.4)式和(7.5)式中的方程都称为**似然方程**.

例 7.6　某种电子管的使用寿命 X 服从指数分布,其概率密度为

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

试求参数 θ 的极大似然估计。

解　设$(x_1,x_2,\cdots,x_n)$为 X 的一组样本观察值,由(7.2)式,似然函数为

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\prod_{i=1}^{n}\frac{1}{\theta}\mathrm{e}^{-\frac{x_i}{\theta}}=\frac{1}{\theta^n}\mathbf{exp}\left\{-\frac{1}{\theta}\sum_{i=1}^{n}x_i\right\},$$

将上式取对数得

$$\ln(L(\theta))=-n\ln\theta-\frac{1}{\theta}\sum_{i=1}^{n}x_i,$$

关于 θ 求导,得似然方程

$$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=\frac{-n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}x_i=0,$$

解方程得 θ 的极大似然估计值为

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

例 7.7　从一批产品中任取 50 件,发现有 2 件废品,试求这批产品的废品率 p 的极大似然估计.

解　废品率 p 即从此批产品中任取一件为次品的概率. 故 p 为两点分布总体 $X\sim B(1,p)$,亦即

$$X=\begin{cases}1, & \text{任取一件为废品,}\\ 0, & \text{任取一件为正品}\end{cases}$$

的参数. 设$(X_1,X_2,\cdots,X_n)$为来自 X 的样本,则

$$P(X_i=x_i)=p^{x_i}(1-p)^{1-x_i}, \quad x_i=0,1, \quad i=1,2,\cdots,n.$$

于是,似然函数为

$$L(p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{n-\sum\limits_{i=1}^{n}x_i}.$$

对上式取对数,并记 $\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$, 得

$$\ln L(p)=n\bar{x}\ln p+n(1-\bar{x})\ln(1-p).$$

关于 p 求导,得似然方程

$$\frac{n\bar{x}}{p}-\frac{n(1-\bar{x})}{1-p}=0,$$

解此方程得 p 的极大似然估计为

$$\hat{p}=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

现 $n=50$, $\sum\limits_{i=1}^{n}x_i=2$,故此时废品率的极大似然估计值为

$$\hat{p}=\bar{x}=\frac{2}{50}=4\%.$$

例 7.8　设总体 $X\sim N(\mu,\sigma^2)$,其中 μ,σ^2 为未知参数,$(X_1,X_2,\cdots,X_n)$为总体 X 的样本,求 μ 和 σ^2 的极大似然估计.

解　设$(x_1,x_2,\cdots,x_n)$为样本$(X_1,X_2,\cdots,X_n)$的观测值,则似然函数是 μ,σ^2 的函数:

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}f(x_i;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\},$$

$$\ln L(\mu,\sigma^2)=\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2,$$

令
$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0,\\[2ex]\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2=0,\end{cases}$$

解得 μ,σ^2 的极大似然估计为

$$\begin{cases}\hat{\mu}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i=\overline{X},\\[2ex]\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\widetilde{S}^2.\end{cases}$$

例 7.9　设总体 $X\sim U(0,\theta)$,$\theta>0$ 未知,$(x_1,x_2,\cdots,x_n)$为 X 的样本观察值,求 θ 的极大似然估计.

解　X 的概率密度为

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

似然函数为

$$L(x_1,x_2,\cdots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n},\ 0\leqslant x_1,x_2,\cdots,x_n\leqslant\theta,\theta>0,\\ 0,\ \text{其他}.\end{cases}$$

记 $x_{(n)}^{*}=\max\{x_1,x_2,\cdots,x_n\}$,由

$$\frac{\mathrm{d}L}{\mathrm{d}\theta}=\begin{cases}-n\theta^{-(n+1)}<0,\ \theta\geqslant x_{(n)}^{*},\\ 0,\qquad 0<\theta<x_{(n)}^{*}\end{cases}$$

知似然方程$\dfrac{\mathrm{d}L}{\mathrm{d}\theta}=0$ 无解.由于当 $0<\theta<x_{(n)}^{*}$ 时 $L(\theta)=0$,当 $\theta\geqslant x_{(n)}^{*}$ 时 $L(\theta)$关于 θ 严格降,故

$L(\theta)$在 $x_{(n)}^*$ 取得最大值,即 θ 的极大似然估计为

$$\hat{\theta}=x_{(n)}^*=\max\{x_1,x_2,\cdots,x_n\}.$$

极大似然估计法有许多优良性质,其中一个简单有用的性质是:若 $u=u(\theta)(\theta\in\Theta)$有唯一反函数,且$\hat{\theta}$是 θ 的极大似然估计量,则 $u(\hat{\theta})$是 $u(\theta)$的极大似然估计量.

事实上,因为$\hat{\theta}$是 θ 的极大似然估计,所以

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta),$$

其中$(x_1,x_2,\cdots,x_n)$是总体 X 的样本观察值. 令 $\theta=\theta(u)$($u\in\mathscr{U}$, $\mathscr{U}$ 为 $u(\theta)$的值域)是 $u=u(\theta)$的唯一反函数,记 $\hat{u}=u(\hat{\theta})$,则$\hat{\theta}=\theta(\hat{u})$,因此

$$L(x_1,x_2,\cdots,x_n;\theta(\hat{u}))=\max_{u\in\mathscr{U}}L(x_1,x_2,\cdots,x_n;\theta(u)).$$

由此可见,$\hat{u}=u(\hat{\theta})$是 $u=u(\theta)$的极大似然估计.

例 7.10　设总体 $X\sim N(\mu,\sigma^2)$,其中 μ,σ^2 未知,试求标准差 σ 的极大似然估计.

解　设$(X_1,X_2,\cdots,X_n)$为来自 X 的样本. 由例 7.8 知,σ^2 的极大似然估计为

$$\hat{\sigma}^2=\widetilde{S}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

由于 $\sigma=\sqrt{\sigma^2}$关于 σ^2 的严格单调,故 σ 的极大似然估计为

$$\hat{\sigma}=\sqrt{\hat{\sigma}^2}=\sqrt{\widetilde{S}^2}=\widetilde{S}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}.$$

7.3　估计量的评选标准

从上节的讨论可以看到,参数的点估计方法有多种,对于同一参数,用不同的估计法可能得到不同的估计量,这时就存在着采用哪一个估计量为好的问题. 另一方面,即使用不同的估计法得到了某参数的相同估计量,也存在一个评判该估计量好不好的问题,这就需要先明确什么是"好","好"的标准是怎样的. 本节介绍三种常用的评价估计量优良性的标准.

7.3.1　无偏性

估计量是随机变量,相应于不同的样本观测值而取到不同的估计值. 由于随机性,其取值难免与参数的真值有或大或小的偏差,但我们希望估计量的所有可能取值按概率加权的平均值与参数真值没有偏差,亦即要求估计量的数学期望等于参数的真值,这就是无偏性标准.

