概率论与数理统计

赵子轩

2023 年 3 月 30 日

目录

1 随机事件及其概率

1.1 随机试验

在一定条件下必然出现的现象叫做**必然现象**. 在相同的条件下，可能出现不同的结果，而在试验或观测之前不能预知确切结果的现象叫做**随机现象**.

随机现象具有随机性和统计规律性.

- 随机性：对随机现象进行观测时，不能预先确定其结果.
- 统计规律性：对随机现象进行大量重复观测后，其结果往往会表现出某种规律性.

为了研究和揭示随机现象的统计规律性，需要对随机现象进行大量重复的观察、测量或试验，统称为试验.

如果试验具有以下特点：

1. 可重复性：试验可以在相同条件下重复进行多次，甚至进行无限次；
2. 可观测性：每次试验的所有可能结果都是明确的、可以观测的，并且试验的可能结果有两个或两个以上；
3. 随机性：每次试验出现的结果是不确定的，在试验之前无法预先确定究竟会出现哪一个结果，

则称之为**随机试验**，简称为**试验**.

通常用字母 E 表示一个随机试验. 随机试验 E 的基本结果称为**样本点**，用 ω 表示. 随机试验 E 的所有基本结果的集合称为**样本空间**，用 $\Omega = \{\omega\}$ 表示.

1.2 随机事件

1.2.1 随机事件的概念

随机试验 E 的样本空间 $\Omega = \{\omega\}$ 的子集称为随机试验 E 的**随机事件**（random event），简称为**事件**（event），用大写字母 A, B, C 等表示.

设 $A \subseteq \Omega$，如果试验结果 $\omega \in A$，则称在这次试验中事件 A 发生；如果 $\omega \notin A$，则称事件 A 不发生.

由一个样本点 ω 组成的事件称为**基本事件**.

样本空间 Ω 本身也是 Ω 的子集，它包含 Ω 的所有样本点，在每次试验中 Ω 必然发生，称为**必然事件**.

空集 $\varnothing$ 也是 Ω 的子集，它不包含任何样本点，在每次试验中都不可能发生，称为**不可能事件**.

在一个样本空间中，如果只有有限个样本点，则称它为**有限样本空间**；如果有无限个样本点，则称它为**无限样本空间**.

1.2.2 随机事件的关系

1 事件的包含

如果当事件 A 发生时事件 B 一定发生，则称事件 B **包含**事件 A，记作 $A \subseteq B$.

对于任意事件 A，有 $\varnothing \subseteq A \subseteq \Omega$.

如果 $A \subseteq B, B \subseteq C$，则 $A \subseteq C$.

2 事件的相等

如果事件 A 和事件 B 相互包含，即 $A \subseteq B$ 且 $B \subseteq A$，则称事件 A 与事件 B **相等**，记作 $A = B$.

3 事件的互不相容

如果事件 A 和事件 B 在同一次试验中不能同时发生，则称事件 A 与事件 B 是**互不相容**的，或称事件 A 与事件 B 是**互斥**的.

任意两个基本事件一定互斥.

4 事件的互逆

如果在每一次试验中事件 A 和事件 B 必有一个且仅有一个发生，则称事件 A 与事件 B 是**互逆**的或**对立**的，称其中的一个事件是另一个事件的**逆事件**，记作 $\overline{A}=B$，或 $\overline{B}=A$.

对于任意事件 A，有 $\overline{\overline{A}}=A$.

如果事件 A 与事件 B 互逆，则事件 A 与事件 B 一定互斥. 反之，如果事件 A 与事件 B 互斥，事件 A 与事件 B 不一定互逆.

1.2.3 随机事件的运算

1 事件的并

如果事件 A 和事件 B 至少有一个发生，则这样的一个事件称为事件 A 与事件 B 的**并事件**或**和事件**，记作 $A\cup B$.

$$A\cup B=\{\omega\mid\omega\in A\text{ 或 }\omega\in B\}$$

事件 A 和事件 B 作为样本空间 Ω 的子集，并事件 $A\cup B$ 就是子集 A 与 B 的并集.

对于任何事件 A 与 B，事件的并运算有如下性质：

1. $A\cup A=A$
2. $A\cup\varnothing=A$
3. $A\cup B=B\cup A$
4. $A\cup\overline{A}=\Omega$
5. $A\subseteq A\cup B$
6. $B\subseteq A\cup B$
7. 如果 $A\subseteq B$，则有 $A\cup B=B$.

