概率论与数理统计笔记

作者：KoSaking

组织：计合二刺螈根据地

时间：2023 年 2 月 2 日

版本：5.0.0

教材：刘建亚、吴臻《概率论与数理统计》

我们所经历的每个平凡的日常，实际上可能是接连不断发生的奇迹。 ——《日常》

目 录

第 1 章 随机事件及其概率

1.1 随机事件及相关定义

日常生活中的现象可分为**确定性现象**和**随机现象**。确定性现象又称**必然现象**，指一定发生或一定不发生的现象；随机现象又称**偶然现象**，指可能发生也可能不发生的现象。我们把由大量同类随机现象所呈现出来的规律性称为**统计规律性**，而概率论与数理统计就是研究这种统计规律性的一门数学学科。

在正式学习概率论前，我们需要了解一些基本概念。

1. **试验**：为确定随机现象的规律性而进行多次的观察、调查、实验等工作的统称。

   性质 试验具有以下特征。

   (a). 在相同条件下可多次进行。

   (b). 试验结果不止一个。

   (c). 无法预测会出现哪一个结果。

2. **随机试验**：即试验，记为 E。
3. **事件**：试验中的每一种可能的结果称为一个事件。

   注 事件本质上是样本空间（即必然事件）的子集。

4. **随机事件**：可能发生也可能不发生的事件，通常用大写字母 $A, B, C, \cdots$ 来表示。

笔记 随机事件有两种极端情况：一是必然事件，指在每次试验中都发生的事件，记为 Ω；二是不可能事件，指在每次试验中都不可能发生的事件，记为 $\varnothing$。

5. **基本事件**：又称为样本点，指相对于试验目的来说，不能再分或不必再分的事件，记为 ω。
6. **复合事件**：由若干个基本事件复合而成的事件。
7. **样本空间**：所有基本事件组成的集合称为试验 E 的样本空间，记为 Ω。

   注 样本空间可以是无限集。

例题 1.1 投掷一颗骰子并观察其正面朝上的点数，共有 6 种等可能结果，样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$。

例题 1.2 在一批日光灯中随机抽取一只并测试其寿命，以 t（单位：小时）表示日光灯的使用寿命，样本空间为 $\Omega = \{t \mid t \geqslant 0\}$。

例题 1.3 向区间 $[a, b]$ 投掷一质子，样本空间为 $\Omega = [a, b]$。

例题 1.4 向平面中投掷一质子，样本空间为 $\Omega = \{(x, y) \mid x, y \in \mathbf{R}\}$。

例题 1.5 投掷一颗骰子，正面朝上的点数为偶数的事件记为 $A = \{2, 4, 6\}$。

1.2 随机事件的运算

在同一随机试验中，事件不止一个，其间或多或少存在着联系。类似于算术、几何与集合，随机事件也可以进行运算，下面将会详细介绍。

1. **包含**：若事件 A 发生必定有事件 B 发生，则称事件 A 包含于事件 B，记为 $A \subset B$。

笔记 对任意事件，都有 $\varnothing \subset A \subset \Omega$ 成立。

2. **相等**：若 $A \subset B$ 且 $B \subset A$，则称 A 和 B 相等，记为 $A = B$。
3. **并**：若事件 A 或事件 B 之中至少有一个发生，则称该事件为 A 和 B 的并事件或和事件，记为 $A \cup B$ 或 $A + B$。

笔记 对任意事件 A 和 B，都有 $A \subset A + B$、$A + A = A$、$A + \varnothing = A$、$A + \Omega = \Omega$ 成立。

4. **交**：若事件 A 和事件 B 同时发生，则称该事件为 A 和 B 的交事件或积事件，记为 $A \cap B$ 或 AB。

笔记 对任意事件 A 和 B，都有 $AB \subset A$、$A \cap A = A$、$A \cap \varnothing = \varnothing$、$A \cap \Omega = A$ 成立。

5. **差**：若事件 A 发生但事件 B 不发生，则称该事件为 A 和 B 的差事件，记为 $A - B$。

笔记 对任意事件 A 和 B，都有 $A-B=A-AB=A-A\overline{B}$ 成立。

6. **互不相容事件**：若事件 A 和事件 B 不能同时发生，则称 A 和 B 为互不相容事件，即 $AB=\varnothing$。
7. **对立事件**：若事件 A 和事件 B 必有一个发生且仅有一个发生，则称 A 和 B 为对立事件或互逆事件，即 $AB=\varnothing$ 且 $A+B=\Omega$。

笔记 若事件 A 和事件 B 对立，则 A 和 B 互不相容，反之不然。

8. **逆**：对于对立的事件 A 和事件 B，B 为 A 的逆事件，记为 $B=\overline{A}$。
9. **完备事件组**：设事件 $A_1,A_2,\cdots,A_n$ 之间两两互不相容，且 $\bigcup\limits_{i=1}^{n}A_i=\Omega$，则这一系列事件称为完备事件组。

注 互不相容事件与对立事件的区别。

1. 互不相容适用于多个事件，而对立只适用于两个事件。
2. 对于两个互不相容的事件，二者可以同时不发生。

基于上述运算，我们可以得到如下的事件运算律：

公式	名称
$A\cup B=B\cup A$ $A\cap B=B\cap A$	交换律
$(A\cup B)\cup C=A\cup(B\cup C)$ $(A\cap B)\cap C=A\cap(B\cap C)$	结合律
$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$ $(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$	分配律
$\overline{A\cap B}=\overline{A}\cup\overline{B}$ $\overline{A\cup B}=\overline{A}\cap\overline{B}$ $\overline{\bigcup\limits_{i=1}^{n}A_i}=\bigcap\limits_{i=1}^{n}\overline{A_i}$（$\overline{\bigcup\limits_{i=1}^{\infty}A_i}=\bigcap\limits_{i=1}^{\infty}\overline{A_i}$） $\overline{\bigcap\limits_{i=1}^{n}A_i}=\bigcup\limits_{i=1}^{n}\overline{A_i}$（$\overline{\bigcap\limits_{i=1}^{\infty}A_i}=\bigcup\limits_{i=1}^{\infty}\overline{A_i}$）	德摩根律

表 1.1: 事件的运算律

1.3 随机事件的概率

在研究随机现象时，我们常用**概率**来刻画事件出现的可能性大小，其为事件本身固有而不随人的意志改变，且一定符合一般常情。下面先给出**频率**的概念。

定义 1.1 (频率)

设在相同的条件下，进行了 n 次试验，其中事件 A 出现了 m 次，则称

$$f_n(A)=\frac{m}{n}=\frac{\text{事件 } A \text{ 出现的次数}}{\text{试验次数}}$$

为随机事件 A 在 n 次试验中出现的频率，m 称为频数。♣

由定义 1.1 易得，对任一事件 A，都有 $0\leqslant f_n(A)\leqslant 1$，且 $f_n(\Omega)=1$，$f_n(\varnothing)=0$。若设 $A_1,A_2,\cdots,A_n$ 为完备事件组，则有

$$f_n\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}f_n(A_i)$$

实际上，频率随试验的具体结果而定，一般不是确定的值，但当试验的次数相当大时，频率总稳定于某一常数附近，即频率具有**稳定性**，而我们称该常数为概率。

定义 1.2 (概率的统计定义)

在大量重复试验中，若事件 A 发生的频率稳定地在某一常数 p 的附近摆动，则称该常数 p 为事件 A 发生的概率，记为 $P(A)$，即 $P(A)=p$。♣

笔记 频率是变动的，概率是不变的，且概率是先于试验而稳定存在的；在试验次数相当大的情况下，可近似认为 $f_n(A)\approx P(A)$。

对于本身具有某种“对称性”的事件，我们可直接计算其概率，其随机试验具有以下特点：

1. 试验的样本空间中的样本点只有有限个，即基本事件的总数有限；
2. 在每一次试验中，每个基本事件发生的可能性相等。

通常我们把具有上述特点的随机现象的数学模型称为**古典概型**，由此可得出概率的古典定义。

定义 1.3 (概率的古典定义)

设试验结果共有 n 个基本事件 $\omega_1,\omega_2,\cdots,\omega_n$，而且这些事件发生的可能性相等. 事件 A 由其中的 m 个基本事件组成，则事件 A 的概率为

$$P(A)=\frac{\text{事件 } A \text{ 包含的基本事件数 } (m)}{\text{基本事件总数 } (n)}$$

♣

例题 1.6 书架上有一套五卷的选集，求自左向右或自右向左排列为第一、二、三、四、五卷的概率。

解 五卷排列共 $5!=120$ 种情况，故概率为 $\dfrac{2}{120}=\dfrac{1}{60}$。

古典概型的概率定义只适用于试验的可能结果为有限个，且具有等可能性的情况，因此有一定的局限性。若将概率的古典定义推广到样本空间中有无穷个样本点的情况，便可得到**几何概率**的定义。

定义 1.4 (几何概率)

如果试验 E 的可能结果可以几何地表示某一维、二维、三维甚至更高维的区域 Ω 中的一个点，并且点落在 Ω 中某区域 A 的概率与 A 的测度（长度、面积、体积等）成正比，而与 A 的位置与形状无关，则随机点落在区域 A 的概率为

$$P(A)=\frac{A\text{ 的测度}}{\Omega\text{ 的测度}}=\frac{mA}{m\Omega}$$

称上式定义的概率为几何概率。♣

例题 1.7 平面上画着一些平行线，它们之间的距离都为 d，向此平面上任投一长度为 $l<d$ 的针，求此针与任一平行线相交的概率。

解 设事件 $A=\{$针与线相交$\}$，x 为针中点离最近线的距离，φ 为针与线的夹角，则有

$$\Omega=\left\{(x,\varphi)\,\middle|\,0\leqslant x\leqslant\frac{d}{2},\,0\leqslant\varphi\leqslant\pi\right\}$$

满足题意条件为 $\dfrac{x}{\sin\varphi}\leqslant\dfrac{l}{2}$，即条件区域为

$$A=\left\{(x,\varphi)\,\middle|\,0\leqslant x\leqslant\frac{l}{2}\sin\varphi,\,0\leqslant\varphi\leqslant\pi\right\}$$

在平面上做出 Ω 和 A 的区域，可得

$$P(A)=\frac{\displaystyle\int_0^{\pi}\frac{l}{2}\sin\varphi\mathrm{d}\varphi}{\pi\times\dfrac{d}{2}}=\frac{2l}{\pi d}$$

笔记 设做了 N 次投针试验，其中有 n 次针与线相交，则由已知的概率可得 $\frac{n}{N} \approx \frac{2l}{\pi d}$，可估得圆周率为 $\pi \approx \frac{2lN}{nd}$，且试验次数越多精度越高。此方法称为随机模拟法，也称为蒙特卡洛法。

例题 1.8 甲乙两人约定在 6 点至 7 点在某一地点见面，先到者等 15 分钟，超过 15 分钟则不再等待。两人在一小时内的任意时刻到达，问甲乙两人能见面的概率。

解 设事件 $A = \{$两人见面$\}$，x 为甲到达的时刻，y 为乙到达的时刻，易得两人见面条件为 $|y - x| \leqslant 15$，在平面直角坐标系上做出满足该条件的区域及全集正方形，故两人见面的概率为

$$P(A) = \frac{60 \times 60 - 2 \times \frac{1}{2} \times 45 \times 45}{60 \times 60} = 0.43756$$

由于随机事件概率的统计定义、古典定义和几何概率都有其局限性，苏联数学家柯尔莫哥洛夫于 1933 年给出了概率的公理化定义。

定义 1.5 (概率的公理化定义)

设随机试验 E 的样本空间为 Ω，若对于 E 的每一个事件 A，都有一实数 $P(A)$ 与之对应，并且满足以下三条公理，则称 $P(A)$ 为事件 A 的概率。

1. 非负性：对于任意一个事件 A，$0 \leqslant P(A) \leqslant 1$。
2. 规范性：$P(\Omega) = 1$。
3. 可列可加性：对于两两互不相容事件 $A_1, A_2, \cdots, A_n, \cdots$，有 $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$。

♣

性质 公理化概率具有以下性质。

1. $P(\varnothing) = 0$。
2. $P(\overline{A}) = 1 - P(A)$。
3. 有限可加性：设 $A_1, A_2, \cdots, A_n$ 两两互不相容，则 $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$。
4. 下连续性：若 $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$，则 $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to +\infty} P(A_n)$。
5. 上连续性：若 $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$，则 $P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \to +\infty} P(A_n)$。

注 若 A 是不可能事件，则 $P(A) = 0$；反之，若 $P(A) = 0$，A 不一定是不可能事件。

由上述公理化概率的性质，可推得以下计算公式。

1. 加法公式：$P(A + B) = P(A) + P(B) - P(AB)$。

笔记 将加法公式推广到多事件情况可得容斥原理，即

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{1 \leqslant i \leqslant n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

通常取 $n = 3$ 进行计算，即

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

2. 减法公式：$P(A - B) = P(A) - P(AB)$。

笔记 若 $B \subset A$，则 $P(A - B) = P(A) - P(B)$，且 $P(A) \geqslant P(B)$。

例题 1.9 设有事件 A 和 B，$P(A) = 0.4$，$P(B) = 0.3$，$P(A + B) = 0.6$，求 $P(A\overline{B})$。

解 $P(AB) = P(A) + P(B) - P(A + B) = 0.1$，则 $P(A\overline{B}) = P(A - B) = P(A) - P(AB) = 0.3$。

1.4 概率的基本运算法则

对任何概率问题的讨论都必须建立在 $P(\Omega)=1$ 的前提之上，而有时除了此总前提之外，还会出现附加前提，我们将这种问题称为**条件概率**问题。

定义 1.6 (条件概率)

设 Ω 为样本空间，A 和 B 是两个事件，且 $P(B)>0$，则称在 B 发生的条件下 A 发生的概率为 A 对 B 的条件概率，记为 $P(A\,|\,B)$。

笔记 在 $P(A)$ 中，样本空间为 Ω，而在 $P(A\,|\,B)$ 中，样本空间为 B，可记为 Ω_B。

性质 条件概率具有以下性质。

1. 非负性：$P(A\,|\,B)\geqslant 0$。
2. 规范性：$P(\Omega\,|\,B)=1$。
3. 可列可加性：设 $A_1,A_2,\cdots,A_n,\cdots$ 两两不相容，则 $P\left(\bigcup\limits_{i=1}^{\infty}A_i\,|\,B\right)=\sum\limits_{i=1}^{\infty}P(A_i\,|\,B)$。

条件概率有样本点型与定义型两种计算公式：设 B 中的样本点数为 n_B，在 AB 中的样本点数为 n_{AB}，B 发生的概率为 $P(B)$，AB 发生的概率为 $P(AB)$，则有

$$P(A\,|\,B)=\frac{n_{AB}}{n_B}=\frac{P(AB)}{P(B)}$$

由条件概率的定义与计算公式，可得到如下三种**乘法公式**。

定理 1.1 (乘法公式)

1. 两个事件的乘法公式由条件概率定义式变化而来，即
$$\begin{aligned}P(AB)&=P(B)P(A\,|\,B)\quad (P(B)>0)\\&=P(A)P(B\,|\,A)\quad (P(A)>0)\end{aligned}$$
2. 含有三个事件 A,B,C 的乘法公式为 $P(ABC)=P(A)P(B\,|\,A)P(C\,|\,AB)$。
3. 乘法公式可以扩展到有限多个事件的情况，即
$$\begin{aligned}P(A_1A_2\cdots A_n)&=P(A_1)P(A_2\,|\,A_1)P(A_3\,|\,A_1A_2)\cdots P(A_n\,|\,A_1A_2\cdots A_{n-1})\\&=P(A_1)\cdot\prod_{i=2}^{n}P\left(A_i\,|\,\bigcap_{j=1}^{i-1}A_j\right)\end{aligned}$$

1.5 全概率公式

若将一个复杂事件分解为若干个较简单事件，并利用概率的加法公式和乘法公式，便可得到如下的**全概率公式**。

定理 1.2 (全概率公式)

设 $A_1,A_2,\cdots,A_n$ 是试验 E 的完备事件组，且 $P(A_i)>0$ ($i=1,2,\cdots,n$)，则对任意事件 B，其发生的概率为

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B\,|\,A_i)$$

该公式称为全概率公式。

例题 1.10 某工厂有四条生产线，相关信息如下表：

生产线编号	一	二	三	四
产量占比	15%	20%	30%	35%
不合格率	0.05	0.04	0.03	0.02

表 1.2: 某工厂生产线信息

现从该厂产品中随机抽取一件，求其为不合格品的概率。

解 设事件 $A_i=\{$抽到第 i 条生产线的产品$\}$，其中 $i=1,2,3,4$，事件 $B=\{$抽到的产品为不合格品$\}$，则

$$
\begin{aligned}
P(B) &= \sum_{i=1}^{4} P(A_i)P(\,B\mid A_i\,) \\
&= 0.05\times 0.15+0.04\times 0.2+0.03\times 0.3+0.02\times 0.35 \\
&= 0.0315
\end{aligned}
$$

1.6 贝叶斯公式

将条件概率公式、概率的乘法公式与全概率公式相结合，可得如下所示的**贝叶斯公式**。

定理 1.3 (贝叶斯公式)

设 $A_1,A_2,\cdots,A_n$ 是完备事件组，B 是任意事件，且 $P(A_i)>0$，$P(B)>0$，则

$$
P(\,A_k\mid B\,)=\frac{P(A_k)P(\,B\mid A_k\,)}{\sum\limits_{i=1}^{n} P(A_i)P(\,B\mid A_i\,)},\qquad k=1,2,\cdots,n
$$

称上述公式为贝叶斯公式。

笔记 一般地，我们将形如 $P(A_i)$ 的概率称为先验概率，而将 $P(\,A_i\mid B\,)$ 称为后验概率。

1.7 事件的独立性

在一般情况下，无条件概率并不等于条件概率，但在事件**相互独立**的前提下二者可以相等，下面将给出其定义。

定义 1.7 (事件的独立性)

若事件 A 的概率不受事件 B 发生与否的影响，即 $P(\,A\mid B\,)=P(A)$，称为 A 对 B 独立。

性质 独立事件具有以下性质。

1. 若事件 A 对事件 B 独立，则事件 B 也对事件 A 独立。
2. 若 A 与 B 独立，则 A 与 $\overline{B}$、$\overline{A}$ 与 B、$\overline{A}$ 与 $\overline{B}$ 也独立。
3. 若 $P(A)=0$ 或 $P(A)=1$，则 A 与任意事件独立。

设有事件 A 和事件 B，且 $P(A)>0$，$P(B)>0$，则 A 和 B 相互独立的充分条件为 $P(AB)=P(A)P(B)$，且当 $P(A)=0$ 或 $P(B)=0$ 时，该公式仍成立。由此我们可以对事件的独立性进行再定义：**若事件 A 和事件 B 满足 $P(AB)=P(A)P(B)$，则称 A 与 B 独立**。若不满足上述定义，称 A 与 B **不独立**或**相依**。