定义 7.2　设$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是参数 θ 的估计量,若对于任意的 $\theta\in\Theta$,有

$$E(\hat{\theta})=\theta,$$

则称$\hat{\theta}$是 θ 的**无偏估计量**(或称估计量$\hat{\theta}$是**无偏的**);记

$$b_n=E(\hat{\theta})-\theta,$$

称 b_n 为估计量$\hat{\theta}$的**偏差**;当 $b_n\neq0$ 时,称$\hat{\theta}$是 θ 的**有偏估计**;若

$$\lim_{n\to\infty}b_n=0,$$

则称$\hat{\theta}$是 θ 的**渐近无偏估计**.

例 7.11　设总体 X 的 k 阶原点矩 $\alpha_k=E(X^k)(k\geqslant1)$存在,$(X_1,X_2,\cdots,X_n)$是 X 的样本,试证明样本 k 阶原点矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$ 是 α_k 的无偏估计.

证明　由样本的定义知 $X_1,X_2,\cdots,X_n$ 与 X 同分布，因此

$$E(X_i^k)=E(X^k)=\alpha_k,\quad k\geqslant 1,\ i=1,2,\cdots,n.$$

故
$$E(A_k)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=\alpha_k.$$

上例给出了一个一般性结论：不论总体 X 服从什么分布，其样本 k 阶原点矩必是对应的总体 k 阶原点矩 α_k（若存在）的无偏估计. 特别地，样本均值 $\overline{X}$ 是总体均值 $E(X)$ 的无偏估计.

例 7.12　设总体方差 $DX=\sigma^2<+\infty$，试证明样本方差 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 是 σ^2 的无偏估计量.

证明　设总体均值 $E(X)=\mu$，由 $D(X)=\sigma^2<+\infty$ 知 μ 存在且有限.

$$\begin{aligned}ES^2&=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2\right\}\\&=\frac{1}{n-1}E\{\sum_{i=1}^{n}[(X_i-\mu)^2-2(X_i-\mu)(\overline{X}-\mu)+(\overline{X}-\mu)^2]\}\\&=\frac{1}{n-1}[\sum_{i=1}^{n}E(X_i-\mu)^2-2E\sum_{i=1}^{n}(X_i-\mu)(\overline{X}-\mu)+nE(\overline{X}-\mu)^2]\\&=\frac{1}{n-1}\sum_{i=1}^{n}E(X_i-\mu)^2-\frac{n}{n-1}E(\overline{X}-\mu)^2=\frac{n}{n-1}\sigma^2-\frac{n}{n-1}\cdot\frac{\sigma^2}{n}=\sigma^2,\end{aligned}$$

即样本方差是总体方差的无偏估计量.

值得注意的是，总体方差 σ^2 的矩估计 $\widetilde{S}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$ 是有偏估计，这是因为

$$E(\widetilde{S}^2)=E\left(\frac{n-1}{n}S^2\right)=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}\sigma^2.$$

由此可见，$\widetilde{S}^2$ 为 σ^2 的渐近无偏估计. 于是，当 n 比较大时，取 S^2 或 $\widetilde{S}^2$ 作为 σ^2 的估计皆可.

显然，对于同一未知参数，可以构造许多无偏估计. 例如，若 $(X_1,X_2,\cdots,X_n)$ 是总体 X 的一个样本，则对任意满足 $\sum\limits_{i=1}^{n}c_i=1$ 的常数 $c_i(1\leqslant i\leqslant n)$ 而言，估计量

$$\sum_{i=1}^{n}c_iX_i$$

总是总体期望 $E(X)$ 的无偏估计.

例 7.13　设总体 $X\sim E(\lambda)$，其中 $\lambda=\dfrac{1}{\theta}>0$，$(X_1,X_2,\cdots,X_n)$ 为 X 的样本，证明 $nX_{(1)}^*$ 是 θ 的无偏估计，其中 $X_{(1)}^*=\min\{X_1,X_2,\cdots,X_n\}$.

证明　依题意，X 的概率密度为

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

由 3.5 节的(3.5)式知，$X_{(1)}^*$ 的分布函数为

$$F_{X_{(1)}^*}(x;\theta)=[1-F(x;\theta)]^n,$$

其中 $F(x;\theta)$ 是 X 的分布函数，故 $X_{(1)}^*$ 的概率密度为

$$f_{X^*_{(1)}}(x;\theta)=n[1-F(x;\theta)]^{n-1}f(x;\theta)=\begin{cases}\dfrac{n}{\theta}\mathrm{e}^{-\frac{n}{\theta}x}, & x>0,\\ 0, & x\leqslant 0,\end{cases}$$

即 $X^*_{(1)}\sim E(n/\theta)$,故

$$E(X^*_{(1)})=\int_0^{+\infty}x\cdot\frac{n}{\theta}\mathrm{e}^{-\frac{n}{\theta}x}\mathrm{d}x=\theta/n,$$

从而

$$E(nX^*_{(1)})=nE(X^*_{(1)})=\theta,$$

即 $nX^*_{(1)}$ 是 θ 的无偏估计.

由指数分布 $E(\lambda)$ 的数学期望 $EX=\dfrac{1}{\lambda}=\theta$ 知,$\overline{X}$ 是上例中 θ 的另一无偏估计,这表明,一个未知参数可以有多个不同的无偏估计量.

还值得注意的是,若 $\hat{\theta}$ 是 θ 的无偏估计,对于 θ 的函数 $g(\theta)$,$g(\hat{\theta})$ 不一定是 $g(\theta)$ 的无偏估计.

例 7.14 试证明样本标准差 S 不是总体标准差 σ 的无偏估计.

证明 由 $\sigma^2=E(S^2)=D(S)+[E(S)]^2$ 及 $D(S)\geqslant 0$,知

$$E(S)=\sqrt{\sigma^2-D(S)}\leqslant\sigma.$$

这个例子表明,虽然样本方差是总体方差的无偏估计,但它的平方根亦即样本标准差却不是总体标准差的无偏估计.

7.3.2 有效性

既然同一未知参数可以有多个无偏估计,那么就存在对这多个无偏估计作优劣比较的问题.虽然它们取值都在参数真值 θ 附近,但集中程度会有差异.我们当然希望集中的程度越高越好.由于无偏估计量的数学期望就是参数真值,这意味着,要求无偏估计量偏离其数字期望的程度越小越好,亦即方差越小越好,这就产生了无偏估计量的有效性标准.

定义 7.3 设 $\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$ 与 $\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$ 都是待估参数 θ 的无偏估计量,若有

$$D(\hat{\theta}_1)<D(\hat{\theta}_2),$$

则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**.

例如,设 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的样本,则 X_1 与 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 都是总体均值 $E(X)$ 的无偏估计量,但当 $n\geqslant 2$ 时,

$$D(X_1)=D(X)>D(\overline{X})=\frac{1}{n}D(X),$$

所以 $\overline{X}$ 比 X_1 有效.

定义 7.4 设 $\hat{\theta}_0(X_1,X_2,\cdots,X_n)$ 是参数 θ 的无偏估计量,若对 θ 的任一无偏估计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$,都有

$$D(\hat{\theta}_0)\leqslant D(\hat{\theta}),$$

则称 $\hat{\theta}_0$ 是 θ 的**最小方差无偏估计量**(又称**最优无偏估计量**).