事件的并可以推广到多个事件的情形：

$$\bigcup_{i=1}^{n}A_i=\{\text{事件}A_1,A_2,\cdots,A_n\text{中至少有一个发生}\}$$

$$\bigcup_{i=1}^{\infty}A_i=\{\text{事件}A_1,A_2,\cdots,A_n,\cdots\text{中至少有一个发生}\}$$

2 事件的交

如果事件 A 和事件 B 同时发生，则这样的一个事件称为事件 A 与事件 B 的**交事件**或**积事件**，记作 $A \cap B$ 或 AB.

$$A \cap B = \{\omega \mid \omega \in A \text{ 且 } \omega \in B\}$$

事件 A 和事件 B 作为样本空间 Ω 的子集，交事件 $A \cap B$ 就是子集 A 与 B 的交集.

对于任何事件 A 与 B，事件的交运算有如下性质：

1. $A \cap A = A$
2. $A \cap \emptyset = \emptyset$
3. $A \cap B = B \cap A$
4. $A \cap \overline{A} = \emptyset$
5. $A \cap B \subseteq A$
6. $A \cap B \subseteq B$
7. 如果 $A \subseteq B$，则有 $A \cap B = A$.
8. 如果 A 与 B 互不相容，则有 $A \cap B = \emptyset$.

事件的交可以推广到多个事件的情形：

$$\bigcap_{i=1}^{n} A_i = \{\text{事件}A_1, A_2, \cdots, A_n\text{同时发生}\}$$

$$\bigcap_{i=1}^{\infty} A_i = \{\text{事件}A_1, A_2, \cdots, A_n, \cdots\text{同时发生}\}$$

3 事件的差

如果事件 A 发生而事件 B 不发生，则这样的一个事件称为事件 A 与事件 B 的**差事件**，记作 $A - B$.

$$A - B = \{\omega \mid \omega \in A \text{ 且 } \omega \notin B\}$$

对于任何事件 A 与 B，事件的差运算有如下性质：

1. $A - A = \emptyset$
2. $A - \emptyset = A$
3. $A - B = A - AB = A\overline{B}$
4. $\Omega - A = \overline{A}$
5. $A - \Omega = \emptyset$
6. $(A - B) \cup B = (B - A) \cup A = A \cup B$

7. $A \cup B = A \cup (B - AB) = B \cup (A - AB)$
8. $A - B, AB, B - A$ 两两互斥，且

（1）$A \cup B = (A - B) \cup AB \cup (B - A)$

（2）$A = (A - B) \cup AB$

（3）$B = (B - A) \cup AB$

4　随机事件的运算规律

1. 交换律：$A \cup B = B \cup A$，$AB = BA$.
2. 结合律：$(A \cup B) \cup C = A \cup (B \cup C)$，$(AB)C = A(BC)$.
3. 分配律：$A(B \cup C) = (AB) \cup (AC)$，$A \cup (BC) = (A \cup B)(A \cup C)$.
4. 对偶律：$\overline{A \cup B} = \overline{A}\,\overline{B}$，$\overline{AB} = \overline{A} \cup \overline{B}$.

对于多个随机事件，以上的运算规律也成立.

1.3　随机事件的概率

1.3.1　频率

定义 1.1　设在相同的条件下进行的 n 次试验中，事件 A 发生了 n_A 次，则称 n_A 为事件 A 发生的**频数**，称比值 $\dfrac{n_A}{n}$ 为事件 A 发生的**频率**，记作 $f_n(A)$，即

$$f_n(A) = \frac{n_A}{n}$$

事件 A 发生的频率反映了事件 A 在 n 次试验中发生的频繁程度. 频率越大，表明事件 A 的发生越频繁，从而可知事件 A 在一次试验中发生的可能性越大.

性质 1（非负性）　对于任意事件 A，有 $f_n(A) \geqslant 0$.

性质 2（规范性）　对于必然事件 Ω，有 $f_n(\Omega) = 1$.