同理，我们可以将两个事件相互独立的概念推广到多个事件的独立性。

定义 1.8 (多个事件的独立性)

设有事件 $A_1, A_2, \cdots, A_n$，对任意的 $1 \leqslant i < j < k < \cdots \leqslant n$，若以下 $2^n - n - 1$ 个等式

$$\begin{cases} P(A_iA_j) = P(A_i)P(A_j) \\ P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k) \\ \cdots \\ P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n) \end{cases}$$

均成立，则称 n 个事件 $A_1, A_2, \cdots, A_n$ 相互独立。

与上面给出的两个事件相互独立的性质类似，n 个相互独立的事件中任意 $1 < k \leqslant n$ 个事件仍然是相互独立的，且若将 n 个相互独立的事件中任意 $1 \leqslant m \leqslant n$ 个换成其对立事件，所得到的 n 个事件仍然是相互独立的。

对于多个事件的独立性问题，我们常取 $n = 3$ 的情况进行讨论：若事件 A, B, C 满足

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{cases}$$

则称 A, B, C **两两独立**；若还有

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

则称 A, B, C 相互独立。

笔记 由事件 A, B, C 两两独立不能推得 A, B, C 相互独立；若 $P(A) > 0$，$P(B) > 0$，则 A 与 B 独立和 A 与 B 互不相容不能同时发生。

由上述多个事件的相互独立问题，我们给出如下**独立事件序列**的定义。

定义 1.9 (独立事件序列)

设有事件序列 $A_1, A_2, \cdots, A_n, \cdots$，若其中任意有限个事件都相互独立，则称该事件序列是相互独立的，并称其为独立事件序列。

当 $A_1, A_2, \cdots, A_n$ 是 n 个相互独立的事件，有

1. $A_1, A_2, \cdots, A_n$ 至少有一个发生的概率为
$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$
2. $A_1, A_2, \cdots, A_n$ 都发生的概率为
$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

例题 1.11 $P(A+B) = 0.9$，$P(A) = 0.4$，求分别在 A 与 B 互不相容和 A 与 B 独立的情况下 $P(B)$ 的值。

解 由题意，可得

1. 当 A 与 B 互不相容时，$AB = \varnothing$，故 $P(AB) = 0$，由 $P(A+B) = P(A) + P(B) - P(AB)$ 有 $P(B) = 0.5$；
2. 当 A 与 B 独立时，有 $P(AB) = P(A)P(B)$，代入 $P(A+B) = P(A) + P(B) - P(A)P(B)$，有 $P(B) = \dfrac{5}{6}$。

例题 1.12 现有一串密码，在某情报机构中每人各自独立地破译出该密码的概率为 0.6，若想以不低于 99% 的概率成功破译，求至少需要的人数。

解 设至少需要 n 个人，事件 $A_i = \{$第 i 个人破译出密码$\}$，其中 $0 \leqslant i \leqslant n$，$B = \{$密码被成功破译$\}$，易得

$B = \bigcup_{i=1}^{n} A_i$，则有

$$\begin{aligned} P(B) &= 1 - P\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \\ &= 1 - \prod_{i=1}^{n} P(\overline{A_i}) \\ &= 1 - 0.4^n \end{aligned}$$

由题意，有 $1 - 0.4^n \geqslant 0.99$，解得 $n \geqslant \log_{0.4} 0.01 \approx 5.026$，即至少需要 6 个人。

1.8 伯努利概型

下面先给出伯努利概型中的一些基本概念。

1. 独立试验序列：若试验 $E_1, E_2, \ldots, E_n$ 的试验结果彼此相互独立，则称其为独立试验序列。
2. n 重独立试验：将一个试验 E 重复进行 n 次，且每次试验结果相互独立，记为 E^n。
3. 伯努利试验：试验结果只有两种的试验。
4. n 重伯努利试验：将伯努利试验重复进行 n 次，且每次试验结果相互独立。

由于伯努利试验只有两种结果，故可以利用二项公式对其结果概率进行计算。

定理 1.4 (伯努利概型的计算公式)

设事件 A 的概率为 p，其中 $0 < p < 1$，则在 n 重伯努利试验中 A 恰好发生 k 次的概率为

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

上述公式也成为二项概率公式。

笔记 小概率事件的实际不可能原理：概率极小的事件在一次试验中几乎不可能发生。

第 1 章 练习

1. 现有 a 个白球和 b 个黑球，从中接连取出 m 个球，其中 $1 \leqslant m \leqslant a+b$，求取出的第 m 个球是白球的概率。

解 在此提供四种解法，记 $A_m = \{$取出的第 m 个球是白球$\}$。

(a). 我们可以只对第 m 个位置进行讨论，则共有 $a+b$ 种取出方法，其中满足题意的有 a 种方法，故概率为 $\dfrac{a}{a+b}$。

(b). 可以先将 $a+b$ 个球全部取出并排列，指定第 m 位为白球，剩下的 $a+b-1$ 个球的位置我们并不关心，进行全排列即可，故概率为

$$\frac{a \cdot (a+b-1)!}{(a+b)!} = \frac{a}{a+b}$$

(c). 将同色的球视为相同元素，则 $a+b$ 个球的取出排列由 a 个白球的位置决定，由古典概型有总情况数为 C_{a+b}^{a}，满足题意情况数为 C_{a+b-1}^{a-1}，故可求得概率为

$$\frac{\mathrm{C}_{a+b-1}^{a-1}}{\mathrm{C}_{a+b}^{a}} = \frac{a}{a+b}$$

(d). 只对前 m 个球的位置进行讨论，样本点总数为从 $a+b$ 个球中取出 m 个的排列数 A_{a+b}^{m}，A_m 相当于在第 m 个位置放白球，共 a 种，每种放法又对应前面 $m-1$ 个位置由余下 $a+b-1$ 个球取出 $m-1$

个球进行排列，由此可得 A_m 包含的样本数为 $a\mathrm{A}_{a+b-1}^{m-1}$，故概率为

$$\frac{a\mathrm{A}_{a+b-1}^{m-1}}{\mathrm{A}_{a+b}^{m}} = \frac{a}{a+b}$$

2. 设有事件 A,B,C，$P(A)=P(B)=P(C)=\frac{1}{4}$，$P(AB)=0$，$P(BC)=P(AC)=\frac{1}{16}$，求 A,B,C 至少发生一个的概率和 A,B,C 都不发生的概率。

解 由于 $ABC \subset AB$，故 $0 \leqslant P(ABC) \leqslant P(AB)=0$，即 $P(ABC)=0$，则有

(a). A,B,C 至少发生一个的概率 $P(A+B+C)=3\times\frac{1}{4}-2\times\frac{1}{16}=\frac{5}{8}$；

(b). A,B,C 都不发生的概率 $P(\overline{A}\cdot\overline{B}\cdot\overline{C})=1-P(A+B+C)=\frac{3}{8}$。

3. 设有 20 件奖品，其中一等奖 6 件，二等奖 10 件，三等奖 4 件，从中任取 3 件，求至少有两件奖品是同等级的概率。

解 设事件 $A=\{$至少有两件奖品是同等级$\}$，则 $\overline{A}=\{$三件奖品的等级两两不相同$\}$，故有

$$P(A)=1-P(\overline{A})=1-\frac{\mathrm{C}_6^1\mathrm{C}_{10}^1\mathrm{C}_4^1}{\mathrm{C}_{20}^3}=\frac{15}{19}$$

4. 求在一个平年里，在 n 个人中至少有两人生日相同的概率。

解 设事件 $A=\{n$ 个人中至少有两人生日相同$\}$，则 $\overline{A}=\{n$ 个人的生日两两不相同$\}$，故有

$$P(A)=1-P(\overline{A})=1-\frac{\overbrace{365\times364\times363\times\cdots}^{n\text{ 个数}}}{365^n}$$

5. 现有 100 件产品，次品率为 10%，从中不放回地抽取检查，求第 3 次才取到合格品的概率。

解 设事件 $A_i=\{$第 i 次抽取的产品为合格品$\}$，其中 $i=1,2,3$，则第 3 次才取到合格品的概率为

$$\begin{aligned} P(\overline{A_1}\cdot\overline{A_2}\cdot A_3) &= P(\overline{A_1})P(\,\overline{A_2}\,|\,\overline{A_1}\,)P(\,A_3\,|\,\overline{A_1}\cdot\overline{A_2}\,) \\ &= \frac{10}{100}\times\frac{9}{99}\times\frac{90}{98}=0.00835 \end{aligned}$$

6. 设箱子中有 a 个红球和 b 个黑球，每次从中随机抽取一个并放回，且再放入 c 个颜色与之相同的球，求连续三次都抽到红球的概率。

解 设事件 $A_i=\{$第 i 次抽到红球$\}$，其中 $i=1,2,3$，则连续三次都抽到红球的概率为

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)P(\,A_2\,|\,A_1\,)P(\,A_3\,|\,A_1A_2\,) \\ &= \frac{a}{a+b}\cdot\frac{a+c}{a+b+c}\cdot\frac{a+2c}{a+b+2c} \end{aligned}$$

7. 有 10 台收音机，其中有 3 台是次品，现售出 2 台，求在其中任取一台是正品的概率。

解 设事件 $B=\{$第三次抽到次品$\}$，$A_i=\{$在售出的两台中有 i 台正品$\}$，其中 $i=0,1,2$，则

$$\begin{aligned} P(B) &= \sum_{i=1}^{3}P(A_i)P(\,B\,|\,A_i\,) \\ &= \frac{\mathrm{C}_3^2}{\mathrm{C}_{10}^2}\times\frac{7}{8}+\frac{\mathrm{C}_3^1\mathrm{C}_7^1}{\mathrm{C}_{10}^2}\times\frac{6}{8}+\frac{\mathrm{C}_7^2}{\mathrm{C}_{10}^2}\times\frac{5}{8} \\ &= 0.7 \end{aligned}$$

8. 在 10 件产品中，次品的数量在 0 件至 2 件之间。在其中随机抽出一件，若检验出该产品为次品，则这批产品不通过检验，其中正品被检验成次品的概率为 0.02，次品被检验成正品的概率为 0.05，求这批产品能通

过检验的概率。

解 设 $B=\{$这批产品通过检验$\}$，$A_i=\{$这批产品中有 i 件产品$\}$（$i=0,1,2$），$B_1=\{$所抽的产品是正品$\}$，易得 $P(A_0)=P(A_1)=P(A_2)=\dfrac{1}{3}$，$P(B\,|\,A_0)=1$，$P(B\,|\,A_1)=\dfrac{9}{10}$，$P(B\,|\,A_2)=\dfrac{8}{10}$，故有

$$P(B_1)=\sum_{i=0}^{2}P(A_i)P(B_1\,|\,A_i)=0.9$$

这批产品通过验证有两种情况，第一种是抽到了正品并且能通过检验，第二种是虽然抽到了次品但是能通过检验，故有

$$P(B)=P(B_1)P(B\,|\,B_1)+P(\overline{B_1})P(B\,|\,\overline{B_1})=0.887$$

9. 某病发病率为 0.0004，但由于检验技术限制，该病患者被检测成健康人的概率为 1%，而健康人被检验成患者的概率为 0.1%。现有一人被检验成患者，求其确实是患者的概率。

解 设事件 $A=\{$该人为患者$\}$, $B=\{$该人检验结果为患者$\}$, 则由题意可得 $P(A)=0.0004$, $P(\overline{A})=0.0096$, $P(B\,|\,A)=0.99$，$P(B\,|\,\overline{A})=0.001$，则

$$\begin{aligned}P(B)&=P(A)P(B\,|\,A)+P(\overline{A})P(B\,|\,\overline{A})\\&=0.0004\times 0.99+0.9996\times 0.001\\&=0.0013956\end{aligned}$$

利用贝叶斯公式，可得该人确实是患者的概率为

$$P(A\,|\,B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B\,|\,A)}{P(B)}=\frac{0.0004\times 0.99}{0.0013956}=0.284$$

10. 已知事件 A 和 B，$0<P(A)<1$，$0<P(B)<1$，且 $P(A\,|\,B)+P(\overline{A}\,|\,\overline{B})=1$，求证 A 与 B 独立。

证明 由于有恒等式 $P(A\,|\,\overline{B})+P(\overline{A}\,|\,\overline{B})=1$ 成立，故有 $P(A\,|\,B)=P(A\,|\,\overline{B})$ 成立；
因为有

$$P(A\,|\,B)=\frac{P(AB)}{P(B)}\tag{1.1}$$

和

$$P(A\,|\,\overline{B})=\frac{P(A\overline{B})}{P(\overline{B})}=\frac{P(A-AB)}{1-P(B)}=\frac{P(A)-P(AB)}{1-P(B)}\tag{1.2}$$

联立 (1.1) 式和 (1.2) 式并交叉相乘有

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

即有 $P(AB)=P(A)P(B)$ 成立，故 A 与 B 独立。

第 2 章 随机变量及其分布

2.1 随机变量

我们可以用一个数 X 表示一个试验中的各类结果，由于 X 的取值随机，故称 X 为**随机变量**。

定义 2.1 (随机变量)

设试验 E 的样本空间为 Ω，若对每一个样本点 $\omega \in \Omega$，都有唯一实数 $X(\omega)$ 与之对应，则称 $X = X(\omega)$ 为样本空间 Ω 上的随机变量。♣

笔记 通常，随机变量用 $X, Y, Z, \xi, \eta, \zeta$ 等表示。

我们可以利用随机变量对第一章中的相关概念进行重新表示：事件为样本点的集合，记作 $\{\omega \mid X(\omega) = a\}$ 或 $\{X = a\}$；相应地，事件的概率记作 $P\{X = a\}$ 或 $P(X = a)$。

在实际生活中，随机变量主要有两大类，一类是**离散型随机变量**，其取值结果为有限个或无限可列个的随机变量，而另一类是**非离散型随机变量**，其中以**连续型随机变量**较为常见。

2.2 离散型随机变量及其概率分布

由定义可知，离散型随机变量的每一个离散取值都有一个对应的概率，在此给出**概率分布**的定义。

定义 2.2 (概率分布)

设离散型随机变量 X 所有可能取值为 x_k（$k = 1, 2, \cdots$），而 X 取值 x_k 的概率为 p_k，即

$$P(X = x_k) = p_k, \qquad k = 1, 2, \cdots$$

称上式为离散型随机变量 X 的概率分布或分布律，也称概率函数。♣

若将离散型随机变量 X 的所有可能取值及相应概率列成如表 (2.1) 所示的表，则称其为 X 的**概率分布表**，或称为**分布列**。我们也可以用如图 (2.1) 所示的**概率函数图**来表示随机变量的分布律，且对于离散型随机变量需要满足

$$\begin{cases} p_k \geqslant 0, \qquad k = 1, 2, \ldots & (2.1) \\ \sum_k p_k = 1 & (2.2) \end{cases}$$

反过来，满足 (2.1) 式和 (2.2) 式的数 p_k 也一定可以作为某离散型随机变量的概率分布。

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

表 2.1: 离散型随机变量 X 的分布列

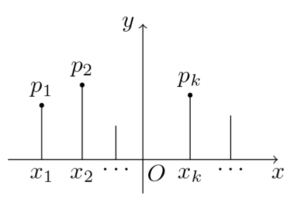

图 2.1: 随机变量的概率函数图

例题 2.1 在一个箱子内有 5 个黑球和 3 个白球，每次从中随机抽取一个并且不放回，直到取到黑球为止。设 X 为取到白球的数目，试求 $P(-1<X<0)$、$P(1<X<3)$ 与 $P(X\leqslant 3)$ 的值。

解 由题意，X 的所有可能取值为 $0,1,2,3$，则有

$$
\begin{aligned}
P(X=0)&=\frac{5}{8}\\
P(X=1)&=\frac{3}{8}\times\frac{5}{7}=\frac{15}{56}\\
P(X=2)&=\frac{3}{8}\times\frac{2}{7}\times\frac{5}{6}=\frac{5}{56}\\
P(X=3)&=1-P(X=0)-P(X=1)-P(X=2)=\frac{1}{56}
\end{aligned}
$$

X 的分布列如表 2.2 所示：

X	0	1	2	3
P	$\frac{5}{8}$	$\frac{15}{56}$	$\frac{5}{56}$	$\frac{1}{56}$

表 2.2: 例题 2.1 X 的分布列

由分布列可知，$P(-1<X<0)=0$，$P(1<X<3)=P(X=2)=\frac{15}{56}$，$P(X\leqslant 3)=1$。

2.3 连续型随机变量及其概率密度函数

在初高中时期我们已经学习过频率密度直方图（即频率分布直方图），其中每个矩形的面积等于该组对应的频率，所有矩形的面积总和为 1，且介于 $x=a$ 与 $x=b$ 之间的面积近似于 $(a,b]$ 的频率。若将频率密度直方图中的横轴区间等分成无数的小区间并作出上方的矩形，则此时连接每个矩形的上中点可近似得到一条光滑的曲线，该曲线称为**概率密度函数**。

定义 2.3 (概率密度函数的严格定义)

设变量 X 的所能取值为某区间上的所有实数，若存在非负可积函数 $f(x)$，使得对任意的实数 $a\leqslant b$，都有

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

则 X 称为连续型随机变量，$f(x)$ 称为 X 的概率分布密度函数，简称概率密度函数或密度函数，可记为 $X\sim f(x)$。

♣

性质 概率密度函数具有以下性质。

1. $f(x)\geqslant 0$ 且 $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$。
2. $P(a<X<b)=P(a\leqslant X<b)=P(a<X\leqslant b)=P(a\leqslant X\leqslant b)$。
3. 连续型随机变量取个别值的概率为 0。

在概率密度函数定义域中取一点 x_0，其函数值 $f(x_0)$ 为 X 取 x_0 附近值时的概率。

证明 对 x_0 取一个极小的增量 Δx，则有

$$\lim_{\Delta x\to 0}\frac{P(x_0<X<x_0+\Delta x)}{\Delta x}=\lim_{\Delta x\to 0}\frac{\int_{x_0}^{x_0+\Delta x}f(x)\mathrm{d}x}{\Delta x}=f(x)$$

即 $P(x_0<X<x_0+\Delta x)\approx f(x_0)\cdot\Delta x$。

笔记 可以将概率密度函数类比为不均匀物体的密度函数，在某一体积区间上对其求积分即为该物体的部分质量。

例题 2.2 设一连续型随机变量的概率密度函数为

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

试求 k 以及 $P(X\leqslant 2)$ 的值。

解 由题意有 $\displaystyle\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=\int_0^2(kx+1)\mathrm{d}x=1$，解得 $k=-\dfrac{1}{2}$，故有

$$P(X\leqslant 2)=\int_{-\infty}^{2}f(x)\mathrm{d}x=\int_0^2\left(-\frac{1}{2}x+1\right)\mathrm{d}x=1$$