如何寻找最小方差无偏估计量,读者可参考有关书籍.

例 7.15 设总体 X 服从区间 $[0,\theta]$ 上的均匀分布,$(X_1,X_2,\cdots,X_n)$ 为取自 X 的样本,对

未知参数 θ 的两个估计量：

$$\hat{\theta}_1=2\overline{X},\quad \hat{\theta}_2=\frac{n+1}{n}\max_{1\leqslant i\leqslant n}\{X_i\},$$

（1）试验证 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 均为 θ 的无偏估计；

（2）指出哪一个更有效.

解　（1）因

$$E(\hat{\theta}_1)=2E(\overline{X})=2\times\frac{\theta}{2}=\theta,$$

由

$$E(X_{(n)}^*)=\int_0^\theta x\,\frac{n\,x^{n-1}}{\theta^n}\mathrm{d}x=\frac{n}{n+1}\theta,$$

有

$$E(\hat{\theta}_2)=\frac{n+1}{n}E(X_{(n)}^*)=\theta.$$

故 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 均为 θ 的无偏估计.

（2）因

$$D(\hat{\theta}_1)=4D(\overline{X})=\frac{4}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{4}{n^2}\times n\,\frac{\theta^2}{12}=\frac{\theta^2}{3n},$$

而由

$$E[(X_{(n)}^*)^2]=\int_0^\theta x^2\,\frac{nx^{n-1}}{\theta^n}\mathrm{d}x=\frac{n}{n+2}\theta^2,$$

及 $D(X_{(n)}^*)=E[(X_{(n)}^*)^2]-E^2[X_{(n)}^*]=\dfrac{n}{n+2}\theta^2-\left(\dfrac{n}{n+1}\theta\right)^2=\dfrac{n}{(n+2)(n+1)^2}\theta^2$，

得

$$D(\hat{\theta}_2)=\left(\frac{n+1}{n}\right)^2D(X_{(n)}^*)=\frac{\theta^2}{(n+2)n},$$

显然，当 $n>1$ 时 $\dfrac{1}{n(n+2)}<\dfrac{1}{3n}$，故 $\hat{\theta}_2$ 比 $\hat{\theta}_1$ 有效.

7.3.3　一致性

上面的无偏性与有效性标准是在样本容量 n 固定的前提下提出的.当样本容量 n 增大时，样本中包含的总体信息将增多，估计量作为样本的函数，其包含的总体信息也会增多.一个基本而自然的要求是，随着 n 的增大，估计量越来越逼近所估参数的真值.这就是一致性标准.

定义 7.5　设 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为参数 θ 的估计量，若对任意的 $\theta\in\Theta$ 及任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P(|\hat{\theta}(X_1,X_2,\cdots,X_n)-\theta|\leqslant\varepsilon)=1,\tag{7.6}$$

则称 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 θ 的**一致估计量**.

例 7.16　设总体 X 的 k 阶原点矩 $E(X^k)=\alpha_k$ 存在，$(X_1,X_2,\cdots,X_n)$ 是 X 的样本，试证明样本 k 阶原点矩 $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$ 是 α_k 的一致估计量.

证明　因为 $X_1,X_2,\cdots,X_n$ 相互独立且与 X 同分布，故 $X_1^k,X_2^k,\cdots,X_n^k$ 相互独立且与 X^k 同分布，于是有

$$E(X_1^k)=E(X_2^k)=\cdots=E(X_n^k)=\alpha_k.$$

由辛钦大数定律知，对任意 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i^k-\alpha_k\right|\leqslant\varepsilon\right)=1,$$

即样本 k 阶原点矩是参数 α_k 的一致估计量.

可以证明，总体原点矩的连续函数 $u=g(\alpha_1,\alpha_2,\cdots,\alpha_k)$ 的矩估计 $\hat{u}=g(A_1,A_2,\cdots,A_k)$

$\left(\text{其中 } A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k\right)$ 是一致估计量;在一定的条件下,待估参数的极大似然估计量也具有一致性.

7.4 区间估计

7.4.1 区间估计的概念

点估计法直接给出未知参数的具体估计值,简单明确而便于应用.但对于估计值与参数真值偏差多少、可信程度又如何这样的问题,点估计法却无法回答.解决这一问题的一个直观想法是,给定一个可信程度(概率)要求,然后基于样本来寻求一个未知参数值的范围(区间),使能以给定的可信概率相信该区间包含了未知参数的真值,从而也能以所给概率相信该区间中任一点作为参数估计值时与参数真值的偏差不超过该区间的长度,这就是区间估计的基本思想.一般定义如下.

定义 7.6 设总体 X 的分布函数为 $F(x;\theta)$,其中 θ 是未知参数,$(X_1,X_2,\cdots,X_n)$为 X 的样本,给定 $\alpha(0<\alpha<1)$,若统计量$\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$和$\bar{\theta}=\bar{\theta}(X_1,X_2,\cdots,X_n)$满足

$$P(\underline{\theta}<\theta<\bar{\theta})=1-\alpha, \tag{7.7}$$

则称区间$(\underline{\theta},\bar{\theta})$是参数 θ 的**置信水平**为 $1-\alpha$ 的**置信区间**,$\underline{\theta}$和$\bar{\theta}$分别称为**置信下限**和**置信上限**,$1-\alpha$ 称为**置信水平**(或置值度).

值得注意的是,区间$(\underline{\theta},\bar{\theta})$的上、下限都是统计量,从而置信区间是随机区间.随着样本观察值的不同,随机区间$(\underline{\theta},\bar{\theta})$产生不同的具体区间.(7.7)式的意义是,随机区间$(\underline{\theta},\bar{\theta})$包含 θ 真值的概率为 $1-\alpha$,而不是 θ 的真值落在区间$(\underline{\theta},\bar{\theta})$内的概率为$1-\alpha$.换句话说,若反复抽样多次(各次抽取的样本容量都是 n),则每组样本观测值都确定一个具体区间$(\underline{\theta},\bar{\theta})$,每个这样的区间要么包含 θ 的真值,要么不包含 θ 的真值,按伯努利大数定理,在所得的几个区间中,包含 θ 真值的约占 $100(1-\alpha)\%$,不包含 θ 真值的约占 $100\alpha\%$.例如,若 $\alpha=0.05$,反复抽样 1000 次,则得到的 1000 个区间中包含 θ 真值的约占 95%,对每组样本观察值确定的具体区间而言,它属于包含 θ 真值的区间的置信概率为 $100\times(1-0.05)\%=95\%$,而不能说一个具体的区间以 95%的概率包含 θ.

下面着重讨论正态总体均值和方差的区间估计问题.

7.4.2 单个正态总体均值的区间估计

设已给定置信水平为 $1-\alpha$,并设$(X_1,X_2,\cdots,X_n)$是正态总体 $N(\mu,\sigma^2)$的样本,$\bar{X}$ 和 S^2 分别是样本均值和样本方差.对总体均值 μ 作区间估计,分总体方差 σ^2 已知和未知两种情况.

1. σ^2 已知情形

由定义 7.6 知,对于给定的置信水平 $1-\alpha$,求 μ 的置信区间的关键是求满足

$$P(\underline{\mu}<\mu<\bar{\mu})=1-\alpha$$

的统计量$\underline{\mu}(X_1,X_2,\cdots,X_n)$和 $\bar{\mu}(X_1,X_2,\cdots,X_n)$,上式中的不等式可视为由含有参数 μ 和某统计量的适当不等式解得.因此,取 μ 的估计量 $\bar{\mu}=\bar{X}$,则有

$$U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1). \tag{7.8}$$

对给定的 $\alpha(0<\alpha<1)$，据正态分布的上侧 α 分位点定义，存在双侧分位点 $\pm u_{\alpha/2}$（见图7.1），使

$$P(|U|<u_{\alpha/2})=P\left(\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right|<u_{\alpha/2}\right)=1-\alpha. \tag{7.9}$$

(7.9)式等价于

$$P\left(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right)=1-\alpha.$$

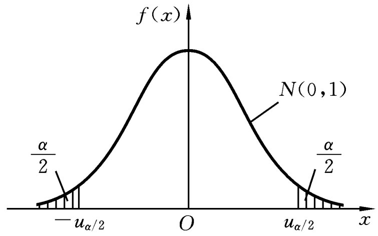

图 7.1

由此得到在 σ^2 已知情形下 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right). \tag{7.10}$$

参考上面的讨论，可归纳区间估计的一般步骤如下：

(1) 取待估参数 θ 的一个点估计 $\hat{\theta}(X_1,X_2,\cdots,X_n)$，并构造一个包含 θ 与 $\hat{\theta}$ 的随机变量 $G(\theta,\hat{\theta})$，使得 $G(\theta,\hat{\theta})$ 的分布为已知且与 θ 无关.

(2) 对于给定的置信水平 $1-\alpha$，定出两个常数 a,b，使得

$$P(a<G(\theta,\hat{\theta})<b)=1-\alpha.$$

(3) 将(2)中式子等价变形为

$$P(\underline{\theta}(X_1,X_2,\cdots,X_n)<\theta<\underline{\theta}(X_1,X_2,\cdots,X_n))=1-\alpha.$$

则得 θ 为置信水平为 $1-\alpha$ 的置信区间 $(\underline{\theta},\overline{\theta})$.

通常称(1)中的 $G(\theta,\hat{\theta})$ 为**枢轴变量**，而称这种利用枢轴变量构造置信区间的方法为**枢轴变量法**.

例 7.17　某企业生产的滚珠直径 X 服从正态分布 $N(\mu,0.0006)$. 现从产品中随机抽取 6 颗检测，测得它们的直径（单位：mm）如下：1.46，1.51，1.49，1.48，1.52，1.51. 试求滚珠平均直径 μ 的置信水平为 95% 的置信区间.

解　这里 $1-\alpha=0.95$，$\alpha=0.05$，$\sigma^2=0.0006$，$n=6$. 查 $N(0,1)$ 分布表，得 $u_{\alpha/2}=u_{0.025}=1.96$，由给出的数据算得 $\overline{x}=1.495$. 故由(7.10)式得平均直径 μ 的置信水平为 95% 的置信区间为

$$(1.495-1.96\times\sqrt{0.0006/6},1.495+1.96\ \sqrt{0.0006/6}),$$

即

$$(1.4754,1.5146), \tag{7.11}$$

这就是说估计滚珠平均直径的值在 1.4754 mm 与 1.5146 mm 之间，这个估计的可信度为 95%，若以此区间的任一值作为 μ 的近似值，其误差不大于

$$2\times1.96\times\sqrt{0.0006/6}\ \text{mm}=0.0392\ \text{mm},$$

这个误差估计的可信度为 95%.

记置信区间(7.10)式的长度为 l，则

$$l=\frac{2\sigma}{\sqrt{n}}u_{\alpha/2}.$$

由此可见 l 与 n,α 的关系是：

(1) 对给定 α，区间长度 l 随 n 的增加而减小；

(2) 对给定 n，区间长度 l 随 α 的减小（即置信水平 $1-\alpha$ 增大）而增大.

注意　在同一置信水平下，置信区间的选取不唯一，例如，在例 7.17 中，令 $\alpha=0.01+$

0.04,由

$$P\left(-u_{0.04}<\frac{\overline{X}-\mu}{\sigma}<u_{0.01}\right)=95\%,$$

查 $N(0,1)$分布表,得 $u_{0.04}=2.33$,$u_{0.01}=1.75$,所以 μ 的另一个置信水平为 95%的置信区间为

$$(\overline{X}-2.33\sigma/\sqrt{n},\overline{X}+1.75\sigma/\sqrt{n}),$$

其区间长度 $l_2=4.08\sigma/\sqrt{n}=0.0408$. 由于(7.11)式的区间长度 $l_1=0.0392$,故 $l_2>l_1$.

在同一置信水平下,置信区间的长度越小,意味着估计的精度越高. 由于随机变量 U 的概率密度函数的图形是单峰且对称的,故当样本容量 n 固定时,(7.10)式所示置信区间是置信水平为 $1-\alpha$ 的所有置信区间中长度最短的,因此我们用它作为 μ 的置信水平为 $1-\alpha$ 的置信区间.

2. σ^2 未知情形

此时不能采用(7.10)式给出的区间,因为其中含未知参数 σ,由于 S^2 是 σ^2 的无偏估计,将(7.8)式中的 σ 换成 $S=\sqrt{S^2}$,由第六章 6.1 节推论 1 知,随机变量

$$t=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1). \qquad (7.12)$$

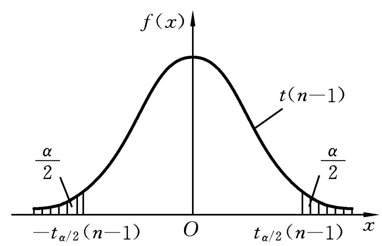

图 7.2

对给定 $\alpha(0<\alpha<1)$,查 t 分布表,可得 $t(n-1)$分布的双侧分位点 $t_{\alpha/2}(n-1)$的值(见图 7.2),使得

$$P(|t|<t_{\alpha/2}(n-1))=P\left(\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right|<t_{\alpha/2}(n-1)\right)=1-\alpha,$$

即

$$P\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)<\mu<\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right)=1-\alpha.$$

所以,在 σ^2 未知情形下 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right). \qquad (7.13)$$

在例 7.17 中,若总体方差未知,则可按(7.13)式求 μ 的置信区间. 这时,$\alpha=0.05$,查 t 分布表得 $t_{\alpha/2}(n-1)=t_{0.025}(5)=2.57$. 由所给数据算出 $s=0.0226$,将上述值代入(7.13)式,得参数 μ 的置信水平为 95%的置信区间为

$$(1.4715,1.5187).$$

实际问题中,总体方差 σ^2 未知的情况居多,故(7.13)式较(7.10)式有更大的实用价值.