2.4 分布函数

无论是离散型随机变量还是连续型随机变量，我们都希望能够求出其在给定区间内取值时的概率，由此便有了**分布函数**的概念。

定义 2.4 (分布函数)

设有随机变量 X，x 是任意一实数，则函数

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

称为随机变量 X 的概率分布函数，简称分布函数。

笔记 分布函数 $F(x)$ 的含义为随机变量 X 的取值不超过 x 的概率。

性质 分布函数具有以下性质。

1. $0\leqslant F(x)\leqslant 1$，其中 $x\in(-\infty,+\infty)$。
2. $F(x)$ 为非减函数，且 $\lim\limits_{x\to+\infty}F(x)=1$，$\lim\limits_{x\to-\infty}F(x)=0$。
3. $F(x)$ 至多有可列个间断点，且对于离散型随机变量，$F(x)$ 是右连续的，而对于连续型随机变量，$F(x)$ 是连续的。

借助于分布函数，我们可以求出随机变量在一定区间内取值时的概率。

1. $P(X>a)=1-F(a)$。
2. $P(a<X\leqslant b)=F(b)-F(a)$。
3. $P(X=a)=F(a)-F(a-0)$。
4. $P(a\leqslant X\leqslant b)=F(b)-F(a-0)$。

注 上述公式中的 $a-0$ 指的是从左端逼近 a，其值不一定等于 a。

对于离散型随机变量 X，其分布函数的间断点 x_k 即为 X 的取值，且 $P(X=x_k)=F(x_k)-F(x_k-0)$；对于连续型随机变量 X，有 $F(x)=P(X\leqslant x)=\displaystyle\int_{-\infty}^{x}f(t)\mathrm{d}t$，即 $F'(x)=f(x)$，其中 $f(x)$ 为 X 的概率密度函数。

例题 2.3 已知某随机变量的分布函数为

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

其中 $\lambda>0$，求 a 的值。

解 根据分布函数的性质可知 $\lim\limits_{x\to+\infty}F(x)=1$，故 $\lim\limits_{x\to+\infty}(a-\mathrm{e}^{-\lambda x})=1$，即 $a=1$。

2.5 概率分布的类型

0-1 分布是最简单而又最常见的概率分布，下面给出其定义。

定义 2.5 (0-1 分布)

若在一次试验中，离散型随机变量 X 的分布列形式如表 2.3 所示，则称这样的 X 服从 0-1 分布，也称两点分布。♣

X	0	1
P	$1-p$	p

表 2.3: 服从 0-1 分布的 X 的分布列

若将两点分布推广到更一般的情况，则可得到**二项分布**。

定义 2.6 (二项分布)

在 n 重伯努利试验中，若事件 A 出现的次数记为 X，则随机变量 X 的可能的取值是 $0,1,2,\cdots,n$，相应概率分布为

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

其中 $p=P(A)$，$0<p<1$，称 X 服从参数为 n,p 的二项分布，记作 $X\sim B(n,p)$。♣

对于二项分布 $B(n,p)$，其**最可能出现次数**（或称最可能值，即二项分布达到最大值处）为

1. 若 $(n+1)p$ 不为整数，则最可能出现次数为 $\lfloor(n+1)p\rfloor$；
2. 若 $(n+1)p$ 为整数，则最可能出现次数为 $(n+1)p$ 和 $(n+1)p-1$。

例题 2.4 设某款火灾烟雾报警器在遇到险情时报警的概率为 0.8，现要保证此情况下报警的概率大于 99%，求至少需要报警器的台数。

解 设 X 为发生火灾时报警的报警器台数，n 为报警器的总台数，显然 $X\sim B(n,0.8)$，则有

$$P(X\geqslant 1)=1-P(X=0)=1-0.2^n\geqslant 0.99$$

解得 $n\geqslant\dfrac{\ln 0.01}{\ln 0.2}\approx 3$，即至少需要 3 台报警器。

当二项分布中的 n 较大时，计算随机变量的取值概率将会非常棘手，此时可采用**泊松定理**来近似计算。

定理 2.1 (泊松定理)

设随机变量 $X_n\sim B(n,p_n)$（$n=1,2,\cdots$），若 $\lim\limits_{n\to\infty}np_n=\lambda$，则有

$$\lim_{n\to\infty}P(X_n=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

♡

由泊松定理可以引出下面**泊松分布**的定义。

定义 2.7 (泊松分布)

若随机变量 X 的概率分布为

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

其中 $\lambda>0$ 为常数，则称 X 服从以 λ 为为参数的泊松分布，记作 $X\sim P(\lambda)$。♣

在二项分布中，若遇到 $n\geqslant 100$, $np\leqslant 10$ 或 $n>10$, $p<0.1$ 等情况，即当 n 比较大、p 比较小、np 比较适中时，则可以泊松分布来近似代替二项分布进行计算。

例题 2.5 有一电话交换台，每分钟接收到用户呼叫的次数 $X \sim P(3)$，写出 X 的概率函数，并求出该电话交换台一分钟内接收到的呼叫次数不超过 5 次的概率。

解 由于 $X \sim P(3)$，则有 $\lambda = 3$，故 X 的概率函数为

$$P(X=k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \frac{3^k}{\mathrm{e}^3 k!}, \qquad k = 0, 1, \cdots$$

经查表，可得该电话交换台一分钟内接收到的呼叫次数不超过 5 次的概率为

$$P(X \leqslant 5) = \sum_{k=0}^{5} P(X=k) = 0.916$$

例题 2.6 一证券营业部拥有 1000 个账户，其中每个账户都存有 10 万元，且各户前来提出 20% 本金的概率为 0.006，那么至少需要准备多少现金能有 95% 以上的概率以满足前来提款账户的要求？

解 设 X 为提款的账户数，需要准备的现金为 x 万元，则可得 $X \sim B(1000, 0.006)$，且需要满足 $P(2X \leqslant x) \geqslant 0.95$，即 $P\left(X \leqslant \frac{x}{2}\right) \geqslant 0.95$，由于 $np = 6$，则可令 $\lambda = np = 6$，采用泊松分布近似代替二项分布，则有

$$\sum_{k=0}^{x/2} \frac{6^k}{k!}\mathrm{e}^{-6} \geqslant 0.95$$

查表可解得 $\frac{x}{2} \geqslant 10$，即 $x \geqslant 20$，故至少需要准备 20 万元现金。

例题 2.7 某传染病的发病率为 $\frac{1}{1000}$，在一单位内有 5000 人，求至少两人得病的概率。

解 设 X 为得病人数，$n = 5000$，$p = \frac{1}{1000}$，则有 $X \sim B\left(5000, \frac{1}{1000}\right)$，考虑到 $np = 5$，选择采用泊松分布近似代替二项分布，设 $\lambda = np = 5$，则至少两人得病的概率为

$$\begin{aligned} P(X \geqslant 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - 0.006738 - 0.03369 \\ &= 0.959572 \end{aligned}$$

下面我们将给出**几何分布**与**超几何分布**的定义。

定义 2.8 (几何分布)

设事件 A 发生的概率为 p，且 A 于第 k 次试验中首次发生，此时设 X 为 A 发生时已经进行的试验次数，则满足

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

称 X 服从几何分布，记为 $X \sim G(p)$。

♣

定义 2.9 (超几何分布)

设有 N 个元素，根据某种规则将其划分出一指定种类，其中含有 M 个元素。从所有元素中取出 n 个，记 X 为属于指定种类的元素个数，则有

$$P(X=k) = \frac{\mathrm{C}_M^k \mathrm{C}_{N-M}^{n-k}}{\mathrm{C}_N^n}, \qquad k = 0, 1, 2, \cdots, \min\{n, M\}$$

称 X 服从参数为 n, M, N 的超几何分布，记作 $X \sim H(n, M, N)$。

♣

注 超几何分布分布的模型是不放回抽样。

在超几何分布中，记指定种类元素占比 $p = \frac{M}{N}$，当 N 较大而 n 相对其较小时，超几何分布可近似看为二项

分布，即

$$P(X=k)=\frac{\mathrm{C}_M^k\mathrm{C}_{N-M}^{n-k}}{\mathrm{C}_N^n}\approx \mathrm{C}_n^k p^k(1-p)^{n-k}$$

例题 2.8 有 20 名学生，其中 5 名是女生，现要挑选 4 名学生去参赛，设 X 为被挑选的学生中女生的人数，求 X 的概率函数。

解 显然 $X\sim H(4,5,20)$，则有

$$P(X=k)=\frac{\mathrm{C}_5^k\mathrm{C}_{15}^{4-k}}{\mathrm{C}_{20}^4},\qquad k=0,1,2,3,4$$

上面介绍了几种离散型随机变量的概率分布，接下来将着重介绍连续型随机变量常见的概率分布。

定义 2.10 (均匀分布)

若连续型随机变量 X 的概率密度函数为

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

则称 X 在区间 $[a,b]$ 上服从均匀分布，记作 $X\sim U[a,b]$。♣

若连续型随机变量 $X\sim U[a,b]$，则其分布函数为

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

且若设 $[c,d]\subset[a,b]$，则有 $P(c\leqslant X\leqslant d)=\dfrac{d-c}{b-a}$。

定义 2.11 (指数分布)

若连续型随机变量 X 的概率密度函数为

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

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

若连续型随机变量 $X\sim E(\lambda)$，则其分布函数为

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

指数分布有一项重要的性质，称为“**无记忆性**”。

定理 2.2 (指数分布的“无记忆性”)

设随机变量 $X\sim E(\lambda)$，则对于任意 $s>0$ 和 $t>0$，有

$$P(X>s+t\,|\,X>s)=P(X>t)$$

♡

证明 直接利用条件概率定义式，可得

$$
\begin{aligned}
P(\,X>s+t\,|\,X>s\,) &= \frac{P((X>s+t)\cap(X>s))}{P(X>s)} \\
&= \frac{P(X>s+t)}{P(X>s)} \\
&= \frac{\displaystyle\int_{s+t}^{+\infty}\lambda \mathrm{e}^{-\lambda x}\mathrm{d}x}{\displaystyle\int_{s}^{+\infty}\lambda \mathrm{e}^{-\lambda x}\mathrm{d}x} \\
&= \frac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda s}} \\
&= \mathrm{e}^{-\lambda t} \\
&= \int_{t}^{+\infty}\lambda \mathrm{e}^{-\lambda x}\mathrm{d}x \\
&= P(X>t)
\end{aligned}
$$

正态分布又称为高斯分布或高斯–拉普拉斯分布，是自然界中最常见的一种分布，于 1733 年由棣莫弗提出。

定义 2.12 (正态分布)

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

其中 μ,σ 为常数，且 $\sigma>0$，则称随机变量 X 服从参数为 μ,σ 的正态分布，记作 $X\sim N(\mu,\sigma^2)$，并称 X 为正态变量。

若连续型随机变量 $X\sim N(\mu,\sigma^2)$，则其分布函数为

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

性质 正态分布具有以下性质。

1. $f(x)$ 是以 $x=\mu$ 为对称轴的钟形函数，且当 $x=\mu$ 时 $f(x)$ 取到最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$。
2. $f(x)$ 以 x 轴为渐近线，并以 $x=\mu\pm\sigma$ 为拐点。
3. μ 称为位置参数，σ 称为离散参数：保持 σ 不变，当 μ 减小或增大时，正态曲线向左或向右移动；保持 μ 不变，当 σ 减小或增大时，正态曲线变得更陡峭或更平缓。

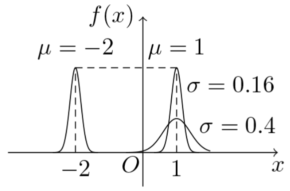

图 2.2: 参数 μ, σ 对正态曲线的影响

例题 2.9 求证：$\displaystyle\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x=\sqrt{\pi}$。

证明 由正态分布的概率密度函数有

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

上式变化可得

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

故 $\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x=\sqrt{\pi}$。

若 $\mu=0$，$\sigma=1$，则这样的正态分布称为**标准正态分布**，记作 $X\sim N(0,1)$，并将相应的概率密度函数与分布函数写为 $\varphi(x)$ 和 $\Phi(x)$，即

$$\begin{aligned}\varphi(x)&=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, && -\infty<x<+\infty\\ \Phi(x)&=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t, && -\infty<x<+\infty\end{aligned}$$

性质 标准正态分布具有以下性质。

1. $\varphi(x)$ 为偶函数。
2. $\Phi(-x)=1-\Phi(x)$。

正态分布与标准正态分布具有如下联系：若 $X\sim N(\mu,\sigma^2)$，则 $\dfrac{X-\mu}{\sigma}\sim N(0,1)$，且有

$$f(x)=\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

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

定理 2.3 (3σ 原则)

$$\begin{cases}P(\mu-\sigma<X<\mu+\sigma)=0.6826\\P(\mu-2\sigma<X<\mu+2\sigma)=0.9546\\P(\mu-3\sigma<X<\mu+3\sigma)=0.9974\end{cases}$$

由这组值可以看出，服从正态分布 $N(\mu,\sigma^2)$ 的随机变量 X 的取值几乎全部落在区间 $(\mu-3\sigma,\mu+3\sigma)$ 内，此特征称为正态分布的 3σ 原则。

为便于数理统计的应用，对于标准正态随机变量，我们引入**上 α 分位点**的定义。

定义 2.13 (上 α 分位点)

设 $X\sim N(0,1)$，若 u_α 满足

$$P(X>u_\alpha)=\alpha,\qquad 0<\alpha<1$$

则称点 u_α 为标准正态分布的上 α 分位点或上 α 分位数。

2.6 随机变量函数的分布

在实际问题中，我们可以由已知的离散型随机变量的分布来推导出与之相关变量的分布。

定理 2.4 (离散型随机变量函数的分布)

设离散型随机变量 X 的概率函数为

$$P(X=x_k)=p_k,\qquad k=1,2,\cdots$$

并设 Y 是 X 的函数，即 $Y=g(X)$，Y 的所有可能取值为 $g(x_1),g(x_2),\cdots,g(x_k),\cdots$，则

1. 若 $g(x_k)$ 两两互不相同，则 Y 的概率函数为

$$P(Y=g(x_k))=p_k,\qquad k=1,2,\cdots$$

2. 若存在 $g(x_{i1})=g(x_{i2})=\cdots=g(x_{im})=y_i$，其中 $i=1,2,\cdots$，则 Y 的概率函数为

$$P(Y=y_i)=\sum_{k=1}^{m}P(X=x_{ik})$$

♡

设 X 是连续型随机变量，若要求随机变量函数 $Y=g(X)$ 的分布，我们可以通过以下方式解决。

定理 2.5 (连续型随机变量函数的分布)

设连续型随机变量 X 的分布函数为 $F_X(x)$，即

$$F_X(x)=P(X\leqslant x)$$

可求得 Y 的分布函数为

$$F_Y(x)=P(Y\leqslant x)\leqslant P(g(X)\leqslant x)$$

解出不等式 $g(x)\leqslant x$ 并代入 $F_Y(x)$，对等式两边求导即可得 Y 的概率密度函数 $f_Y(x)$。

♡

若连续型随机变量 $X\sim U[a,b]$，则对于随机变量函数 $Y=kX+c$（$k\neq 0$），有

$$Y\sim U[\min\{ka+c,kb+c\},\max\{ka+c,kb+c\}]$$

若连续型随机变量 $X\sim N(\mu,\sigma^2)$，则对于随机变量函数 $Y=aX+b$（$a\neq 0$），有 $Y\sim N(a\mu+b,a^2\sigma^2)$ 成立，即

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

对于线性随机变量函数，有如下定理成立。

定理 2.6 (线性随机变量函数的概率密度函数公式)

设连续型随机变量 X 的概率密度函数为 $f_X(x)$，线性随机变量函数 $Y=kX+b$（$k\neq 0$），则 Y 的概率密度函数为

$$f_Y(x)=\frac{1}{|k|}f_X\left(\frac{x-b}{k}\right)$$

♡

上面都只是给出了较简单的随机变量函数的求解方法。若将定理 2.6 中的线性随机变量函数推广到非线性的情况，就得到下面更一般的公式。

定理 2.7 (随机变量函数的概率密度函数公式)

设连续型随机变量 X 具有概率密度函数 $f_X(x)$，又设函数 $y=g(x)$ 严格单调，其反函数 $g^{-1}(x)$ 有连续导数，则 $Y=g(X)$ 是连续型随机变量，其概率密度函数为

$$f_Y(x)=\begin{cases}f_X[g^{-1}(x)]\cdot|[g^{-1}(x)]'|, & \alpha<x<\beta\\ 0, & \text{其他}\end{cases}$$

其中 $\alpha=\min\{g(-\infty),g(+\infty)\}$，$\beta=\max\{g(-\infty),g(+\infty)\}$。

♡

柯西分布是概率论中重要的分布之一，下面给出标准柯西分布的定义。

定义 2.14 (标准柯西分布)

若随机变量 X 的概率密度函数为

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

则称 X 服从标准柯西分布。

第 2 章 练习

1. 已知离散型随机变量 X 的分布列如下所示，求 X 的分布函数。

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

解 根据 X 的取值对分布函数 $F(x)$ 进行分段讨论。

(a). 当 $x<-1$ 时，$F(x)=P(X\leqslant x)=0$；

(b). 当 $-1\leqslant x<2$ 时，$F(x)=P(X\leqslant x)=P(x=-1)=\frac{1}{2}$；

(c). 当 $2\leqslant x<3$ 时，$F(x)=P(X\leqslant x)=P(x=-1)+P(x=2)=\frac{5}{6}$；

(d). 当 $x\geqslant 3$ 时，$F(x)=1$。

故求得的分布函数为

$$F(x)=\begin{cases}0, & x<-1\\ \frac{1}{2}, & -1\leqslant x<2\\ \frac{5}{6}, & 2\leqslant x<3\\ 1, & x\geqslant 3\end{cases}$$

笔记 离散型随机变量的分布函数大致呈阶梯形，且各区间（除无穷大外）都为左闭右开。

2. 已知连续型随机变量 X 的概率密度函数为 $f(x)=\frac{1}{\pi(1+x^2)}$，求其分布函数。

解 由题意可知 $F(x)=\int_{-\infty}^{x}\frac{1}{\pi(1+t^2)}\mathrm{d}t=\frac{1}{\pi}\arctan t\Big|_{-\infty}^{x}=\frac{1}{\pi}\arctan x+\frac{1}{2}$。

3. 已知连续型随机变量 X 的概率密度函数为

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

求 X 的分布函数。

解 对概率密度函数进行分段讨论有

(a). 当 $x<0$ 时，$F(x)=\int_{-\infty}^{x}0\mathrm{d}t=0$；

(b). 当 $0\leqslant x<2$ 时，$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t=\int_{-\infty}^{0}f(t)\mathrm{d}t+\int_{0}^{x}\left(-\frac{1}{2}t+1\right)\mathrm{d}t=-\frac{1}{4}x^2+x$；

(c). 当 $x\geqslant 2$ 时，$F(x)=1$。

故求得的分布函数为

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

4. 现有分布函数

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

试求出概率密度函数 $f(x)$。

解 由分布函数的右连续性可得 $\lim\limits_{x\to 1^-}Ax^2=1$，解得 $A=1$，再由 $F'(x)=f(x)$ 可得概率密度函数为

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

5. 现有 10000 颗种子，发芽率为 99%，从中取出 200 颗，求至多一颗不发芽的概率。

解 设 X 为取出的种子中不发芽的颗数，由题意可知 X 服从超几何分布，记 $N=10000$，$M=100$，$n=200$，则至多一颗不发芽的概率为

$$\begin{aligned}P(X\leqslant 1)&=P(X=0)+P(X=1)\\&=\frac{\mathrm{C}_{100}^{0}\mathrm{C}_{9900}^{200}}{\mathrm{C}_{10000}^{200}}+\frac{\mathrm{C}_{100}^{1}\mathrm{C}_{9900}^{199}}{\mathrm{C}_{10000}^{200}}\end{aligned}$$

此数据计算较复杂，考虑采用二项分布近似代替超几何分布，设 $p=\dfrac{M}{N}=0.01$，则有

$$P(X\leqslant 1)=\mathrm{C}_{200}^{0}\cdot 0.01^0\cdot 0.99^{200}+\mathrm{C}_{200}^{1}\cdot 0.01\cdot 0.99^{199}$$

注意到 $n=200$，$p=0.01$，而 $np=2$，故采用泊松分布进一步代替二项分布：设 $\lambda=np=2$，则经查表可得

$$P(X\leqslant 1)=0.1353+0.2707=0.406$$

6. 公共汽车从 7:00 开始每 15 分钟发一次车，一乘客从 7:00 到 7:30 到达车站的时刻服从均匀分布，求其等车不超过 5 分钟以及等车超过 10 分钟的概率。

解 设该乘客到达车站时距离 7:00 已过去 X 分钟，则 $X\sim U[0,30]$，概率密度函数为

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

故其等车不超过 5 分钟的概率为

$$P(10\leqslant X\leqslant 15)+P(25\leqslant X\leqslant 30)=\frac{5}{30}+\frac{5}{30}=\frac{1}{3}$$

等车超过 10 分钟的概率为

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

7. 设某种机器元件的寿命为 X，其服从指数分布，概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{1000}\mathrm{e}^{-\frac{x}{1000}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

某台机器由 3 个该种元件构成，只要其中一个元件损坏其便不能正常工作，求该机器能工作 1000 小时的概率。

解 一个元件的寿命超过 1000 小时的概率为

$$P(X>1000)=\int_{1000}^{+\infty}\frac{1}{1000}\mathrm{e}^{-\frac{x}{1000}}\,\mathrm{d}x=\mathrm{e}^{-1}$$

故该机器能工作 1000 小时的概率为 $(\mathrm{e}^{-1})^3=\mathrm{e}^{-3}$。

8. 已知 $X\sim N(1,4)$，求 $P(0<X<1.6)$ 的表达式。

解 由题意，$\mu=1$，$\sigma=2$，则有

$$\begin{aligned}P(0<X<1.6)=F(1.6)-F(0)&=\Phi\left(\frac{1.6-1}{2}\right)-\Phi\left(\frac{0-1}{2}\right)\\&=\Phi(0.3)-\Phi(-0.5)\\&=\Phi(0.3)+\Phi(0.5)-1\end{aligned}$$

9. 某台生产机床所生产的零件长度满足 $X\sim N(50,1)$，其合格范围在 50 ± 1 内，试求由该机床生产的零件的合格率。

解 由题意，$\mu=50$，$\sigma=1$，则合格率为

$$\begin{aligned}P(49\leqslant X\leqslant 51)=F(51)-F(49)&=\Phi\left(\frac{51-50}{1}\right)-\Phi\left(\frac{49-50}{1}\right)\\&=\Phi(1)-\Phi(-1)\\&=2\Phi(1)-1\\&=0.6826\end{aligned}$$

10. 设 $X\sim N(0,1)$，$Y=X^2$，求 Y 的概率密度函数。

解 对函数区间进行分类讨论，有

(a). 当 $x<0$ 时，$F_Y(x)=P(Y\leqslant x)=P(X^2\leqslant x)=0$，故 $f_Y(x)=0$；

(b). 当 $x\geqslant 0$ 时，

$$\begin{aligned}F_Y(x)=P(Y\leqslant x)=P(X^2\leqslant x)&=P(-\sqrt{x}\leqslant X\leqslant \sqrt{x})\\&=\int_{-\sqrt{x}}^{\sqrt{x}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\\&=2\int_{0}^{\sqrt{x}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\end{aligned}$$

故 $f_Y(x)=F'_Y(x)=\dfrac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}\mathrm{e}^{-\frac{x}{2}}$。

综上，Y 的概率密度函数为

$$f_Y(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}\mathrm{e}^{-\frac{x}{2}}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

笔记 一般地，我们称概率密度函数为上式的连续型随机变量 X 服从自由度 $n = 1$ 的卡方分布，可记作 $X \sim \chi^2(1)$。

11. 设随机变量 X 的概率密度函数为

$$f_X(x) = \begin{cases} \dfrac{1}{x+1}, & 0 < x < \mathrm{e}-1 \\ 0, & \text{其他} \end{cases}$$

且有 $Y = \sqrt{X}$，求 Y 的概率密度函数。

解 对函数区间进行分类讨论，有

(a). 当 $x < 0$ 时，$F_Y(x) = P(Y \leqslant x) = P(\sqrt{X} < x) = 0$，故 $f_Y(x) = 0$；

(b). 当 $x \geqslant 0$ 时，

$$\begin{aligned} F_Y(x) = P(Y \leqslant x) = P(\sqrt{X} \leqslant x) = P(X \leqslant x^2) &= \int_{-\infty}^{x^2} f_X(t)\mathrm{d}t \\ &= \int_{-\infty}^{0} f_X(t)\mathrm{d}t + \int_{0}^{x^2} f_X(t)\mathrm{d}t \\ &= \begin{cases} \displaystyle\int_0^{x^2} \frac{1}{t+1}\mathrm{d}t, & 0 \leqslant x^2 < \mathrm{e}-1 \\ 1, & x^2 \geqslant \mathrm{e}-1 \end{cases} \\ &= \begin{cases} \ln(1+x^2), & 0 \leqslant x < \sqrt{\mathrm{e}-1} \\ 1, & x \geqslant \sqrt{\mathrm{e}-1} \end{cases} \end{aligned}$$

故

$$f_Y(x) = \begin{cases} \dfrac{2x}{1+x^2}, & 0 \leqslant x < \sqrt{\mathrm{e}-1} \\ 0, & x \geqslant \sqrt{\mathrm{e}-1} \end{cases}$$

综上，Y 的概率密度函数为

$$f_Y(x) = \begin{cases} \dfrac{2x}{1+x^2}, & 0 \leqslant x < \sqrt{\mathrm{e}-1} \\ 0, & \text{其他} \end{cases}$$

第 3 章 多维随机变量及其分布

我们通常会把与同一个随机现象相联系的多个随机变量看作一个整体，称为**多维随机变量**或**随机向量**，如由 n 个随机变量 $X_1, X_2, \cdots, X_n$ 构成的 n 维随机变量 $(X_1, X_2, \cdots, X_n)$，其中第 i 个随机变量 X_i 称为 n 维随机变量的第 i 个分量。

3.1 二维随机变量及其分布

在多维随机变量中，最常见的是二维随机变量，本章将针对其进行重点介绍。

定义 3.1 (二维随机变量分布函数)

设 (X, Y) 为二维随机变量，对任意实数 x, y，称二元函数

$$F(x, y) = P(X \leqslant x, Y \leqslant y)$$

为二维随机变量 (X, Y) 的联合分布函数，简称为 (X, Y) 的分布函数。♣

性质 二维随机变量分布函数具有以下性质。

1. $0 \leqslant F(x, y) \leqslant 1$。
2. $F(x, y)$ 是关于 x 或 y 的非减函数。
3. $F(-\infty, -\infty) = F(x, -\infty) = F(-\infty, y) = 0$，$F(+\infty, +\infty) = 1$。
4. $F(x, y)$ 关于 x 或 y 都为右连续，即 $F(x+0, y) = F(x, y)$，$F(x, y+0) = F(x, y)$。

设 $x_1 < x_2$，$y_1 < y_2$，则有

$$P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, x_1)$$

二维随机变量 (X, Y) 的一维分量 X 和 Y 都有各自的分布，故分别称 X 和 Y 的分布为 (X, Y) 关于 X 和 Y 的**边缘分布**。

定义 3.2 (边缘分布函数)

设二维随机变量 (X, Y) 的分布函数为 $F(X, Y)$，分别记关于 X 和 Y 的边缘分布函数为 $F_X(x)$ 和 $F_Y(y)$，则有

$$F_X(x) = P(X \leqslant x) = P(X \leqslant x, Y < +\infty) = F(x, +\infty)$$

$$F_Y(y) = P(Y \leqslant y) = P(X < +\infty, Y \leqslant y) = F(+\infty, y)$$

♣

3.2 二维离散型随机变量的联合概率分布与边缘分布

定义 3.3 (二维离散型随机变量)

若二维离散型随机变量 (X, Y) 只取有限对或可数无限对不同值 (x_i, y_j)，其中 $i, j = 1, 2, \cdots$，则称 (X, Y) 为二维离散型随机变量，同时称

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

为 (X, Y) 的联合概率分布，简称为概率分布或分布律。♣

笔记 二维离散型随机变量 (X, Y) 的联合概率分布也可用表 3.1 表示，称为 (X, Y) 的联合分布列。

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

表 3.1: 二维离散型随机变量 (X,Y) 的联合分布列

在二维离散型随机变量概率分布中，p_{ij}（$i,j=1,2,\cdots$）满足

$$\begin{cases} p_{ij} \geqslant 0 & (3.1) \\ \sum\limits_i \sum\limits_j p_{ij} = 1 & (3.2) \end{cases}$$

由 (3.1) 式、(3.2) 式与二维随机变量分布函数的定义可得二维离散型随机变量 (X,Y) 的分布函数以及 (X,Y) 落入平面区域 G 的概率为

$$F(x,y) = \sum_{x_i \leqslant x} \sum_{y_j \leqslant y} p_{ij}$$

$$P((X,Y) \in G) = \sum_{(x_i,y_j)\in G} p_{ij}$$

下面将给出二维离散型随机变量的边缘概率分布的定义。

定义 3.4 (二维离散型随机变量的边缘概率分布)

设二维离散型随机变量 (X,Y) 的概率分布为

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

则称

$$P(X=x_i) = \sum_j p_{ij}, \qquad i=1,2,\cdots$$

和

$$P(Y=y_j) = \sum_i p_{ij}, \qquad j=1,2,\cdots$$

为二维离散型随机变量 (X,Y) 的关于 X 和 Y 的边缘概率分布，分布列如表 3.2 所示，其中

$$p_{i\cdot} = \sum_j p_{ij},\ p_{\cdot j} = \sum_i p_{ij}$$

♣

X	x_1	x_2	$\cdots$	x_i	$\cdots$
$P(X=x_i)$	$p_{1\cdot}$	$p_{2\cdot}$	$\cdots$	$p_{i\cdot}$	$\cdots$

(a) 关于 X 的边缘概率分布列

Y	y_1	y_2	$\cdots$	y_j	$\cdots$
$P(Y=y_j)$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$

(b) 关于 Y 的边缘概率分布列

表 3.2: 二维离散型随机变量 (X,Y) 的边缘概率分布列

笔记 在表 3.1 的基础上加入边缘概率分布项可得到如表 3.3 所示的新联合分布列。

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	$P(X=x_i)=p_{i\cdot}$
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}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$P(Y=y_j)=p_{\cdot j}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

表 3.3: 加入边缘概率分布项的 (X,Y) 新联合分布列

在一般情况下，联合分布唯一确定其边缘分布，反之不然，除非 X 与 Y 独立。

3.3 二维连续型随机变量的联合密度函数与边缘密度函数

定义 3.5 (二维连续型随机变量的联合概率密度函数)

设 $F(x,y)$ 是二维随机变量 (X,Y) 的联合分布函数，如果存在非负二元函数 $f(x,y)$，对于任意的实数 x,y 都有

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

则称 (X,Y) 为二维连续型随机变量，$f(x,y)$ 称为 (X,Y) 的联合概率密度函数或联合密度函数。♣

性质 联合密度函数 $f(x,y)$ 具有以下性质。

1. $f(x,y)\geqslant 0$。
2. $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}x\mathrm{d}y=1$。
3. 若 D 是 xOy 平面上的任一区域，则随机点 (X,Y) 落在 D 内的概率为

$$P((X,Y)\in D)=\iint_D f(x,y)\mathrm{d}x\mathrm{d}y$$

定义 3.6 (二维连续型随机变量的边缘概率密度函数)

设二维连续型随机变量 (X,Y) 的联合密度函数为 $f(x,y)$，记

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

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

则分别称 $f_X(x)$ 和 $f_Y(y)$ 为 (X,Y) 关于 X 和 Y 的边缘概率密度函数，简称边缘密度函数。♣

定义 3.7 (二维均匀分布)

设 D 为平面上的有界区域，其面积为 A，若二维随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in D\\ 0, & \text{其他}\end{cases}$$

则称 (X,Y) 服从区域 D 上的均匀分布。♣

定义 3.8 (二维正态分布)

若二维随机变量 (X,Y) 的联合密度函数为

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

其中 $\mu_1,\mu_2,\sigma_1>0,\sigma_2>0,\rho$（$|\rho|<1$）均为常数，$x,y\in\mathbf{R}$，则称 (X,Y) 服从参数 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 的二维正态分布，记作 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$。 ♣

笔记 设二维正态随机变量 $(X,Y)\sim N(\mu_1,\sigma_1^2;\mu_2,\sigma_2^2;\rho)$，则其边缘密度函数为

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\qquad -\infty<x<+\infty$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},\qquad -\infty<y<+\infty$$

可以看出，二维正态分布的两个边缘分布都是一维正态分布，但反之不然。

例题 3.1 某二维连续型随机变量的联合密度函数为

$$f(x,y)=\begin{cases}C, & (x,y)\in G\\ 0, & \text{其他}\end{cases}$$

其中 $G:x^2+y^2\leqslant r^2$，求 C 的值。

解 由题意有 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\mathrm{d}y=C\iint\limits_G\mathrm{d}x\mathrm{d}y=C\pi r^2=1$，故 $C=\dfrac{1}{\pi r^2}$。

例题 3.2 某二维连续型随机变量 (X,Y) 的联合密度函数为 $f(x,y)=\dfrac{1}{\pi^2(1+x^2)(1+y^2)}$，求 $f_X(x)$ 与 $f_Y(y)$。

解 由题意有 $f_X(x)=\displaystyle\int_{-\infty}^{+\infty}\frac{1}{\pi^2(1+x^2)(1+y^2)}\mathrm{d}y=\frac{1}{\pi(1+x^2)}$，同理 $f_Y(y)=\dfrac{1}{\pi(1+y^2)}$。

例题 3.3 已知二维连续型随机变量 (X,Y) 的联合密度函数为

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

求边缘密度函数 $f_X(x)$ 与 $f_Y(y)$。

解 对于 x 而言，需要对区域进行分类讨论。

1. 当 $|x|\leqslant r$ 时，$f_X(x)=\displaystyle\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}}\frac{1}{\pi r^2}\mathrm{d}y=\frac{2\sqrt{r^2-x^2}}{\pi r^2}$；
2. 当 $|x|>r$ 时，$f_X(x)=0$。

故有

$$f_X(x)=\begin{cases}\dfrac{2\sqrt{r^2-x^2}}{\pi r^2}, & |x|\leqslant r\\ 0, & \text{其他}\end{cases}$$

同理，有

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

注 在三维空间中作出边缘密度函数 $f_X(x)$ 的大致示意图，可看出在区域 $x^2+y^2\leqslant r^2$ 上，其值恰好是投影在 yOz 平面上的矩形的面积，$f_Y(y)$ 同理。

3.4 条件分布

定义 3.9 (一维条件分布)

设有事件 A 和随机变量 X，其中 X 的分布函数为 $F(x)$，则在事件 A 发生的前提下 X 的分布称为 X 的条件分布，记作 $F(x\,|\,A)$，即有 $F(x\,|\,A)=P(X\leqslant x\,|\,A)$。♣

定义 3.10 (二维离散型随机变量的条件分布)

设离散型随机变量 (X,Y) 的联合概率分布为

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

关于 X 和 Y 的边缘分布列为

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

对某一固定的 i，若 $P(X=x_i)=p_{i\cdot}>0$，则称

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

为在 $X=x_i$ 的条件下随机变量 Y 的条件分布律；同理地，对于某一固定的 j，若 $P(Y=y_j)>0$，则称

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

为在 $Y=y_j$ 的条件下随机变量 X 的条件分布律。♣

性质 二维离散型随机变量的条件分布具有以下性质。

1. $P(\,X=x_i\,|\,Y=y_j\,)\geqslant 0$，$P(\,Y=y_j\,|\,X=x_i\,)\geqslant 0$。
2. $\sum\limits_i P(\,X=x_i\,|\,Y=y_j\,)=1$，$\sum\limits_j P(\,Y=y_j\,|\,X=x_i\,)=1$。

定义 3.11 (二维连续型随机变量的条件分布)

对于任意 $\varepsilon>0$，若 $P(x-\varepsilon<X\leqslant x+\varepsilon)>0$，对任意实数 y，有极限

$$\lim_{\varepsilon\to 0}P(\,Y\leqslant y\,|\,x-\varepsilon<X\leqslant x+\varepsilon\,)=\lim_{\varepsilon\to 0}\frac{P(x-\varepsilon<X\leqslant x+\varepsilon,Y\leqslant y)}{P(x-\varepsilon<X\leqslant x+\varepsilon)}$$

存在，则称其为在给定条件 $X=x$ 下随机变量 Y 的条件分布函数，记作 $F_{Y|X}(\,y\,|\,x\,)$ 或 $F(\,y\,|\,x\,)$。♣

由上述定义及积分中值定理可推得二维连续型随机变量的条件分布的计算公式为

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

分别记在 $X=x$ 条件下 Y 的**条件密度函数**和在 $Y=y$ 条件下 X 的条件密度函数为 $f(\,y\,|\,x\,)$ 和 $f(\,x\,|\,y\,)$，则有

$$f(\,y\,|\,x\,)=\frac{f(x,y)}{f_X(x)},\qquad f(\,x\,|\,y\,)=\frac{f(x,y)}{f_Y(y)}$$