7.4.3 单个正态总体方差的区间估计

设总体 $X\sim N(\mu,\sigma^2)$,$(X_1,X_2,\cdots,X_n)$是 X 的样本,对总体方差 σ^2 作区间估计,同样分成 μ 已知和未知两种情形,此处,根据实际情况的需要,只讨论 μ 未知的情况.

由第六章定理 6.2 知,随机变量

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$$

给定 $\alpha(0<\alpha<1)$,依 χ^2 分布的上侧 α 分位点定义(见图 7.3),有

$$P\left(\chi^2_{1-\alpha/2}(n-1)<\frac{(n-1)S^2}{\sigma^2}<\chi^2_{\alpha/2}(n-1)\right)=1-\alpha.$$

查 $\chi^2(n-1)$ 分布表，可得 $\chi^2_{\alpha/2}(n-1)$ 和 $\chi^2_{1-\alpha/2}(n-1)$ 的值，于是，

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}<\sigma^2<\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)=1-\alpha. \tag{7.14}$$

故 σ^2 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

由(7.14)式，可以得到标准差 σ 的一个置信水平为 $1-\alpha$ 的置信区间

$$\left(\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{\alpha/2}(n-1)}},\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}}\right). \tag{7.15}$$

应该指出，当概率密度曲线的图形不对称时，如 χ^2 分布和 F 分布，习惯上仍取对称的分位点(如用图 7.3 中的分位点 $\chi^2_{1-\alpha/2}(n-1)$ 与 $\chi^2_{\alpha/2}(n-1)$)来确定置信区间，但这样确定的置信区间的长度并不是最短，由于求最短置信区间的计算太繁杂，通常采用(7.14)式与(7.15)式来估计方差与标准差的置信水平为$1-\alpha$ 的置信区间.

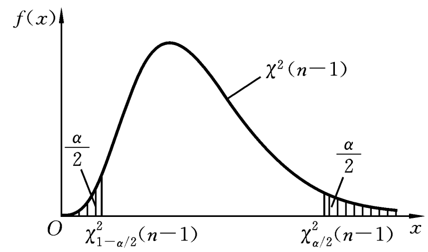

图 7.3

例 7.18　从自动车床加工的一批零件中随机地抽取 16 件，测得各零件的长度如下(单位：cm)：

2.15　2.10　2.12　2.10　2.14　2.11　2.15　2.13

2.13　2.11　2.14　2.13　2.12　2.13　2.10　2.14

设零件长度服从正态分布，试求零件长度标准差 σ 的置信水平为 95%的置信区间.

解　这里 $1-\alpha=95\%$，$\frac{\alpha}{2}=0.025$，$n=16$，查 χ^2 分布表得 $\chi^2_{0.025}(15)=27.488$，$\chi^2_{0.975}=6.262$，经计算得

$$\overline{X}=2.125,\quad S^2=0.000293.$$

由(7.15)式得零件长度标准差的置信水平为 95%的置信区间为

$$\left(\frac{\sqrt{15}}{\sqrt{27.488}}\times\sqrt{0.000293},\frac{\sqrt{15}}{\sqrt{6.262}}\times\sqrt{0.000293}\right),$$

也即(0.0127,0.0265).

7.4.4　两个正态总体均值差的区间估计

在实际中常遇到这样的情况，产品的某一质量指标 X 服从正态分布 $N(\mu_1,\sigma_1^2)$，由于原料、设备、工艺及操作人员的变动，可能引起总体均值、方差有所变动. 设变化后的指标 Y 服从正态分布 $N(\mu_2,\sigma_2^2)$. 为知道这些变化有多大，需要考虑两个正态总体均值差 $\mu_1-\mu_2$ 或方差比 σ_1^2/σ_2^2 的区间估计问题.

设已给定置信水平为 $1-\alpha$，并设$(X_1,X_2,\cdots,X_{n_1})$是总体 $X\sim N(\mu_1,\sigma_1^2)$的样本，$(Y_1,Y_2,\cdots,Y_{n_2})$是总体 $Y\sim N(\mu_2,\sigma_2^2)$的样本，且这两个样本相互独立. 记 $\overline{X}$，$\overline{Y}$ 分别是总体 X 和 Y 的样本均值，S_1^2，S_2^2 分别是总体 X 和 Y 的样本方差. 对总体均值 $\mu_1-\mu_2$ 作区间估计，分如下两种情况讨论.

1. σ_1^2 **和** σ_2^2 **均为已知**

由于 $\overline{X}$ 和 $\overline{Y}$ 分别是 μ_1 和 μ_2 的无偏估计,故 $\overline{X}-\overline{Y}$ 是 $\mu_1-\mu_2$ 的无偏估计,由 $\overline{X}$ 与 $\overline{Y}$ 相互独立及 $\overline{X}\sim N(\mu_1,\sigma_1^2/n_1)$,$\overline{Y}\sim N(\mu_2,\sigma_2^2/n_2)$得

$$\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,\sigma_1^2/n_1+\sigma_2^2/n_2),$$

所以随机变量

$$U=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1),$$

类似于单个总体的情况,可求得 $\mu_1-\mu_2$ 的一个置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\overline{Y}-\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\frac{\alpha}{2}},\overline{X}-\overline{Y}+\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}u_{\frac{\alpha}{2}}\right). \tag{7.16}$$

2. σ_1^2 **和** σ_2^2 **均未知,但** $\sigma_1^2=\sigma_2^2=\sigma^2$

此时,由第六章定理 6.2 的推论 3,有

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2),$$

其中,$S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$,$S_w=\sqrt{S_w^2}$.从而得 $\mu_1-\mu_2$ 的一个置信水平为 $1-\alpha$ 的置信区间为

$$\left(\overline{X}-\overline{Y}-t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},\overline{X}-\overline{Y}+t_{\alpha/2}(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right). \tag{7.17}$$

在(7.16)式、(7.17)式所示的 $\mu_1-\mu_2$ 的置信区间中,若置信下限大于零,则可认为 $\mu_1>\mu_2$;若置信上限小于零,则可认为 $\mu_1<\mu_2$;若置信区间含零,则认为 μ_1 与 μ_2 没有显著差别.

例 7.19　为提高某一化工生产过程的得率,试采用一种新的催化剂.为慎重起见,在实验工厂进行试验.设采用原来的催化剂进行了 $n_1=8$ 次试验,得到得率的平均值 $\bar{x}=91.73$,样本方差 $s_1^2=3.98$;又采用了新的催化剂进行了 $n_2=8$ 次试验,得到得率的平均值 $\bar{y}=93.75$,样本方差 $s_2^2=4.02$,假定两总体都服从正态分布,且方差相等,试求总体均值差 $\mu_1-\mu_2$ 的置信水平为 95%的置信区间.