其中 $f(x,y)$ 是 X 和 Y 的联合密度函数。

设二维随机变量 $(X,Y)\sim N(\mu_1,\sigma_1;\mu_2,\sigma_2;\rho)$，则在 $X=x$ 条件下 Y 的条件分布为正态分布

$$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2)\right)$$

同理，在 $Y=y$ 条件下 X 的条件分布为正态分布

$$N\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\sigma_1^2(1-\rho^2)\right)$$

例题 3.4 设随机变量 X 的概率密度函数为 $f(x)=\dfrac{1}{\pi(1+x^2)}$，求在 $X>1$ 的条件下 X 的条件分布。

解 由题意有 $F(\,x\,|\,X>1\,)=P(\,X\leqslant x\,|\,X>1\,)=\dfrac{P(\,X\leqslant x,\,X>1\,)}{P(X>1)}$；

1. 当 $x\leqslant 1$ 时，$F(\,x\,|\,X>1\,)=0$；
2. 当 $x>1$ 时，$P(1<X\leqslant x)=\displaystyle\int_1^x\frac{1}{\pi(1+t^2)}\mathrm{d}t=\frac{1}{\pi}\arctan x-\frac{1}{4}$，$P(X>1)=\displaystyle\int_1^{+\infty}\frac{1}{\pi(1+t^2)}\mathrm{d}t=\frac{1}{4}$，故 $F(\,x\,|\,X>1\,)=\dfrac{4}{\pi}\arctan x-1$。

综上，在 $X>1$ 的条件下 X 的条件分布为

$$F(\,x\,|\,X>1\,)=\begin{cases}\dfrac{4}{\pi}\arctan x-1, & x>1\\[2ex] 0, & x\leqslant 1\end{cases}$$

例题 3.5 某二维连续型随机变量 (X,Y) 的联合密度函数为 $f(x,y)=\dfrac{1}{\pi^2(1+x^2)(1+y^2)}$，求 $f(\,x\,|\,y\,)$ 和 $f(\,y\,|\,x\,)$。

解 由例 3.2 可知 $f_X(x)=\dfrac{1}{\pi(1+x^2)}$ 且 $f_Y(y)=\dfrac{1}{\pi(1+y^2)}$，故有

$$f(\,x\,|\,y\,)=\frac{f(x,y)}{f_Y(y)}=\frac{1}{\pi(1+x^2)}$$

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

3.5 随机变量的独立性

对事件的独立性进行拓展，可得到如下二维随机变量独立性的定义。

定义 3.12 (随机变量的独立性)

设二维随机变量 (X,Y) 的联合分布函数为 $F(x,y)$，关于 X 和 Y 的边缘分布函数为 $F_X(x)$ 和 $F_Y(y)$，若对任意的实数 x 和 y 有 $F(x,y)=F_X(x)F_Y(y)$，则称随机变量 X 和 Y 相互独立。

♣

笔记 上述定义中 $F(x,y)=F_X(x)F_Y(y)$ 等价于

1. 由分布函数定义有

$$P(X\leqslant x,Y\leqslant y)=P(X\leqslant x)P(Y\leqslant y)$$

2. 当 (X,Y) 为离散型随机变量时，

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j),\qquad i,j=1,2,\cdots$$

3. 当 (X,Y) 为连续型随机变量时，

$$f(x,y)=f_X(x)f_Y(y)$$

若随机变量 X 与 Y 独立，则由这二者构造的函数 $g_1(X)$ 与 $g_2(Y)$ 也独立。

笔记 若 X 与 Y 独立，则 X^2 与 Y^2 独立，a_1X+b_1 与 a_2Y+b_2 独立（$a_1,a_2\neq 0$）。

例题 3.6 二维随机变量 (X,Y) 的联合分布列如表 3.4 所示，试判断 X 和 Y 是否相互独立。

$X \backslash Y$	0	1
0	0.2	0.3
1	0.2	0.3

表 3.4: 例 3.6 中 (X,Y) 的联合分布列

解 由题意，$P(X=0)=P(X=1)=0.5$，$P(Y=0)=0.4$，$P(Y=1)=0.6$，且有

$$P(X=0,Y=0)=0.2=0.5\times 0.4=P(X=0)P(Y=0)$$

$$P(X=0,Y=1)=0.3=0.5\times 0.6=P(X=0)P(Y=1)$$

$$P(X=1,Y=0)=0.2=0.5\times 0.4=P(X=1)P(Y=0)$$

$$P(X=1,Y=1)=0.3=0.5\times 0.6=P(X=1)P(Y=1)$$

故 X 和 Y 相互独立。

例题 3.7 一经理到达办公室的时刻服从均匀分布，时间范围为 8:00 至 12:00，其秘书到达办公室的时刻也服从均匀分布，时间范围为 7:00 至 9:00。现已知二人到达的时刻是相互独立的，求他们到达的时刻之差不超过 5 分钟的概率。

解 设 X 为经理到达办公室的时刻，Y 为秘书到达办公室的时刻，则有

$$f_X(x)=\begin{cases}\dfrac{1}{4}, & 8\leqslant x\leqslant 12\\[2ex] 0, & \text{其他}\end{cases}$$

$$f_Y(y)=\begin{cases}\dfrac{1}{2}, & 7\leqslant y\leqslant 9\\[2ex] 0, & \text{其他}\end{cases}$$

由于二人到达的时刻相互独立，则有

$$\begin{aligned} f(x,y)&=f_X(x)f_Y(y)\\ &=\begin{cases}\dfrac{1}{8}, & 8\leqslant x\leqslant 12,\ 7\leqslant y\leqslant 9\\[2ex] 0, & \text{其他}\end{cases}\end{aligned}$$

故他们到达的时刻之差不超过 5 分钟的概率为

$$\begin{aligned} P\left(|X-Y|\leqslant\frac{1}{12}\right)&=\iint\limits_G f(x,y)\mathrm{d}x\mathrm{d}y\\ &=\frac{1}{8}\iint\limits_G \mathrm{d}x\mathrm{d}y\\ &=\frac{1}{8}\times\frac{1}{2}\times\left[\left(\frac{13}{12}\right)^2-\left(\frac{11}{12}\right)^2\right]\\ &=\frac{1}{48}\end{aligned}$$

其中区域 G 由 $8 \leqslant x \leqslant 12$、$7 \leqslant y \leqslant 9$、$y = x - \frac{1}{12}$ 和 $y = x + \frac{1}{12}$ 围成。

3.6 二维随机变量函数的分布

与一维随机变量函数同理，二维随机变量函数的分布也需要从离散型和随机型两方面进行讨论。

定理 3.1 (二维离散型随机变量函数的分布)

设二维离散型随机变量 (X,Y) 的分布律为

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

$Z=g(X,Y)$ 的可能取值为 z_k（$k=1,2,\cdots$），则 Z 的分布律为

$$P(Z=z_k) = P(g(X,Y)=z_k) = \sum_{g(x_i,y_i)=z_k} p_{ij}, \qquad k=1,2,\cdots$$

定理 3.2 (二维连续型随机变量函数的分布)

设二维连续型随机变量 (X,Y) 的联合密度函数为 $f(x,y)$，若 $Z=g(X,Y)$ 为连续型变量，则 Z 的分布函数为

$$F_Z(z) = P(Z \leqslant z) = P(g(X,Y) \leqslant z) = \iint_{g(x,y)\leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y$$

求得 Z 的分布函数 $F_Z(z)$ 后，经求导可得到 Z 的概率密度函数 $f_Z(z) = F_Z'(z)$。

相比于一维随机变量，二维连续型随机变量函数在某些情况下有着如下所示的特殊的分布。

1. 当 $Z = X+Y$ 时，有

$$\begin{aligned} F_Z(z) = P(Z \leqslant z) = P(X+Y \leqslant z) \\ &= \iint_{x+y\leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z-x} f(x,y)\mathrm{d}y \xlongequal{t=x+y} \int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z} f(x,t-x)\mathrm{d}t \\ &= \int_{-\infty}^{z}\mathrm{d}t\int_{-\infty}^{+\infty} f(x,t-x)\mathrm{d}x \end{aligned}$$

故 $f_Z(z) = \int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x$，同理有 $f_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y$；

特别地，当 X 与 Y 独立时有

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\mathrm{d}y \tag{3.3}$$

公式 (3.3) 被称作概率密度函数的**卷积公式**。

2. 当 $Z = \dfrac{X}{Y}$ 时，有

$$\begin{aligned} F_Z(z) = P(Z \leqslant z) = P\left(\frac{X}{Y} \leqslant z\right) \\ &= \iint_{\frac{x}{y}\leqslant z} f(x,y)\mathrm{d}x\mathrm{d}y \end{aligned}$$

对上式进行变换有 $F_Z(z)=\int_{-\infty}^{z}\mathrm{d}u\int_{-\infty}^{+\infty}f(yu,y)\cdot|\,y\,|\mathrm{d}y$，从而得到 Z 的概率密度函数为

$$f_Z(z)=\int_{-\infty}^{+\infty}f(yz,y)\cdot|\,y\,|\mathrm{d}y$$

3. 当 $M=\max\{\,X,Y\,\}$ 和 $N=\min\{\,X,Y\,\}$ 时，设随机变量 X 与 Y 相互独立，则有

$$\begin{aligned}F_M(z)=P(M\leqslant z)=P(\max\{\,X,Y\,\}\leqslant z)&=P(X\leqslant z,Y\leqslant z)\\&=P(X\leqslant z)P(Y\leqslant z)\\&=F_X(z)F_Y(z)\end{aligned}$$

即有 $F_M(z)=F_X(z)F_Y(z)$，同理有 $F_N(z)=1-[\,1-F_X(z)\,][\,1-F_Y(z)\,]$。

例题 3.8 设二维离散型随机变量 (X,Y) 的联合分布列如表 3.5 所示，求随机变量函数 $Z=XY$ 的分布列。

X \\ Y	4	4.2
5	0.2	0.4
5.1	0.3	0.1

表 3.5: 例 3.8 中 (X,Y) 的联合分布列

解 由题意可知 Z 的所有可能取值为 20、20.4、21、21.42，则 Z 的分布列如表 3.6 所示。

Z	20	20.4	21	21.42
P	0.2	0.3	0.4	0.1

表 3.6: 例 3.8 所求的 Z 的分布列

例题 3.9 设随机变量 X_1 与 X_2 独立，且均服从 0-1 分布，参数均为 p，求随机变量函数 X_1+X_2 的分布列。

解 由题意及 0-1 分布的定义可知 X_1+X_2 的所有可能取值为 0、1、2，则 X_1+X_2 的分布列如表 3.7 所示。

X_1+X_2	0	1	2
P	$(1-p)^2$	$2p(1-p)$	p^2

表 3.7: 例 3.9 所求的 X_1+X_2 的分布列

注 由 X_1+X_2 的分布列可知 $X_1+X_2\sim B(2,p)$。

例题 3.10 设随机变量 X 与 Y 独立，且分别服从参数为 λ_1 和 λ_2 的泊松分布，求随机变量函数 $Z=X+Y$ 的分布律。

解 由题意及泊松分布的定义可知

$$\{\,Z=k\,\}=\sum_{i=0}^{k}\{\,X=i,Y=k-i\,\},\qquad k=0,1,2,\cdots$$

故 Z 的分布律为

$$\begin{aligned}P(Z=k)=\sum_{i=0}^{k}P(X=i,Y=k-i)&=\sum_{i=0}^{k}P(X=i)P(Y=k-i)\\&=\sum_{i=0}^{k}\frac{\lambda_1^i}{i!}\mathrm{e}^{-\lambda_1}\cdot\frac{\lambda_2^{k-i}}{(k-i)!}\mathrm{e}^{-\lambda_2}\\&=\frac{(\lambda_1+\lambda_2)^k}{k!}\mathrm{e}^{-(\lambda_1+\lambda_2)},\qquad k=0,1,2,\cdots\end{aligned}$$

注 若 $X \sim P(\lambda_1)$，$Y \sim P(\lambda_2)$，则对于 $Z = X + Y$ 有 $Z \sim P(\lambda_1 + \lambda_2)$，即泊松分布具有可加性。

例题 3.11 设随机变量 (X,Y) 的联合密度函数为 $f(x,y) = \frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}}$，求随机变量函数 $Z = \sqrt{X^2+Y^2}$ 的概率密度函数。

解 需要对 z 的区间进行分类讨论。

1. 当 $z < 0$ 时，$F_Z(z) = 0$;
2. 当 $Z \geqslant 0$ 时，

$$
\begin{aligned}
F_Z(z) = P(X^2+Y^2 \leqslant z^2) &= \iint\limits_G \frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}} \\
&= \int_0^{2\pi} \mathrm{d}\theta \int_0^z \frac{1}{2\pi}\mathrm{e}^{-\frac{r^2}{2}} r\mathrm{d}r = 1 - e^{-z^2}
\end{aligned}
$$

即 Z 的分布函数为

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

故 Z 的概率密度函数为

$$
f_Z(z) = \begin{cases} 2z\mathrm{e}^{-z^2}, & z \geqslant 0 \\ 0, & z < 0 \end{cases}
$$

第 3 章 练习

1. 已知二维连续型随机变量 (X,Y) 的联合密度函数为

$$
f(x,y) = \begin{cases} \mathrm{e}^{-(x+y)}, & x > 0,\ y > 0 \\ 0, & \text{其他} \end{cases}
$$

试求：

(a). 联合分布函数 $F(x,y)$;

(b). 概率值 $P((X,Y) \in G)$，其中区域 G 由 x 轴、y 轴和直线 $x+y=1$ 围成;

(c). 关于 X 的边缘分布函数 $F_X(x)$。

解 由相关定义有

(a). 对联合分布函数进行讨论，有

I. 当 x 和 y 之中有一个为负数时，$F(x,y) = 0$;

II. 当 x 和 y 均为正数时，

$$
\begin{aligned}
F(x,y) &= P(X \leqslant x, Y \leqslant y) \\
&= \int_{-\infty}^x \int_{-\infty}^y f(s,t)\mathrm{d}s\mathrm{d}t \\
&= \int_0^x \int_0^y \mathrm{e}^{-s}\mathrm{e}^{-t}\mathrm{d}s\mathrm{d}t \\
&= \int_0^x \mathrm{e}^{-s}\mathrm{d}s \cdot \int_0^y \mathrm{e}^{-t}\mathrm{d}t \\
&= (1-\mathrm{e}^{-x})(1-\mathrm{e}^{-y})
\end{aligned}
$$

综上，联合分布函数为

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

(b). 由题意可得

$$\begin{aligned}P((X,Y)\in G)&=\iint\limits_G f(x,y)\mathrm{d}x\mathrm{d}y=\iint\limits_G \mathrm{e}^{-(x+y)}\mathrm{d}x\mathrm{d}y\\&=\int_0^1\mathrm{d}x\int_0^{1-x}\mathrm{e}^{-(x+y)}\mathrm{d}y\\&=1-2\mathrm{e}^{-1}\end{aligned}$$

(c). 当 $x\leqslant 0$ 时，$F_X(x)=0$；当 $x>0$ 时，$F_X(x)=\lim\limits_{y\to+\infty}F(x,y)=1-\mathrm{e}^{-x}$。故有

$$F_X(x)=\begin{cases}1-\mathrm{e}^{-x}, & x>0\\ 0, & \text{其他}\end{cases}$$

2. 已知二维连续型随机变量 (X,Y) 的联合密度函数为

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

求边缘密度函数 $f_X(x)$。

解 对于 x 而言，需要对区域进行分类讨论。

(a). 当 $0\leqslant x\leqslant 1$ 时，$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\int_0^1 2(x+y-2xy)\mathrm{d}y=1$；

(b). 当 $x\notin[0,1]$ 时，$f_X(x)=0$。

故有

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

笔记 二维随机变量 (X,Y) 的各一维分量都服从均匀分布，不能推出其也服从均匀分布。

3. 已知二维连续型随机变量 (X,Y) 的联合密度函数为

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

求条件密度函数 $f(y\,|\,x)$。

解 由例 3.3 可知

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

(a). 当 $|\,y\,|\leqslant r$ 时，$f(x\,|\,y)=\dfrac{f(x,y)}{f_Y(x)}=\dfrac{1}{2\sqrt{r^2-y^2}}$；

(b). 当为其他情况时，$f(x\,|\,y)=0$。

故有

$$f(x\,|\,y)=\begin{cases}\dfrac{1}{2\sqrt{r^2-y^2}}, & -\sqrt{r^2-y^2}\leqslant x\leqslant\sqrt{r^2-y^2}\\ 0, & \text{其他}\end{cases}$$

4. 设有二维随机变量 (X,Y)，且 X 与 Y 相互独立，其中 X 服从在 $[0,1]$ 上的均匀分布，Y 服从参数为 3 的指数分布，求随机变量函数 $M=\max\{X,Y\}$ 与 $N=\min\{X,Y\}$ 的分布函数。

解 由题意及相关定义，可知 $X\sim U[0,1]$ 且 $Y\sim E(3)$，分别求得 X 与 Y 的边缘密度函数为

$$f_X(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & \text{其他}\end{cases},\qquad f_Y(y)=\begin{cases}3\mathrm{e}^{-3y}, & y>0\\ 0, & y\leqslant 0\end{cases}$$

由此可得分布函数 $F_X(x)$ 与 $F_Y(y)$ 为

$$F_X(x)=\begin{cases}0, & x<0\\ x, & 0\leqslant x<1\\ 1, & x\geqslant 1\end{cases},\qquad F_Y(y)=\begin{cases}1-\mathrm{e}^{-3y}, & y>0\\ 0, & y\leqslant 0\end{cases}$$

故有

$$F_M(z)=\begin{cases}0, & z<0\\ z(1-\mathrm{e}^{-3z}), & 0\leqslant z<1\\ 1-\mathrm{e}^{-3z}, & z\geqslant 1\end{cases},\qquad F_N(z)=\begin{cases}0, & z<0\\ 1-(1-z)\mathrm{e}^{-3z}, & 0\leqslant z<1\\ 1, & z\geqslant 1\end{cases}$$

5. 设 $X\sim N(0,1)$，$Y\sim N(0,1)$，X 与 Y 独立，求随机变量函数 $Z=X+Y$ 的概率密度函数。

解 由于 X 与 Y 独立，且有 $Z=X+Y$，则由卷积公式可得

$$\begin{aligned}f_Z(z)&=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)\mathrm{d}x\\&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(z-x)^2}{2}}\mathrm{d}x\\&=\int_{-\infty}^{+\infty}\frac{1}{2\pi}\mathrm{e}^{-\frac{z^2}{4}}\mathrm{e}^{-(x-\frac{z}{2})^2}\mathrm{d}x\\&=\frac{1}{2\pi}\mathrm{e}^{-\frac{z^2}{4}}\int_{-\infty}^{+\infty}\mathrm{e}^{-(x-\frac{z}{2})^2}\mathrm{d}\left(x-\frac{z}{2}\right)\\&=\frac{\sqrt{\pi}}{2\pi}e^{-\frac{z^2}{4}}\end{aligned}$$