解　按实际情况,可以认为两总体的样本是相互独立的.依题意,$\alpha=0.05$,$n_1=n_2=8$.查 t 分布表可得 $t_{0.025}(14)=2.1448$,计算得

$$s_w^2=(7\times 3.89+7\times 4.02)/14=3.955,\quad s_w=\sqrt{s_w^2}=1.989,$$

于是,由(7.17)式得所求两总体均值差 $\mu_1-\mu_2$ 的置信水平为 95%的置信区间为

$$\left(\bar{x}-\bar{y}-s_w\times t_{0.025}(14)\sqrt{\frac{1}{8}+\frac{1}{8}},\bar{x}-\bar{y}+s_w\times t_{0.025}(14)\sqrt{\frac{1}{8}+\frac{1}{8}}\right)$$
$$=(-2.02-2.13,-2.02+2.13)=(-4.15,0.11).$$

由于所得置信区间包含零,因此在实际中我们就认为采用这两种催化剂所得到的得率的均值没有显著差别.

7.4.5　两个正态总体方差比的区间估计

设$(X_1,X_2,\cdots,X_{n_1})$是总体 $X\sim N(\mu_1,\sigma_1^2)$的样本,$(Y_1,Y_2,\cdots,Y_{n_2})$是总体 $Y\sim N(\mu_2,\sigma_2^2)$的

样本，且两组样本相互独立. 要求两总体方差比 σ_1^2/σ_2^2 的置信水平为 $1-\alpha$ 的置信区间，我们仅讨论 μ_1 和 μ_2 均未知的情况.

由第六章定理 6.2 的推论 2 知，随机变量

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1),$$

且分布 $F(n_1-1,n_2-1)$ 不依赖参数 σ_1^2 和 σ_2^2. 由 F 分布的上侧 α 分位点定义知（见图 7.4），

$$P\left\{F_{1-\alpha/2}(n_1-1,n_2-1)<\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}<F_{\alpha/2}(n_1-1,n_2-1)\right\}=1-\alpha,$$

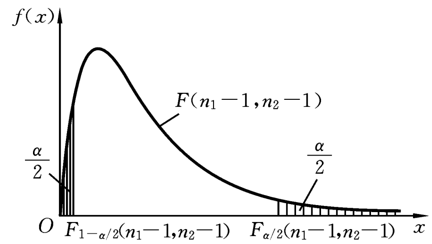

图 7.4

亦即

$$P\left\{\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right\}=1-\alpha.$$

于是得 σ_1^2/σ_2^2 的一个置信水平为 $1-\alpha$ 的置信区间为

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right). \tag{7.18}$$

例 7.20　两台车床生产同一种型号的滚珠，已知两车床生产的滚珠直径 X,Y 分别服从正态分布 $N(\mu_1,\sigma_1^2)$，$N(\mu_2,\sigma_2^2)$，其中 $\mu_i,\sigma_i^2(i=1,2)$ 均未知. 由甲、乙两车床的产品中分别抽出 25 个和 15 个测量直径，并计算得 $s_1^2=6.38$，$s_2^2=5.15$，求两总体方差比 σ_1^2/σ_2^2 的置信水平为 90%的置信区间.

解　依题意，$\alpha=0.1$，$n_1=25$，$n_2=15$，查 F 分布表得

$$F_{0.05}(24,14)=2.35,\quad F_{0.95}(24,14)=\frac{1}{F_{0.05}(14,24)}=\frac{1}{2.13}.$$

故由(7.18)式得 σ_1^2/σ_2^2 的一个置信水平为 90%的置信区间为

$$\left(\frac{6.38}{5.15}\times\frac{1}{2.35},\frac{6.38}{5.15}\times 2.13\right),$$

也即

$$(0.527,2.639).$$

由于该置信区间包含数 1，在实际中我们认为 σ_1^2,σ_2^2 两者没有显著差别，即认为两台机床加工的精度差不多.

7.4.6　单侧置信区间

前面讨论参数区间估计，对未知参数给出的置信区间都是同时具有置信上、下限的双侧置信区间. 在某些实际问题中，往往只需要估计未知参数的下限或上限. 例如，设备、元件等的寿命，越长越好，故我们常会关心平均寿命的下限；相反，产品不合格率越低越好，这时我们又会关心不合格率的上限. 这就需要考虑形如 $(\underline{\theta},+\infty)$ 或 $(-\infty,\overline{\theta})$ 的置信区间. 为此，我们引出单侧置信区间的概念.

定义 7.7　设总体 X 的分布函数为 $F(x;\theta)$，其中 θ 是未知参数. 又设 $(X_1,X_2,\cdots,X_n)$ 是 X 的样本. 对于给定的 $\alpha(0<\alpha<1)$，若统计量 $\underline{\theta}=\underline{\theta}(X_1,X_2,\cdots,X_n)$ 满足

$$P(\theta>\underline{\theta})=1-\alpha,$$

则称随机区间 $(\underline{\theta},+\infty)$ 是 θ 的置信水平为 $1-\alpha$ 的**单侧置信区间**，$\underline{\theta}$ 称为**单侧置信下限**；若统计

量$\bar{\theta}=\bar{\theta}(X_1,X_2,\cdots,X_n)$满足

$$P(\theta<\bar{\theta})=1-\alpha,$$

则称随机区间$(-\infty,\bar{\theta})$是θ的置信水平为$1-\alpha$单侧置信区间,$\bar{\theta}$称为**单侧置信上限**.

求单侧置信区间的方法与求双侧置信区间的方法类似,下面通过一个例子来介绍单侧置信限的求法.

例 7.21　从一批灯泡中随机地取 5 只作寿命试验,测得寿命(单位:h)为

$$1050,\ 1100,\ 1120,\ 1250,\ 1280$$

设灯泡寿命服从正态分布,求灯泡寿命平均值的置信水平为 0.95 的单侧置信下限.

解　设总体$X\sim N(\mu,\sigma^2)$,其中μ,σ^2均未知,因为随机变量

$$\frac{(\bar{X}-\mu)}{S}\sqrt{n}\sim t(n-1),$$

查t分布表,可得满足条件

$$P\left(\frac{\bar{X}-\mu}{S}\sqrt{n}<t_\alpha(n-1)\right)=1-\alpha$$

的分位点值$t_\alpha(n-1)$,于是μ的置信水平为$1-\alpha$的单侧置信区间为

$$\left(\bar{X}-t_\alpha(n-1)\frac{S}{\sqrt{n}},+\infty\right). \tag{7.19}$$

由题设知$\alpha=0.05$,$n=5$,$\bar{x}=1160$,$s^2=9950$,查t分布表,$t_{0.05}(4)=2.1318$,故μ的置信水平为$1-\alpha$的单侧置信下限为

$$\underline{\mu}=\bar{x}-\frac{s}{\sqrt{n}}t_\alpha(n-1)=1065.$$

类似地,可求出μ的单侧置信上限为$\bar{\mu}=\bar{X}+t_\alpha(n-1)\frac{S}{\sqrt{n}}$;单侧置信区间为$\left(-\infty,\bar{X}+t_\alpha(n-1)\frac{S}{\sqrt{n}}\right)$.

注意　对总体方差及两总体的均值差或方差比的单侧置信区间可类似求得.至于对参数是作双侧区间估计还是作单侧区间估计,则要由实际需要来决定.

习　题　七

7.1　填空题.

(1) 设总体X的概率密度为

$$f(x;\theta)=\begin{cases}e^{-(x-\theta)}, & x\geqslant 0,\\ 0, & x<0,\end{cases}$$

而$X_1,X_2,\cdots,X_n$是来自总体X的样本,则未知参数θ的矩估计量为______.

(2) 设 1,0,0,1 是来自两点分布总体$X\sim B(1,p)$的样本观察值,则参数$q=1-p$的矩估计值为______.

(3) 设$\bar{X}$和S^2是来自二项分布总体$X\sim B(m,p)$的样本均值和样本方差,样本容量为n,若用$\bar{X}+kS^2$作为mp^2的无偏估计,则$k=$______.

(4) 设$\hat{\theta}_1$和$\hat{\theta}_2$都是θ的无偏估计,且$D(\hat{\theta}_1)=\sigma^2$,$D(\hat{\theta}_2)=\sigma^2$,构造一个新的无偏估计

$$\hat{\theta}=c\,\hat{\theta}_1+(1-c)\hat{\theta}_2,\quad 0\leqslant c\leqslant 1,$$

如果$\hat{\theta}_1$和$\hat{\theta}_2$相互独立,为使$D(\hat{\theta})$达最小,应有$c=$______.

（5）设总体 X 的概率密度为

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

其中，θ 为未知参数，$X_1,X_2,\cdots,X_n$ 为来自总体 X 的样本，令 $c\sum\limits_{i=1}^{n}X_i^2$ 是 θ^2 的无偏估计量，则 $c=$______。

（6）在天平上重复称量一重为 a 的物品，假设各次称量结果相互独立，且同服从正态分布 $N(a,0.2^2)$．若以 $\overline{X}_n$ 表示 n 次称量结果的算术平均值，则为使

$$P(|\overline{X}_n-a|<0.1)\geqslant 0.95,$$

n 的最小值应不小于自然数______．

（7）设 $x_1,x_2,\cdots,x_n$ 为来自总体 $N(\mu,\sigma^2)$ 的样本观测值，样本均值 $\overline{x}=9.5$，参数 μ 的置信度为 95% 的双侧置信区间的置信上限为 10.8，则 μ 的置信度为 95% 的双侧置信区间为______。

7.2　单项选择题．

（1）设总体 $X\sim N(\mu,\sigma^2)$，其中 μ 已知，$\sigma^2\neq 0$ 为未知参数，$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，$\overline{X}$ 为样本均值，则 σ^2 的极大似然估计量为（　　）．

（A）$\hat{\sigma}^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$　　（B）$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$

（C）$\hat{\sigma}^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\mu)^2$　　（D）$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$

（2）设 n 个随机变量 $X_1,X_2,\cdots,X_n$ 独立同分布，

$$DX_1=\sigma^2,\quad \overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2,$$

则（　　）．

（A）S 是 σ 的无偏估计量　　（B）S 是 σ 的极大似然估计量

（C）S 是 σ 的一致估计量　　（D）S 与 $\overline{X}$ 相互独立

（3）设 $X_1,X_2,\cdots,X_n$ 是总体 $X\sim N(0,\sigma^2)$ 的样本，则未知参数 σ^2 的无偏估计量为（　　）．

（A）$\hat{\sigma}^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}X_i^2$　　（B）$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2$

（C）$\hat{\sigma}^2=\dfrac{1}{n+1}\sum\limits_{i=1}^{n}X_i^2$　　（D）$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$

（4）设正态总体均值 μ 的置信区间长度 $l=2\dfrac{S}{\sqrt{n}}t_\alpha(n-1)$，则其置信水平为（　　）．

（A）$1-\alpha$　　（B）α　　（C）$1-\dfrac{\alpha}{2}$　　（D）$1-2\alpha$

（5）设总体 $X\sim N(\mu,\sigma^2)$，当 σ^2 未知时，μ 的置信区间长度为 l_1，在置信水平不变的条件下，用 σ^2 的无偏估计 S^2 作为 σ^2 的已知值，所得 μ 的置信区间长度为 l_2，则（　　）．

（A）$l_1=l_2$　　（B）$l_1>l_2$　　（C）$l_1<l_2$　　（D）l_1 与 l_2 无一定序关系

7.3　为考察某种轮胎的使用寿命，随机地抽取 12 只轮胎试用，测得它们的寿命（单位：万公里）如下：

4.68　4.85　4.32　4.85　4.61　5.02　4.60　4.58　4.72　4.38　4.70　5.20

试求此种轮胎的平均寿命 μ 和寿命方差 σ^2 的矩估计值，并计算样本方差 S^2．

7.4　设$(X_1,X_2,\cdots,X_n)$为总体 X 的样本，求下述各总体的概率密度或分布列中的未知参数的矩估计量．

（1）

$$f(x)=\begin{cases}\theta c^\theta x^{-(\theta+1)}, & x>c,\\ 0, & \text{其他},\end{cases}$$

其中 $c>0$ 为已知，$\theta>1$ 为未知参数．

(2)
$$f(x)=\begin{cases}\sqrt{\theta}\,x^{\sqrt{\theta}-1}, & 0\leqslant x\leqslant 1,\\ 0, & \text{其他},\end{cases}$$
其中 $\theta>0$ 为未知参数.

(3)
$$f(x)=\begin{cases}\dfrac{x}{\theta^2}\mathrm{e}^{-x^2/(2\theta^2)}, & x>0,\\ 0, & \text{其他},\end{cases}$$
其中 $\theta>0$ 为未知参数.

(4)
$$f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-(x-\mu)/\theta}, & x\geqslant\mu,\\ 0, & \text{其他},\end{cases}$$
其中 $\theta>0,\theta,\mu$ 为未知参数.

(5) $P(X=x)=\dfrac{\lambda^x}{x\,!}\mathrm{e}^{-\lambda},x=0,1,2,\cdots,\lambda,\lambda>0$,

其中 λ 为未知参数.

7.5　求上题中各未知参数的极大似然估计量.

7.6　设总体 X 的分布列为

X	1	2	3
P_k	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中 $\theta(0<\theta<1)$为未知参数,已知取得了样本观察值 $x_1=1,x_2=2,x_3=1$,求 θ 的矩估计值和极大似然估计值.

7.7　设总体 X 的概率密度为
$$f(x;\theta)=\begin{cases}\dfrac{1}{1-\theta}, & \theta<x<1,\\ 0, & \text{其他},\end{cases}$$
其中 θ 为未知参数,$X_1,X_2,\cdots,X_n$为来自总体 X 的样本。求:(1) θ 的矩估计量;(2) θ 的极大似然估计量。

7.8　设$(X_1,X_2,\cdots,X_n)$是正态总体 $N(\mu,\sigma^2)$的样本,其中 μ,σ^2 均未知,试求 $P(\overline{X}<t)$的极大似然估计.

7.9　设一名货运司机在 5 年内发生交通事故的次数服从泊松分布,对 136 名货运司机的调查结果见下表,其中 r 表示一货运司机 5 年内发生交通事故的次数,m 表示观察到的货运司机人数.

r	0	1	2	3	4	5
m	56	43	22	8	5	2

(1) 求一名货运司机在 5 年内发生交通事故的平均次数的极大似然估计;

(2) 求一名货运司机在 5 年内不发生交通事故的概率的极大似然估计.