故 $f_Z(z)=\dfrac{\sqrt{\pi}}{2\pi}e^{-\frac{z^2}{4}}=\dfrac{1}{\sqrt{2\pi}\cdot\sqrt{2}}\mathrm{e}^{-\frac{z^2}{2\cdot(\sqrt{2})^2}}$，可看出 $Z\sim N(0,2)$。

注 若 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，则对于 $Z=X+Y$ 有 $Z\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

第 4 章 随机变量的数字特征

4.1 数学期望

定义 4.1 (离散型随机变量的数学期望)

设离散型随机变量 X 的概率函数为

$$P(X=x_k)=p_k, \qquad k=1,2,\cdots$$

若级数 $\sum\limits_{k=1}^{+\infty} x_k p_k$ 绝对收敛，则称级数 $\sum\limits_{k=1}^{+\infty} x_k p_k$ 的和为离散型随机变量的数学期望或均值，记为 $E(X)$。♣

定义 4.2 (连续型随机变量的数学期望)

设连续型随机变量 X 的概率密度函数为 $f(x)$，若积分 $\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ 绝对收敛，则称该积分值为 X 的数学期望或均值，即 $E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$。♣

下表给出了一些常见概率分布的数学期望。

$X \sim$	$G(p)$	$B(n,p)$	$P(\lambda)$	$H(n,M,N)$	$U[a,b]$	$E(\lambda)$	$N(\mu,\sigma^2)$
$E(X)$	$\frac{1}{p}$	np	λ	$\frac{n}{N}M$	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ

表 4.1: 常见概率分布的数学期望

定理 4.1 (一维随机变量函数的数学期望)

设有随机变量 X 和随机变量连续函数 $Y=g(X)$。

1. 若 X 是离散型随机变量，其概率函数为 $P(X=x_k)=p_k$，且级数 $\sum\limits_k g(x_k)p_k$ 绝对收敛，则有
$$E(Y)=E[g(X)]=\sum_k g(x_k)p_k$$
2. 若 X 是连续型随机变量，其概率密度函数为 $f(x)$，且积分 $\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$ 绝对收敛，则有
$$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$$

♡

定理 4.2 (二维随机变量函数的数学期望)

设 (X,Y) 为二维随机变量，且有 $Z=g(X,Y)$。

1. 若 (X,Y) 是二维离散型随机变量，其联合分布律为 $P(X=x_i,Y=y_j)=p_{ij}$，则有
$$E(Z)=E[g(X,Y)]=\sum_i\sum_j g(x_i,y_i)p_{ij}$$
2. 若 (X,Y) 是二维连续型随机变量，其联合密度函数为 $f(x,y)$，则有
$$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$$

♡

性质 数学期望具有以下性质。

1. 对于常数 C 有 $E(C)=C$。
2. 设有常数 C_1 和 C_2，X 为随机变量，则有 $E(C_1X+C_2)=C_1E(X)+C_2$。
3. 设 $X_1,X_2,\cdots,X_n$ 为任意有限 n 个随机变量，则有 $E\left(\sum\limits_{i=1}^{n}X_i\right)=\sum\limits_{i=1}^{n}E(X_i)$。
4. 设 $X_1,X_2,\cdots,X_n$ 为任意有限 n 个相互独立的随机变量，则有 $E\left(\prod\limits_{i=1}^{n}X_i\right)=\prod\limits_{i=1}^{n}E(X_i)$。
5. 设有随机变量 X 和 Y，若 $E(X^2)$ 和 $E(Y^2)$ 存在，则有
$$[E(XY)]^2\leqslant E(X^2)\cdot E(Y^2)$$
该不等式称为柯西–施瓦茨不等式。

条件分布的数学期望称为**条件数学期望**或**条件期望**，下面给出其定义。

定义 4.3 (条件期望)

设 (X,Y) 为二维随机变量。

1. 若 (X,Y) 为二维离散型随机变量，其联合分布律为 $P(X=x_i,Y=y_j)=p_{ij}$，则有
$$E(X\,|\,Y=y)=\sum_i x_iP(X=x_i\,|\,Y=y)$$
$$E(Y\,|\,X=x)=\sum_j y_jP(Y=y_j\,|\,X=x)$$
2. 若 (X,Y) 为二维连续型随机变量，其联合密度函数为 $f(x,y)$，则有
$$E(X\,|\,Y=y)=\int_{-\infty}^{+\infty}xf(x\,|\,y)\mathrm{d}x$$
$$E(Y\,|\,X=x)=\int_{-\infty}^{+\infty}yf(y\,|\,x)\mathrm{d}y$$

♣

注 条件期望 $E(X\,|\,Y=y)$ 是关于 y 的函数，且条件期望具备数学期望的一切性质。

例题 4.1 设随机变量 X 的概率密度函数为
$$f(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{其他}\end{cases}$$
求 $E(X)$ 的值。

解 由题意可得 $E(X)=\displaystyle\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\int_0^1 2x^2\mathrm{d}x=\frac{2}{3}$。

例题 4.2 设随机变量 X 的概率密度函数为
$$f(x)=\begin{cases}\dfrac{1}{2}, & 0\leqslant x\leqslant 2\\ 0, & \text{其他}\end{cases}$$
且随机变量函数 $Y=4X+1$，求 Y 的数学期望。

解 由题意有 $E(Y)=E(4X+1)=\displaystyle\int_{-\infty}^{+\infty}(4x+1)f(x)\mathrm{d}x=\int_0^2\frac{1}{2}(4x+1)\mathrm{d}x=5$。

例题 4.3 设随机变量 (X,Y) 的联合密度函数为
$$f(x,y)=\begin{cases}x+y, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

求随机变量函数 $Z = XY$ 的数学期望。

解 由题意有 $E(Z) = E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y = \int_0^1\int_0^1 xy(x+y)\mathrm{d}x\mathrm{d}y = \frac{1}{3}$。

4.2 方差

定义 4.4 (方差)

设 X 为随机变量，如果 $E[X-E(X)]^2$ 存在，则称之为 X 的方差，记作 $D(X)$。

1. 若 X 为离散型随机变量，设 $P(X = x_k) = p_k$（$k = 1, 2, \cdots$），则

$$D(X) = \sum_k [x_k - E(X)]^2 p_k$$

2. 若 X 为连续型随机变量，设 X 的概率密度函数为 $f(x)$，则

$$D(X) = \int_{-\infty}^{+\infty} [X - E(X)]^2 f(x)\mathrm{d}x$$

笔记 一般地，我们称 $\sqrt{D(X)}$ 为标准差。

通常，我们并不会使用定义式来计算方差，而是利用下面给出的数学期望与方差的关系式：

$$D(X) = E(X^2) - E^2(X)$$

下表给出了一些常见概率分布的方差。

$X\sim$	$G(p)$	$B(n,p)$	$P(\lambda)$	$H(n,M,N)$	$U[a,b]$	$E(\lambda)$	$N(\mu,\sigma^2)$
$D(X)$	$\frac{1-p}{p^2}$	$np(1-p)$	λ	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$	$\frac{(b-a)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2

表 4.2: 常见概率分布的方差

性质 方差具有以下性质。

1. 对于常数 C 有 $D(C) = 0$。
2. 设有常数 C_1 和 C_2，X 为随机变量，则有 $D(C_1X + C_2) = C_1^2 D(X)$。
3. 设 $X_1, X_2, \cdots, X_n$ 为任意 n 个相互独立的随机变量，则有 $D(X_1 \pm X_2 \pm \cdots \pm X_n) = \sum_{i=1}^{n} D(X_i)$。
4. 设 X 为随机变量，则 $D(X) = 0$ 的充要条件为 $P(X = E(X)) = 1$。

标准化随机变量是指经过处理从而具有一些较好性质的随机变量，下面给出其定义。

定义 4.5 (标准化随机变量)

设 X 为随机变量，则称随机变量

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}$$

为标准化随机变量。

笔记 对于任意的随机变量 X，必有 $E(X^*) = 0$ 且 $D(X^*) = 1$。

例题 4.4 已知随机变量 X 的概率密度函数为

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

求 $D(X)$ 的值。

解 由题意有 $E(X)=\int_0^1 x\cdot 2x\mathrm{d}x=\frac{2}{3}$，$E(X^2)=\int_0^1 x^2\cdot 2x\mathrm{d}x=\frac{1}{2}$，故 $D(X)=E(X^2)-E^2(X)=\frac{1}{18}$。

4.3 协方差与相关系数

定义 4.6 (协方差)

设 (X,Y) 为二维随机变量，若

$$E\{[X-E(X)][Y-E(Y)]\}$$

存在，则称其为随机变量 X 与 Y 的协方差，记作 $\mathrm{Cov}(X,Y)$ 或 σ_{XY}。♣

定义 4.7 (相关系数)

记 $\sigma_X=\sqrt{D(X)}$，$\sigma_Y=\sqrt{D(Y)}$，若 $\sigma_X\neq 0$，$\sigma_Y\neq 0$，则称

$$\rho_{XY}=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

为随机变量 X 与 Y 的相关系数。♣

注 设有随机变量 X 和 Y，则有 $\mathrm{Cov}(X^*,Y^*)=\rho_{XY}$。

定理 4.3 (协方差的计算公式)

设有二维随机变量 (X,Y)，$\mathrm{Cov}(X,Y)$ 为 X 与 Y 的协方差。

1. 一般常用的计算公式为
$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)$$
2. 当 (X,Y) 为二维离散型随机变量时，计算公式为
$$\mathrm{Cov}(X,Y)=\sum_i\sum_j[x_i-E(X)][y_j-E(Y)]p_{ij}$$
3. 当 (X,Y) 为二维连续型随机变量时，计算公式为
$$\mathrm{Cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(X)][y-E(Y)]f(x,y)\mathrm{d}x\mathrm{d}y$$

♡

性质 协方差具有以下性质。

1. 设有常数 C，X 为随机变量，则有 $\mathrm{Cov}(C,X)=0$。
2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$。
3. 设有常数 a_1,a_2,b_1,b_2，则有 $\mathrm{Cov}(a_1X+b_1,a_2Y+b_2)=a_1a_2\mathrm{Cov}(X,Y)$。
4. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$。
5. $D(X\pm Y)=D(X)+D(Y)\pm 2\mathrm{Cov}(X,Y)$，并且当 X 与 Y 相互独立时 $\mathrm{Cov}(X,Y)=0$。

性质 相关系数具有以下性质。

1. $|\rho_{XY}|\leqslant 1$。
2. $|\rho_{XY}|=1$ 的充要条件是 $P(Y=aX+b)=1$，其中 a,b 为常数且 $a\neq 0$。

相关系数刻画的是 X 与 Y 之间的线性关系的程度，$|\rho_{XY}|$ 越大说明 X 与 Y 之间的线性关系程度越好，反之越差，并将 $|\rho_{XY}|=0$ 的情况称为 **X 与 Y 不相关**。若 X 与 Y 独立，则 $\mathrm{Cov}(X,Y)=0$，推出 $\rho_{XY}=0$，即 X 与 Y 不相关；反之，当 X 与 Y 不相关时，不能保证 X 与 Y 独立。

笔记 对于二维正态随机变量 (X,Y) 而言，X 与 Y 独立和 X 与 Y 不相关是等价的。

例题 4.5 设随机变量 (X,Y) 的联合密度函数为

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

求 $\mathrm{Cov}(X,Y)$ 的值。

解 由题意可知

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

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

故 $E(X)=\int_0^1 x\cdot(x+\frac{1}{2})\mathrm{d}x=\frac{7}{12}$，同理 $E(Y)=\frac{7}{12}$，$E(XY)=\int_0^1\int_0^1 xy(x+y)\mathrm{d}x\mathrm{d}y=\frac{1}{3}$，故有

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y)=-\frac{1}{144}$$

4.4 原点矩与中心矩

定义 4.8 (原点矩、中心矩、混合矩与中心混合矩)

设有随机变量 X 和 Y，且有 $k,l=1,2,\cdots$，则

1. 若 $E(X^k)$ 存在，则称其为 X 的 k 阶原点矩；
2. 若 $E\{[X-E(X)]^k\}$ 存在，则称其为 X 的 k 阶中心矩；
3. 若 $E(X^kY^l)$ 存在，则称其为 X 和 Y 的 $k+l$ 阶混合矩；
4. 若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$ 存在，则称其为 X 和 Y 的 $k+l$ 阶中心混合矩。

♣

笔记 对于随机变量 X,Y，$E(X)$, $E(Y)$ 是一阶原点矩，$D(X),D(Y)$ 是二阶中心矩，而 $\mathrm{Cov}(X,Y)$ 是二阶混合中心距。

注 对于所有的随机变量而言，一阶中心矩均为 0。

在求原点矩与中心矩时，我们需要根据随机变量的类型来确定计算方式。

定理 4.4 (原点矩与中心矩的计算公式)

设有随机变量 X，概率函数为 $P(X=X_k)=p_k$，概率密度函数为 $f(x)$，且有 $k=1,2,\cdots$，则

1. 若 X 为离散型随机变量，则

$$E(X^k)=\sum_i x_i^k p_i$$

$$E\{[X-E(X)]^k\}=\sum_i[x_i-E(X)]^k p_i$$

2. 若 X 为连续型随机变量，则

$$E(X^k)=\int_{-\infty}^{+\infty}x^k f(x)\mathrm{d}x$$

$$E\{[X-E(X)]^k\}=\int_{-\infty}^{+\infty}[X-E(X)]^k f(x)\mathrm{d}x$$

♡

定义 4.9 (协方差矩阵)

设 $(X_1, X_2, \ldots, X_n)$ 为 n 维随机变量，且记 $c_{ij} = \mathrm{Cov}(X_i, X_j)$，其中 $i, j = 1, 2, \cdots, n$，若它们都存在，则称矩阵

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 $(X_1, X_2, \ldots, X_n)$ 的协方差矩阵。♣

笔记 协方差矩阵是非负定对称矩阵。

4.5 大数定律与中心极限定理

定理 4.5 (切比雪夫不等式)

设有随机变量 X，若 $E(X)$ 和 $D(X)$ 均存在，则对任意的 $\varepsilon > 0$ 有

$$P(\,|\,X - E(X)\,| \geqslant \varepsilon\,) \leqslant \frac{D(X)}{\varepsilon^2}$$

♡

定义 4.10 (依概率收敛)

对于随机变量 X 和随机变量列 $\{X_n\}$，若对于任意的 $\varepsilon > 0$，都有

$$\lim_{n \to \infty} P(\,|\,X_n - X\,| < \varepsilon\,) = 1$$

则称 $\{X_n\}$ 依概率收敛于 X，记作 $X_n \xrightarrow{P} X$。♣

概率论中用来阐明大量随机现象平均结果的稳定性的定理统称为**大数定律**。

定理 4.6 (伯努利大数定律)

设 k_n 为 n 重伯努利试验中事件 A 发生的次数，p 为事件 A 在每次试验中发生的概率，则对于任意给定的 $\varepsilon > 0$，都有

$$\lim_{n \to \infty} P\left(\left|\frac{k_n}{n} - p\right| < \varepsilon\right) = 1$$

♡

笔记 伯努利大数定律从理论上说明任一随机事件的频率具有稳定性。

定义 4.11 (独立同分布)

设 $X_1, X_2, \cdots, X_n, \cdots$ 为一系列随机变量，若它们服从同一分布且相互独立，则称其为是独立同分布的。♣

将伯努利大数定律中的分布条件推广到一般情况，可得到下面的**辛钦大数定律**。

定理 4.7 (辛钦大数定律)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的，且具有相同的数学期望 μ，则对于任意的 $\varepsilon > 0$，都有

$$\lim_{n \to \infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right) = 1$$

♡

若在伯努利大数定律和辛钦大数定律的基础上不要求 $X_1, X_2, \cdots, X_n, \cdots$ 同分布，可得到更为一般的**切比雪夫大数定律**。

定理 4.8 (切比雪夫大数定律)

设随机变量 $X_1, X_2, \ldots, X_n, \ldots$ 之间两两不相关，各变量的数学期望 $E(X_i)=\mu_i$ 与方差 $D(X_i)=\sigma_i^2$ 存在，且 $D(X_i)$ 有共同的有限上界 M，则对于任意的 $\varepsilon>0$，都有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}\mu_i\right|<\varepsilon\right)=1$$

注 特别地，若所有变量的数学期望均为 μ，则此时切比雪夫大数定律即为辛钦大数定律。

概率论中有关阐述独立随机变量和的极限分布是正态分布的定理称为**中心极限定理**。

定理 4.9 (列维–林德伯格中心极限定理)

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的，其中有

$$E(X_k)=\mu,\ D(X_k)=\sigma^2 \qquad k=1,2,\cdots$$

则对于任意 x 都有

$$\lim_{n\to\infty} P\left(\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sigma\sqrt{n}}\leqslant x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\Phi(x)$$

注 当 n 充分大时，独立同分布的随机变量之和 $\sum\limits_{k=1}^{n}X_k$ 近似服从正态分布 $N(n\mu, n\sigma^2)$，即

$$\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1)$$

若记 $\overline{X}=\frac{1}{n}\sum\limits_{k=1}^{n}X_k$，则上式等价于 $\overline{X} \overset{\text{近似地}}{\sim} N\left(\mu,\frac{\sigma^2}{n}\right)$。

若令列维–林德伯格中心极限定理中的 X_i 服从 0-1 分布，则 $Y_n=\sum\limits_{k=1}^{n}X_k$ 的极限分布是正态分布，由此可得到下面的**棣莫弗–拉普拉斯定理**。

定理 4.10 (棣莫弗–拉普拉斯定理)

设 $Y_n \sim B(n,p)$，其中 $n=1,2,\cdots$，则对于任意的 x，都有

$$\lim_{n\to\infty} P\left(\frac{Y_n-np}{\sqrt{np(1-p)}}\leqslant x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\Phi(x)$$

注 该定理表明，正态分布是二项分布的极限分布，当 n 充分大时，Y_n 近似服从 $N(np, np(1-p))$。

设随机变量 $X\sim B(n,p)$，且在某些特定背景下 X 只能取整数值，若在 n 相对较大且 np 也相对较大时需要求解 $P(X=k)$ 的值，可以根据棣莫弗–拉普拉斯定理，以正态分布近似代替二项分布进行计算，即

$$\begin{aligned}P(X=k)=P\left(k-\frac{1}{2}<X<k+\frac{1}{2}\right)&=P\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}<X<\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)\\&=\Phi\left(\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)\end{aligned}$$