7.10　某工程师为了解一台天平的质量,用该天平对一物体的质量做 n 次测量,该物体的质量 μ 是已知的。设 n 次测量结果 $X_1,X_2,\cdots,X_n$相互独立且均服从正态分布 $N(\mu,\sigma^2)$。该工程师记录的是 n 次测量的绝对误差 $Z_i=|X_i-\mu|(i=1,2,\cdots,n)$,试利用 $Z_1,Z_2,\cdots,Z_n$估计 σ。(1) 求 Z_i的概率密度;(2) 利用一阶矩求 σ 的矩估计量;(3)求 σ 的极大似然估计量。

7.11　设 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim U(0,\theta)$的样本,求总体均值 μ 与方差 σ^2 的极大似然估计.

7.12　设总体 X 的概率密度为
$$f(x;\theta)=\begin{cases}\dfrac{3x^2}{\theta^3}, & 0<x<\theta,\\ 0, & \text{其他},\end{cases}$$
其中 $\theta\in(0,+\infty)$为未知参数,X_1,X_2,X_3为来自总体 X 的样本,令 $T=\max\{X_1,X_2,X_3\}$。(1) 求 T 的概率密度;(2) 确定 a,使得 aT 为 θ 的无偏估计。

7.13　设总体 X 在区间$(0,2\theta)$上服从均匀分布,X_1,X_2,X_3 为其样本. 试证明样本均值 $\overline{X}$ 与 $\dfrac{2}{3}\max\limits_{1\leqslant i\leqslant 3}X_i$

都是 θ 的无偏估计量，并比较它们谁更有效.

7.14　设总体 X 服从参数为 λ 的泊松分布，即

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

$(X_1,X_2,\cdots,X_n)$ 是 X 的样本，试求 λ^2 的无偏估计.

7.15　设总体 $X\sim N(\mu,\sigma^2)$，$(X_1,X_2,\cdots,X_n)$ 是 X 的样本，试确定常数 c，使 $c\sum\limits_{i=1}^{n-1}(X_{i+1}-X_i)^2$ 为 σ^2 的无偏估计.

7.16　设 $\hat{\theta}$ 是参数 θ 的无偏估计，且有 $D\hat{\theta}>0$，证明 $\hat{\theta}^2=(\hat{\theta})^2$ 不是 θ^2 的无偏估计.

7.17　设从均值为 μ，方差为 $\sigma^2(\sigma^2>0)$ 的总体中，分别抽取容量为 n_1,n_2 的两独立样本，$\overline{X}_1$ 和 $\overline{X}_2$ 分别是两样本的均值. 试证：对于任意常数 $a,b(a+b=1)$，$Y=a\overline{X}_1+b\overline{X}_2$ 都是 μ 的无偏估计，并确定常数 a,b 使 $D(Y)$ 达到最小.

7.18　某旅行社为调查当地每一旅游者的日平均消费额，随机地访问了 100 名旅游者，得知日平均消费额 $\bar{x}=80$ 元，根据经验，已知旅游者日消费额服从正态分布，且方差 $\sigma^2=12^2$，求该地旅游者日平均消费额 μ 的置信水平为 95％的置信区间.

7.19　某批牛奶中被混入了一种有害物质三聚氰胺，现从中随机抽取 220 盒进行检测，得到每公斤牛奶中三聚氰胺的含量（单位：mg/kg）如下：

0.86　1.53　1.57　1.81　0.99　1.09　1.29　1.78　1.29　1.58

假设可以认为这批牛奶中三聚氰胺的含量服从正态分布 $N(\mu,\sigma^2)$，试求：

(1) 含量均值 μ 的置信水平为 90％的置信区间；

(2) 含量标准差 σ 的置信水平为 90％的置信区间.

7.20　某地为研究农业家庭与非农业家庭的人口状况，独立、随机地调查了 50 户农业居民和 60 户非农业居民. 经计算知农业居民家庭平均每户 4.5 人，非农业居民家庭平均每户 3.75 人，已知农业居民家庭人口分布为 $N(\mu,1.8^2)$，非农业居民家庭人口分布为 $N(\mu,2.1^2)$，试求 $\mu_1-\mu_2$ 的置信水平为 99％的置信区间.

7.21　为比较甲、乙两种型号的步枪子弹的枪口速度，随机地取甲型子弹 10 发，算得枪口速度的平均值 $\bar{x}=500$(m/s)，标准差 $s_1=1.10$(m/s)；随机地取乙型子弹 20 发，得枪口速度平均值 $\bar{y}=496$(m/s)，标准差 $s_2=1.20$(m/s). 设两总体近似地服从正态分布，并且方差相等，求两总体均值 $\mu_1-\mu_2$ 的置信水平为 95％的置信区间.

7.22　设某产品的生产工艺发生了改变，在改变前后分别测得了若干件产品的技术指标，其结果为

改变前：21.6，　22.8，　22.1，　21.2，　20.5，　21.9，　21.4

改变后：24.1，　23.8，　24.7，　24.0，　23.7，　24.3，　24.5，　23.9

假设该产品的技术指标服从正态分布，方差 σ^2 未知且在工艺改变前后不变，试估计工艺改变后，该技术指标的置信水平为 95％的平均值变化的范围.

7.23　在上题中若工艺改变前后总体方差 σ^2 有所改变，试计算工艺改变后的总体方差 σ_1^2 与工艺改变前的总体方差 σ_2^2 的比 σ_1^2/σ_2^2 的置信水平为 95％的置信区间.

7.24　研究机器 A 和机器 B 生产的钢管内径，随机抽取机器 A 生产的管子 18 只，测得样本方差 $s_1^2=0.34(\mathrm{mm}^2)$；随机抽取机器 B 生产的管子 13 只，测得样本方差 $s_2^2=0.29(\mathrm{mm}^2)$. 设两样本相互独立，且由机器 A、机器 B 生产的管子的内径分别服从正态分布 $N(\mu_1,\sigma_1^2)$，$N(\mu_2,\sigma_2^2)$，这里 $\mu_i,\sigma_i^2(i=1,2)$ 均未知，试求方差比 σ_1^2/σ_2^2 的置信水平为 0.90 的置信区间.

7.25　从一批灯泡中随机地抽取 16 只灯泡作寿命试验，测得它们的寿命（单位：h）如下：

1502　1480　1485　1511　1514　1527　1603　1480

1532　1580　1490　1470　1520　1505　1485　1540

设灯泡寿命 X 服从正态分布，试求灯泡平均寿命的置信水平为 95％的置信下限.

7.26　测得两个民族中各 5 位成年人的身高（单位：cm）如下：

A 民族	162.6	170.2	172.7	165.1	157.5
B 民族	175.3	177.8	167.6	180.3	182.9

设两样本分别来自总体 $N(\mu_1,\sigma^2)$,$N(\mu_2,\sigma^2)$,μ_1,μ_2,σ^2 未知,两样本独立,求 $\mu_1-\mu_2$ 的置信水平为0.90的置信区间.