例题 4.6 设有随机变量 X，$E(X)=7300$，$\sqrt{D(X)}=700$，求 $P(5200<X<9400)$ 的估计值。

解 由题意有 $E(X)=7300$，$D(X)=490000$，取 $\varepsilon=7300-5200=9400-7300=2100$，由切比雪夫不等式有

$$P(5200<X<9400)=P(|X-7300|<2100)\geqslant 1-\frac{490000}{2100^2}=\frac{8}{9}$$

例题 4.7 某商店每天接待 100 位顾客，其中每人的消费额相互独立且均在区间 $[0,60]$（单位：元）上服从均匀分布，求该超市日收益超过 3500 元的概率。

解 设 X_i 为第 i 位顾客的消费额，则 $E(X_i)=30$，$D(X_i)=300$，$\dfrac{\sum\limits_{i=1}^{100}X_i-3000}{100\sqrt{3}}\overset{\text{近似地}}{\sim}N(0,1)$，则该超市日收益超过 3500 元的概率为

$$\begin{aligned}P\left(\sum_{i=1}^{100}X_i>3500\right)=1-P\left(\sum_{i=1}^{100}X_i\leqslant 3500\right)&=1-P\left(\frac{\sum\limits_{i=1}^{100}X_i-3000}{100\sqrt{3}}\leqslant\frac{3500-3000}{100\sqrt{3}}\right)\\&=1-\Phi(2.887)\\&=0.002\end{aligned}$$

第 4 章 练习

1. 某商店采取“先用后付”的营销策略，即一台新家电先交付顾客使用，待其出故障后由顾客带回维修并付款。设家电的使用寿命为 X 年，且 $X\sim E(0.1)$，当 $X\leqslant 1$ 时顾客需支付 1500 元，当 $1<X\leqslant 2$ 时顾客需支付 2000 元，当 $2<X\leqslant 3$ 时顾客需支付 2500 元，当 $x>3$ 时顾客需支付 3000 元，求该商店售出一台家电所得收益的数学期望。

   解 设 Y 为商店售出一台家电所得收益，由题意有

$$P(X\leqslant 1)=\int_0^1\frac{1}{10}\mathrm{e}^{-\frac{x}{10}}\mathrm{d}x=0.0952$$

   同理可得 $P(1<X\leqslant 2)=0.0861$，$P(2<X\leqslant 3)=0.0779$，$P(X>3)=0.7408$，故得 Y 的分布列如表 4.3 所示，并由此求出数学期望为

$$E(Y)=1500\times 0.0952+2000\times 0.0861+2500\times 0.0779+3000\times 0.7408=2732.15$$

Y	1500	2000	2500	3000
P	0.0952	0.0861	0.0779	0.7408

表 4.3: 本题所求的 Y 的分布列

2. 某国对外出口某种货物，设该货物的需求量为 X 吨，且 $X\sim U[2000,4000]$，每出口一吨可收益三万元，但无法出口则每吨损失一万元，求为使收益最大而应出口货物的吨数。

   解 设 y 为出口货物吨数，Y 为出口收益，则可得

$$Y=g(X)=\begin{cases}3y, & X\geqslant y\\ 3X-(y-X), & X<y\end{cases}$$

   由题意有

$$f(x)=\begin{cases}\dfrac{1}{2000}, & 2000\leqslant x\leqslant 4000\\ 0, & \text{其他}\end{cases}$$

故 Y 的数学期望为

$$\begin{aligned} E(Y) = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x &= \int_{2000}^{4000} \frac{g(x)}{2000}\mathrm{d}x \\ &= \frac{1}{2000}\left[\int_{2000}^{y}(4x-y)\mathrm{d}x + \int_{y}^{4000} 3y\mathrm{d}x\right] \\ &= \frac{1}{1000}(-y^2+7000y-4\times 10^6) \end{aligned}$$

该式为关于 y 的二次函数，易得当 $y=3500$ 时 $E(Y)$ 最大，故应出口 3500 吨货物。

3. 设某连锁超市中某种货物每周的进货量为 X 件，需求量为 Y 件，X 与 Y 独立且均在区间 $[10,20]$ 上服从均匀分布。现售出一件该货物可获利 1000 元，但售出一件从其他超市调派来的货物只能获利 500 元，求该超市每周的平均利润。

解 设 Z 为每周的平均利润，则有

$$Z = g(X,Y) = \begin{cases} 1000Y, & Y \leqslant X \\ 1000X + 500(Y-X), & Y > X \end{cases}$$

由于 X 与 Y 均在区间 $[10,20]$ 上服从均匀分布，则有

$$f_X(x) = \begin{cases} \frac{1}{10}, & 10 \leqslant x \leqslant 20 \\ 0, & \text{其他} \end{cases}, \quad f_Y(y) = \begin{cases} \frac{1}{10}, & 10 \leqslant y \leqslant 20 \\ 0, & \text{其他} \end{cases}$$

由于 X 与 Y 独立，则有

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{100}, & 10 \leqslant x \leqslant 20, 10 \leqslant y \leqslant 20 \\ 0, & \text{其他} \end{cases}$$

故有

$$\begin{aligned} E(Z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y &= \frac{1}{100}\int_{10}^{20}\int_{10}^{20} g(x,y)\mathrm{d}x\mathrm{d}y \\ &= \frac{1}{100}\int_{10}^{20}\mathrm{d}y\int_{10}^{20} g(x,y)\mathrm{d}x \\ &= \frac{1}{100}\int_{10}^{20}\mathrm{d}y\left[\int_{10}^{y} 500(x+y)\mathrm{d}x + \int_{y}^{20} 1000y\mathrm{d}x\right] \\ &\approx 14166.67 \end{aligned}$$

即该超市每周的平均利润为 14166.67 元。

4. 随机变量 (X,Y) 的联合分布列如下表所示，求 $E(X-Y)$ 与 $E(XY)$ 的值。

X \ Y	1	2	3
0	0.1	0.2	0.3
1	0.2	0.1	0.1

解 由题意可算得 $E(X)=0.4$，$E(Y)=2.1$，故可得 $E(X-Y)=E(X)-E(Y)=-1.7$；但由联合概率分布算得 $P(X=0,Y=1)\neq P(X=0)P(Y=1)$，$X$ 与 Y 不独立，不满足数学期望的性质，故有

$$E(XY) = 1\times 0.2 + 2\times 0.1 + 3\times 0.1 = 0.7$$

5. 设在 N 件产品中有 M 件次品，从中任取 n 件，其中 $n \leqslant M \leqslant N$，求取到的次品数的数学期望。

解 给每件次品依次编号为 $1,2,\ldots,M$，并设随机变量 X_i（$i=1,2,\cdots,M$），其满足

$$X_i = \begin{cases} 1, & \text{取到编号 } i \text{ 的次品} \\ 0, & \text{没有取到编号为 } i \text{ 的次品} \end{cases}$$

再设随机变量 X 代表取到的次品数，则有 $X = \sum_{i=1}^{M} X_i$；

由于有 $P(X_i = 1) = \dfrac{C_1^1 C_{N-1}^{n-1}}{C_N^n} = \dfrac{n}{N}$，可得 $E(X_i) = 1 \cdot \dfrac{n}{N} + 0 \cdot \left(1 - \dfrac{n}{N}\right) = \dfrac{n}{N}$，故有

$$E(X) = E\left(\sum_{i=1}^{M} X_i\right) = \sum_{i=1}^{M} E(X_i) = \frac{n}{N} M$$

6. 设随机变量 X 的概率密度函数为

$$f(x) = \begin{cases} ax^2 + bx + c, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

且 $E(X) = 0.5$，$D(X) = 0.15$，求 a,b,c 的值。

解 由题意与概率密度函数的性质有

$$\begin{cases} \displaystyle\int_0^1 (ax^2 + bx + c)\mathrm{d}x = 1 \\ E(X) = \displaystyle\int_0^1 x(ax^2 + bx + c)\mathrm{d}x = \frac{1}{2} \\ E(X^2) = \displaystyle\int_0^1 x^2(ax^2 + bx + c)\mathrm{d}x = \frac{a}{5} + \frac{b}{4} + \frac{c}{3} \end{cases}$$

联立解得

$$\begin{cases} a = 12 \\ b = -12 \\ c = 3 \end{cases}$$

7. 设有随机变量 X，$E(X) = \mu$，$D(X) = \sigma^2$，求 $P(\,|X - \mu| \leqslant 3\sigma\,)$ 的估计值。

解 设 $\varepsilon = 3\sigma$，则由切比雪夫不等式有

$$P(\,|X - \mu| \leqslant 3\sigma\,) \leqslant \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9}$$

8. 设在某险种中每位参保人的死亡率为 0.005，现有 10000 人参保，求死亡人数不超过 70 人的概率。

解 设 X 为死亡人数，易得 $X \sim B(10000, 0.005)$，考虑到 n 相对较大且 np 也相对较大，故根据棣莫弗–拉普拉斯定理，以正态分布近似代替二项分布进行计算，则死亡人数不超过 70 人的概率为

$$\begin{aligned} P(X \leqslant 70) &= P\left(\frac{X - np}{\sqrt{np(1-p)}} \leqslant \frac{70 - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi(2.84) \\ &= 0.9977 \end{aligned}$$

9. 某人训练射击，命中 10 环的概率为 0.5，命中 9 环的概率为 0.3，命中 8 环的概率为 0.2，求其射击 100 次后，总命中环数在 915 环至 945 环的概率。

解 设 X_i 为第 i 次命中的环数，则有 $E(X_i) = 9.3$，$D(X_i) = 0.61$，$\dfrac{\sum\limits_{i=1}^{100} X_i - 930}{\sqrt{61}} \underset{\sim\sim\sim}{\text{近似地}} N(0,1)$，设 $Y = \dfrac{\sum\limits_{i=1}^{100} X_i - 930}{\sqrt{61}}$，则总命中环数在 915 环至 945 环的概率为

$$
\begin{aligned}
P(915 \leqslant \sum_{i=1}^{100} X_i \leqslant 945) &= P\left(\frac{915-930}{\sqrt{61}} \leqslant Y \leqslant \frac{945-930}{\sqrt{61}}\right) \\
&= P(-1.92 \leqslant Y \leqslant 1.92) \\
&= 2\Phi(1.92) - 1 \\
&= 0.9452
\end{aligned}
$$

第 5 章 数理统计的基本知识

5.1 总体与样本

在数理统计中，我们将研究对象的全体称为**总体**，构成总体的每个成员称为**个体**。含有有限个个体的总体称为**有限总体**，含有无限多个个体的总体称为**无限总体**。通常，我们可以用随机变量 X 描述总体，将其称为**总体** X，并定义 X 的分布为**总体分布**。若被研究对象的数量指标为 $k \geqslant 2$ 个，则将该总体称为 k **维总体**。

从总体中抽取的待测个体组成的集合称为**样本**，样本所含的个体数目称为**样本容量**。从总体 X 中抽取的容量为 n 的样本常记作 $X_1, X_2, \cdots, X_n$，其所有可能取值的全体称为**样本空间**，而一组样本观测值 $x_1, x_2, \cdots, x_n$ 称为样本空间的一个**点**或**元素**。一般地，我们称具有以下性质的样本为**简单随机样本**：

1. **独立性**，即 $X_1, X_2, \cdots, X_n$ 为相互独立的随机变量；
2. **代表性**，即 $X_1, X_2, \cdots, X_n$ 与总体 X 具有相同的分布。

性质 简单随机样本 $X_1, X_n, \cdots, X_n$ 具有以下性质。

1. 若总体 X 为离散型随机变量，则有 $P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod\limits_{k=1}^{n} P(X_k = x_k)$。
2. 若总体 X 的分布函数为 $F(x)$，则样本 $X_1, X_2, \cdots, X_n$ 的联合分布函数为 $F(x_1, x_2, \cdots, x_n) = \prod\limits_{k=1}^{n} F(x_k)$。
3. 若总体 X 的密度函数为 $f(x)$，则样本 $X_1, X_2, \cdots, X_n$ 的联合密度函数为 $f(x_1, x_2, \cdots, x_n) = \prod\limits_{k=1}^{n} f(x_k)$。
4. 若总体的均值为 μ，方差为 σ^2，则有

$$E(X_k) = \mu,\ D(X_k) = \sigma^2, \qquad k = 1, 2, \cdots, n$$

例题 5.1 已知总体 X 服从 0-1 分布，且 $P(X = 1) = p$，求 $P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$ 的值。

解 由题意有 $P(X = x) = p^x(1-p)^{1-x}$（$x = 0, 1$），故有

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{k=1}^{n} P(X_k = x_k) = p^{x_1+x_2+\cdots+x_n}(1-p)^{n-(x_1+x_2+\cdots+x_n)}$$

5.2 统计量

定义 5.1 (统计量)

设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一组样本，$T(X_1, X_2, \cdots, X_n)$ 为一实值函数，且不含任何未知参数，称 $T = T(X_1, X_2, \cdots, X_n)$ 为一统计量。♣

由定义可看出，统计量仍为随机变量，当给定观测值 $x_1, x_2, \cdots, x_n$ 后，统计量的取值便完全确定，此时称 $T(x_1, x_2 \cdots, x_n)$ 为 $T(X_1, X_2, \cdots, X_n)$ 的观测值。

设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一组样本，$x_1, x_2, \cdots, x_n$ 为样本观测值。理论上，我们可以构造出无穷多个统计量，但现实中较为常用的统计量为以下几个。

1. **样本均值**：$\bar{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$。
2. **样本方差**：通常指**修正的样本方差**，即 $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 = \dfrac{1}{n-1}\left(\sum\limits_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$。

笔记 未修正的样本方差为 $S_0^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$，且有 $S^2=\frac{n}{n-1}S_0^2$。

3. **样本标准差**：$S=\sqrt{S^2}$。
4. **样本 k 阶原点矩**：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，其中 $k=1,2,\cdots$。
5. **样本 k 阶中心矩**：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$，其中 $k=1,2,\cdots$。

笔记 一阶原点矩即为均值，即 $A_1=\bar{X}$；二阶中心矩即为未修正的样本方差，即 $B_2=S_0^2=\frac{n-1}{n}S^2$。

定义 5.2 (顺序统计量)

将样本观测值 $x_1,x_2,\cdots,x_n$ 按由小到大的次序重新排列为 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，并定义 $X_{(k)}=x_{(k)}$ $(k=1,2,\cdots,n)$，由此得到的统计量 $X_{(1)},X_{(2)},\cdots,X_{(n)}$ 称为样本 $X_1,X_2,\cdots,X_n$ 的顺序统计量，其中 $X_{(1)}$ 称为最小顺序统计量，$X_{(n)}$ 称为最大顺序统计量。

设 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$ 为来自二维总体 (X,Y) 的一组样本，则常用的统计量有

1. **样本协方差**，即 $S_{XY}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$；
2. **样本相关系数**，即 $\rho_{XY}=\frac{S_{XY}^2}{S_XS_Y}$，其中 $S_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，$S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$。

定理 5.1 (样本均值与样本方差的性质)

设总体 X 的均值为 $E(X)=\mu$，方差为 $D(X)=\sigma^2$，样本 $(X_1,X_2,\cdots,X_n)$ 来自总体 X，则有

1. 样本均值为 $E(\bar{X})=\mu$；
2. 样本方差为 $D(\bar{X})=\frac{1}{n}\sigma^2$；
3. 样本方差的均值为 $E(S^2)=\sigma^2$。
4. 未修正的样本方差的均值为 $E(S_0^2)=\frac{n-1}{n}\sigma^2$。

5.3 抽样分布

统计量的分布称为**抽样分布**。鉴于**正态总体**是最常见的一种总体，故下面给出与正态分布有关的三种分布。

定义 5.3 (χ^2 分布)

设随机变量 $X_1,X_2,\cdots,X_n$ 相互独立，且 $X_i\sim N(0,1)$，其中 $1\leqslant i\leqslant n$，则称随机变量 $\chi^2=\sum_{i=1}^{n}X_i^2$ 为服从自由度为 n 的 χ^2 分布（卡方分布），记作 $\chi^2(n)$。

性质 χ^2 分布具有以下性质。

1. $E[\chi^2(n)]=n$，$D[\chi^2(n)]=2n$。
2. 设 $X\sim\chi^2(n)$，则由中心极限定理有

$$\lim_{n\to+\infty}\frac{X-n}{\sqrt{2n}}\underbrace{\text{近似地}}\,N(0,1)$$

即当 n 较大时，χ^2 分布可由正态分布近似代替。

3. 设 $X_1, X_2, \cdots, X_n$ 是相互独立的，且有 $X_i \sim \chi^2(m_i)$，则有 $\sum\limits_{i=1}^{n} X_i \sim \chi^2\left(\sum\limits_{i=1}^{n} m_i\right)$。

$\chi^2(n)$ 的密度函数为

$$f_{\chi^2}(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$$

其中 $\Gamma(x) = \displaystyle\int_0^{+\infty} t^{x-1}\mathrm{e}^{-t}\mathrm{d}t$ 在 $x > 0$ 时收敛，且当 $n \in \mathbf{N}$ 时有 $\Gamma(n+1) = n!$ 成立。由该函数不难看出，$\chi^2(2)$ 即为 $\lambda = \dfrac{1}{2}$ 的指数分布。

性质 $f_{\chi^2}(x)$ 具有以下性质。

1. $f_{\chi^2}(x)$ 在 $n = 1$ 与和 $n = 2$ 时为递减函数，而在 $n \geqslant 3$ 时为先增后减函数（单峰函数）。
2. 当 $n = 2$ 时，$f_{\chi^2}(x)$ 能够取到最大值。
3. 随着 n 的增大，$f_{\chi^2}(x)$ 的峰向 x 轴正半轴方向移动，且曲线更加对称。

χ^2 分布的上 α 分位点 $\chi^2_\alpha(n)$ 满足 $P(\chi^2 > \chi^2_\alpha(n)) = \alpha$，且当 $n \geqslant 40$ 时有 $\chi^2_\alpha(n) \approx \dfrac{1}{2}\left(u_\alpha + \sqrt{2n-1}\right)^2$。

例题 5.2 设相互独立的随机变量 $X_1, X_2, \cdots, X_6$ 均服从正态分布 $N(0,4)$，求 $P\left(\sum\limits_{i=1}^{6} X_i^2 > 6.54\right)$ 的值。

解 由题意有 $X_i \sim N(0, 2^2)$，则 $\dfrac{X_i}{2} \sim N(0,1)$，且可得 $\sum\limits_{i=1}^{6}\left(\dfrac{X_i}{2}\right)^2 \sim \chi^2(6)$，故有

$$P\left(\sum_{i=1}^{6} X_i^2 > 6.54\right) = P\left(\frac{\sum\limits_{i=1}^{6} X_i^2}{4} > \frac{6.54}{4}\right) = P(\chi^2(6) > 1.635)$$

经查 χ^2 分布表可得结果为 0.95。

定义 5.4 (t 分布)

设随机变量 X 与 Y 相互独立，且 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，则称随机变量 $t = \dfrac{X}{\sqrt{Y/n}}$ 为服从自由度为 n 的 t 分布，记作 $t(n)$。♣

性质 t 分布具有以下性质。

1. $E[\,t(n)\,] = 0$（$n > 1$），$D[\,t(n)\,] = \dfrac{n}{n-2}$（$n > 2$）。
2. t 分布的密度函数 $f_t(x)$ 为偶函数。
3. t 分布在 $n \to +\infty$ 时的极限分布为标准正态分布，通常在 $n \geqslant 30$ 时即可将 t 分布视为标准正态分布。

t 分布的上 α 分位点 $t_\alpha(n)$ 满足 $P(T > t_\alpha(n)) = \alpha$，且有 $t_{1-\alpha}(n) = -t_\alpha(n)$。

例题 5.3 设有相互独立的随机变量 X 与 Y_1, Y_2, Y_3, Y_4，其中 $X \sim N(2,1)$，$Y_i \sim N(0,4)$。现令

$$T = \frac{4(X-2)}{\sqrt{\sum\limits_{i=1}^{4} Y_i^2}}$$

并已知 $P(\,|\,T\,| > t_0\,) = 0.01$，求 t_0 的值。

解 由题意有 $\dfrac{X-2}{1} \sim N(0,1)$，$\dfrac{Y_i - 0}{2} \sim N(0,1)$，则可得 $\sum\limits_{i=1}^{4}\left(\dfrac{Y_i}{2}\right)^2 \sim \chi^2(4)$。根据卡方分布与相互独立的随机

变量的性质，有

$$\frac{(X-2)/1}{\sqrt{\sum_{i=1}^{4}\left(\frac{Y_i}{2}\right)^2\Big/4}}=\frac{4(X-2)}{\sqrt{\sum_{i=1}^{4}Y_i^2}}=T\sim t(4)$$

而 $P(|T|>t_0)=0.01$，根据 t 分布密度函数的性质有 $P(T>t_0)=0.005$，经查表有 $t=4.604$。

定义 5.5 (F 分布)

设随机变量 X 与 Y 相互独立，且 $X\sim\chi^2(m)$，$Y\sim\chi^2(n)$，则称随机变量 $F=\dfrac{X/m}{Y/n}$ 为服从自由度为 (m,n) 的 F 分布，记作 $F(m,n)$，其中 m,n 分别称为第一自由度和第二自由度。♣

性质 F 分布具有以下性质。

1. $E[F(m,n)]=\dfrac{n}{n-2}$ $(n>2)$，$D[F(m,n)]=\dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n>4)$。
2. 若 $F\sim F(m,n)$，则 $\dfrac{1}{F}\sim F(n,m)$。

F 分布的上 α 分位点 $F_\alpha(m,n)$ 满足 $P(F>F_\alpha(m,n))=\alpha$，且有 $F_{1-\alpha}(m,n)=\dfrac{1}{F_\alpha(n,m)}$。

例题 5.4 已知随机变量 $F\sim F(10,15)$，$P(F>\lambda_1)=0.01$，$P(F\leqslant\lambda_2)=0.01$，求 λ_1 和 λ_2 的值。

解 由题意有 $\lambda_1=F_{0.01}(10,15)$，且 $P(F\leqslant\lambda_2)=P\left(\dfrac{1}{F}\geqslant\dfrac{1}{\lambda_2}\right)=0.01$，其中 $\dfrac{1}{F}\sim F(15,10)$，故据此查表可得 $\lambda_1=3.8$，$\lambda_2=0.2193$。

最常使用的顺序统计量为 $X_{(1)}$ 与 $X_{(n)}$，下面的定理给出了 $X_{(1)}$、$X_{(n)}$ 及 $(X_{(1)},X_{(n)})$ 的分布。

定理 5.2 (顺序统计量的分布)

设总体 X 具有分布函数 $F(x)$，密度函数为 $f(x)$，$X_{1,2},\cdots,X_n$ 是来自总体 X 的一组样本，则

1. $X_{(1)}$ 的密度函数为 $f_1(x)=nf(x)[1-F(x)]^{n-1}$；
2. $X_{(n)}$ 的密度函数为 $f_n(x)=nf(x)[F(x)]^{n-1}$；
3. $(X_{(1)},X_{(n)})$ 的联合密度函数为

$$f(x,y)=\begin{cases}n(n-1)f(x)f(y)[F(y)-F(x)]^{n-2}, & x\leqslant y\\ 0, & x>y\end{cases}$$

♡

5.4 正态总体的抽样分布

定理 5.3 (单个正态总体条件下的抽样分布)

设 $X_1,X_2,\cdots,X_n$ 是来自正态总体 $X\sim N(\mu,\sigma^2)$ 的一组样本，$\bar{X},S^2$ 分别为样本均值与样本方差，则有

1. $\bar{X}$ 与 S^2 相互独立。
2. $\bar{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$；
3. $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$；
4. $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$。

♡

定理 5.4 (两个正态总体条件下的抽样分布)

设有两组样本相互独立，其中 $X_1, X_2, \cdots, X_m$ 是来自正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 的样本，$Y_1, Y_2, \cdots, Y_n$ 是来自正态总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的样本，$\bar{X}, S_X^2$ 与 $\bar{Y}, S_Y^2$ 分别为两组样本的样本均值与样本方差，则有

1. $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$；
2. $\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(m-1, n-1)$；
3. 当 $\sigma_1^2 = \sigma_2^2$ 时，$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}} \sim t(m+n-2)$；
4. 当 $m = n$ 时，$\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{S_X^2+S_Y^2-2S_{XY}^2}/\sqrt{n}} \sim t(n-1)$。

♡

例题 5.5 已知总体 $X \sim N(\mu, \sigma^2)$，$\bar{X}, S^2$ 分别为 X 的样本均值与样本方差，样本容量为 16，求 k 的值，使得 $P(\bar{X} > \mu + kS) = 0.95$。

解 由题意有 $n = 16$，且 $\dfrac{\bar{X}-\mu}{S}\sqrt{n} = \dfrac{4(\bar{X}-\mu)}{S} \sim t(15)$，故 $P(\bar{X} > \mu + kS) = P\left(\dfrac{4(\bar{X}-\mu)}{S} > 4k\right) = 0.95$，经查表有 $k = -0.438$。

例题 5.6 已知总体 $X \sim N(\mu, \sigma^2)$，$(X_1, X_2, \cdots, X_{n+1})$ 为一组样本，$\bar{X}_n, S_n^2$ 分别为样本 $(X_1, X_2, \cdots, X_n)$ 的样本均值与样本方差，求 $\dfrac{X_{n+1}-\bar{X}}{S_n}\sqrt{\dfrac{n}{n+1}}$ 的分布。

解 由题意有 $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$，$X_{n+1} \sim N(\mu, \sigma^2)$，则可得 $\dfrac{X_{n+1}-\bar{X}}{\sigma\sqrt{\dfrac{n+1}{n}}} \sim N(0,1)$，并由 $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$ 有

$$\frac{\dfrac{X_{n+1}-\bar{X}}{\sigma\sqrt{\dfrac{n+1}{n}}}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2}\Big/(n-1)}} = \frac{X_{n+1}-\bar{X}}{S_n}\sqrt{\frac{n}{n+1}} \sim t(n-1)$$

第 6 章 参数估计

6.1 点估计

参数估计是根据样本对总体未知参数的取值进行估计的一种统计推断方法。在本节中，我们假定 θ 是总体的待估参数，并用 Θ 表示**参数空间**，即 θ 的取值范围。

参数估计有以下两种形式。

1. **点估计**：根据样本构造一个统计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$，并用 $\hat{\theta}$ 估计 θ，其中 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 称为 θ 的**估计量**，$\hat{\theta}(x_1, x_2, \cdots, x_n)$ 称为 θ 的**估计值**。
2. **区间估计**：根据样本构造两个统计量 $\hat{\theta_1}$ 和 $\hat{\theta_2}$，其中 $\hat{\theta_1} < \hat{\theta_2}$，并用区间 $[\hat{\theta_1}, \hat{\theta_2}]$ 估计 θ。

点估计有两种常用的方法，一种是**矩估计法**，另一种是**极大似然估计法**，下面将逐一介绍。

6.1.1 矩估计法

矩估计法是用样本的 k 阶原点矩作为总体的 k 阶原点矩的估计，进而求解未知参数的一种估计方法。在使用矩估计法时，并不需要知道总体的分布类型，因此该方法应用广泛。然而，有时候所要求的总体的某阶原点矩可能不存在，且求解方程困难，故矩估计法并不是万能的。

例题 6.1 设总体 $X \sim P(\lambda)$，$(X_1, X_2, \cdots, X_n)$ 是其一组样本，求 λ 的矩估计量。

解 当总体分布仅含一个参数时，只需用一阶原点矩估计即可，故有 $\hat{\lambda} = E(X) = \bar{X}$。

例题 6.2 设总体 X 的均值为 μ，方差为 σ^2，$(X_1, X_2, \cdots, X_n)$ 是其一组样本，求 μ 与 σ^2 的矩估计量。

解 由题意，我们可以用样本的一阶与二阶原点矩分别代替总体的一阶与二阶原点矩，即

$$\begin{cases} E(X) = \mu = A_1 = \bar{X} \\ E(X^2) = E^2(X) + D(X) = \mu^2 + \sigma^2 = A_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases}$$

求解可得 $\hat{\mu} = \bar{X}$，$\hat{\sigma^2} = A_2 - \hat{\mu} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 - \bar{X}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 = B_2$。

6.1.2 极大似然估计法

设总体 X 含有待估参数 $\theta = (\theta_1, \theta_2, \cdots, \theta_m)$，参数空间为 Θ，若能在 Θ 中取得一个 $\hat{\theta}$，使得 $\theta = \hat{\theta}$ 时，事件 $\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$ 出现的概率 $L(\theta)$ 达到最大值，则称 $\hat{\theta}$ 为 θ 的极大似然估计，其中 $L(\theta)$ 称为样本的**似然函数**。

例题 6.3 设总体 $X \sim P(\lambda)$，$(x_1, x_2, \cdots, x_n)$ 是其一组样本观测值，求 λ 的极大似然估计。

解 由题意可知总体的概率函数为 $P(X = k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$（$k = 0, 1, 2, \cdots$），则 λ 的似然函数为

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}\mathrm{e}^{-\lambda} = \frac{\lambda^{x_1 + x_2 + \cdots + x_n}}{\prod\limits_{i=1}^{n} x_i!}\mathrm{e}^{-n\lambda}$$

对两边取对数有

$$\ln L(\lambda) = -\ln\prod_{i=1}^{n} x_i! + \sum_{i=1}^{n} x_i \ln\lambda - n\lambda$$

两边对 λ 求导有

$$\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda}=\frac{x_1+x_2+\cdots+x_n}{\lambda}-n$$

令上式的值为 0，可得 $\lambda=\dfrac{x_1+x_2+\cdots+x_n}{n}$，即 $\hat{\lambda}=\bar{X}$。

例题 6.4 设总体 $X\sim E(\lambda)$，$(x_1,x_2,\cdots,x_n)$ 是其一组样本观测值，求 λ 的极大似然估计。

解 由题意可知总体的密度函数为

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

则 λ 的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\lambda\mathrm{e}^{-\lambda x_i}=\lambda^n\mathrm{e}^{-\lambda(x_1+x_2+\cdots+x_n)}$$

对两边取对数有

$$\ln L(\lambda)=n\ln\lambda-\lambda(x_1+x_2+\cdots+x_n)$$

两边对 λ 求导有

$$\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda}=\frac{n}{\lambda}-(x_1+x_2+\cdots+x_n)$$

令上式的值为 0，可得 $\lambda=\dfrac{n}{x_1+x_2+\cdots+x_n}$，即 $\hat{\lambda}=\dfrac{1}{\bar{X}}$。

例题 6.5 设总体 $X\sim N(\mu,\sigma^2)$，$(x_1,x_2,\cdots,x_n)$ 是其一组样本观测值，求 μ 与 σ^2 的极大似然估计。

解 由题意可知总体的密度函数为 $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$，则 μ 与 σ^2 的似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}}\right)^n(\sigma^2)^{-\frac{n}{2}}\mathrm{e}^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}}$$

对两边取对数有

$$\ln L(\mu,\sigma^2)=n\ln\frac{1}{\sqrt{2\pi}}-\frac{n}{2}\ln\sigma^2-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}$$

两边分别对 μ 与 σ^2 求偏导有

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\frac{x_1+x_2+\cdots+x_n-n\mu}{\sigma^2}$$

$$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^4}$$

分别令上两式的值为 0，可得 $\mu=\dfrac{x_1+x_2+\cdots+x_n}{n}$，$\sigma^2=\dfrac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{n}$，即 $\hat{\mu}=\bar{X}$，$\hat{\sigma^2}=B_2$。

例题 6.6 设总体 $X\sim U[\theta_1,\theta_2]$，$(x_1,x_2,\cdots,x_n)$ 是其一组样本观测值，求 θ_1 与 θ_2 的极大似然估计。

解 由题意可得 θ_1 与 θ_2 的似然函数为

$$L(\theta_1,\theta_2)=\prod_{i=1}^{n}\frac{1}{\theta_2-\theta_1}=\frac{1}{(\theta_2-\theta_1)^n}$$

要使 $L(\theta_1,\theta_2)$ 取到最大值，则需令区间 $[\theta_1,\theta_2]$ 的长度最小且包含所有的观测值，故 $\hat{\theta_1}=\min\{x_1,x_2,\cdots,x_n\}$，$\hat{\theta_2}=\max\{x_1,x_2,\cdots,x_n\}$。

6.1.3 点估计的优良性准则

定义 6.1 (估计量的无偏性)

设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是未知参数 θ 的估计量，若 $E(\hat{\theta}) = \theta$，则称 $\hat{\theta}$ 为 θ 的无偏估计，否则称为有偏估计。若 $\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$，则称 $\hat{\theta}$ 为 θ 的渐进无偏估计。♣

定理 6.1 (样本均值与样本方差的无偏性)

设总体 X 的均值为 μ，方差为 σ^2，$(X_1, X_2, \cdots, X_n)$ 是其一组样本，则有

1. 样本均值 $\bar{X}$ 是 μ 的无偏估计，即 $E(\bar{X}) = \mu$；
2. 样本方差 S^2 是 σ^2 的无偏估计，即 $E(S^2) = \sigma^2$；
3. 未修正的样本方差 S_0^2 是 σ^2 的有偏估计；
4. 估计量 $\hat{\sigma^2}$ 是 σ^2 的渐进无偏估计。

♡

笔记 若 $\hat{\theta}$ 是 θ 的无偏估计，则 $g(\hat{\theta})$ 不一定是 $g(\theta)$ 的无偏估计。

例题 6.7 求证：S 不是 σ 的无偏估计。

证明 由均值与方差的关系可得 $D(S) = E(S^2) - E^2(S) = \sigma^2 - E^2(S)$，则有 $E(S) = \sqrt{\sigma^2 - D(S)} \leqslant \sigma$，故 S 不是 σ 的无偏估计。

定义 6.2 (估计量的有效性)

设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 均为参数 θ 的无偏估计，若 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$，则称估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。♣

例题 6.8 现对于总体 X 及其一组样本 $(X_1, X_2, \cdots, X_n)$ 有 $\sum\limits_{i=1}^{n} a_i X_i$（$\sum\limits_{i=1}^{n} a_i = 1$）与 $\bar{X}$ 两种无偏估计，试通过计算说明何者更具有效性。

解 由样本均值的性质有 $D(\bar{X}) = \dfrac{\sigma^2}{n}$，$D\left(\sum\limits_{i=1}^{n} a_i X_i\right) = \sum\limits_{i=1}^{n} a_i^2 D(X_i) = \sigma^2 \sum\limits_{i=1}^{n} a_i^2$，并根据均值不等式可得

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \leqslant \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}$$

即 $1 = \left(\sum\limits_{i=1}^{n} a_i\right)^2 \leqslant n \sum\limits_{i=1}^{n} a_i^2$，则有 $D(\bar{X}) = \dfrac{\sigma^2}{n} \leqslant \sigma^2 \sum\limits_{i=1}^{n} a_i^2 = D\left(\sum\limits_{i=1}^{n} a_i X_i\right)$，故 $\bar{X}$ 比 $\sum\limits_{i=1}^{n} a_i X_i$ 有效。

6.2 区间估计

下面给出置信区间的定义。

定义 6.3 (置信区间)

设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 是两个统计量，且 $\hat{\theta}_1 < \hat{\theta}_2$，$\theta$ 是总体 X 的未知参数，若对于给定的 $\alpha \in (0,1)$，有

$$P(\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2) = 1 - \alpha$$

则称随机区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 是 θ 的一个区间估计或置信区间，其中 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 分别称为置信下限与置信上限，$1-\alpha$ 称为置信水平或置信度。♣

由定义可看出，置信区间的长度描述了估计的精确性：长度越小，则估计越精确；长度越大，则估计越粗略。除外，当样本容量固定时，置信水平要求越高，估计精确性越差。

求解参数 θ 的置信水平为 $1-\alpha$ 的置信区间的一般方法如下。

1. 寻找一个与要估计的参数 θ 有关的统计量 T，一般为 θ 的一良好点估计。
2. 寻找 T 和 θ 的某一函数 $H(T,\theta)$，要求 $H(T,\theta)$ 的分布 F 已知且与 θ 无关，其中 $H(T,\theta)$ 称为**枢轴变量**。
3. 寻找适当的常数 c,d，使得 $P(c \leqslant H(T,\theta) \leqslant d) = 1-\alpha$；在正态总体的条件下，通常取 c 为 F 的上 $\dfrac{\alpha}{2}$ 分位数，取 d 为 F 的上 $\left(1-\dfrac{\alpha}{2}\right)$ 分位数。
4. 将不等式 $c \leqslant H(T,\theta) \leqslant d$ 化为 $\hat{\theta}_1 \leqslant \theta \leqslant \hat{\theta}_2$，由此可得置信区间为 $[\hat{\theta}_1,\hat{\theta}_2]$。

单个正态总体的参数区间估计的计算公式如下所示。

待估参数	条件	枢轴变量及其分布	置信区间
均值 μ	方差 σ^2 已知	$\dfrac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$	$\left[\bar{X}-u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}, \bar{X}+u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}\right]$
	方差 σ^2 未知	$\dfrac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$	$\left[\bar{X}-t_{\frac{\alpha}{2}}(n-1)\dfrac{S}{\sqrt{n}}, \bar{X}+t_{\frac{\alpha}{2}}(n-1)\dfrac{S}{\sqrt{n}}\right]$
方差 σ^2	均值 μ 已知	$\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$	$\left[\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right]$
	均值 μ 未知	$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left[\dfrac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \dfrac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right]$

表 6.1: 单个正态总体的参数区间估计