性质 3（有限可加性） 对于两两互不相容的事件 $A_1, A_2, \cdots, A_m$（即当 $i \neq j$ 时，有 $A_iA_j = \varnothing$，$i, j = 1, 2, \cdots, m$），有

$$f_n(\bigcup_{i=1}^{m} A_i) = \sum_{i=1}^{m} f_n(A_i)$$

在相同的条件下重复进行 n 次试验，当 n 增大时，事件 A 发生的频率 $f_n(A)$ 呈现出稳定性，逐渐稳定于某一常数 p. 用这一常数表示事件 A 发生的可能性大小，称为事件 A 的概率，记为 $P(A)$，即 $P(A) = p$.

当 n 很大时，可以用频率 $f_n(A)$ 作为概率 $P(A)$ 的近似值.

1.3.2 概率

定义 1.2 设随机试验 E 的样本空间为 Ω，如果对于 E 的每一个事件 A，有唯一的实数 $P(A)$ 和它对应，并且这个事件的函数 $P(A)$ 满足以下条件：

1. 非负性：对于任意事件 A，有 $P(A) \geqslant 0$；
2. 规范性：对于必然事件 Ω，有 $P(\Omega) = 1$；
3. 可列可加性：对于两两互不相容的事件 $A_1, A_2, \cdots$，有

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

则称 $P(A)$ 为事件 A 的**概率**（probability）.

性质 1 对于不可能事件 $\varnothing$，有 $P(\varnothing) = 0$.

证明：因为 $\varnothing = \varnothing \cup \varnothing \cup \cdots$，根据概率的可列可加性，有

$$P(\varnothing) = P(\varnothing) + P(\varnothing) + \cdots$$

由概率的非负性知 $P(\varnothing) \geqslant 0$，因此 $P(\varnothing) = 0$. □

性质 2（有限可加性） 对于两两互不相容的事件 $A_1, A_2, \cdots, A_n$，有

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

证明：令 $A_i = \varnothing\ (i = n+1, n+2, \cdots)$，根据概率的可列可加性及性质 1，有

$$P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i)$$

□

性质 3 对于任一事件 A，有 $P(\overline{A}) = 1 - P(A)$.

证明：因为 $A \cup \overline{A} = \Omega$，且 $A\overline{A} = \varnothing$，由性质 2 及概率的规范性，得

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

即

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

□

性质 4 如果 $A \subseteq B$，则有 $P(B - A) = P(B) - P(A)$，$P(A) \leqslant P(B)$.

证明：因为 $A \subseteq B$，从而有 $B = A \cup (B - A)$，且 $A(B - A) = \varnothing$，由性质 2 可得

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

所以

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

由于 $P(B - A) \geqslant 0$，因此 $P(A) \leqslant P(B)$. □

性质 5 对于任一事件 A，有 $P(A) \leqslant 1$.

证明：因为 $A \subseteq \Omega$，由性质 4 及概率的规范性，可得

$$P(A) \leqslant P(\Omega) = 1$$

□

性质 6（概率的减法公式） 对于任意两个事件 A 与 B，有

$$P(B-A) = P(B) - P(AB) \tag{1-1}$$

证明：由于 $B-A = B-AB$，而 $AB \subseteq B$，根据性质 4 可得

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

□

性质 7 对于任意两个事件 A 与 B，有

$$P(A \cup B) = P(A) + P(B) - P(AB) \tag{1-2}$$

$$P(A \cup B) \leqslant P(A) + P(B)$$

证明：因为 $A \cup B = A \cup (B-AB)$，且 $A(B-AB) = \varnothing$, $AB \subseteq B$，由性质 2 及性质 4 可得

$$\begin{aligned} P(A \cup B) &= P(A \cup (B-AB)) \\ &= P(A) + P(B-AB) \\ &= P(A) + P(B) - P(AB) \end{aligned}$$

由于 $P(AB) \geqslant 0$，因此 $P(A \cup B) \leqslant P(A) + P(B)$. □

式 (1-2) 称为概率的**加法公式**.

加法公式可以推广到任意有限个事件的情形：设 $A_1, A_2, \cdots, A_n$ 是 n 个随机事件，则有

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) = &\sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i<j \leqslant n} P(A_i A_j) + \sum_{1 \leqslant i<j<k \leqslant n} P(A_i A_j A_k) + \cdots \\ &+ (-1)^{n-1} P(A_1 A_2 \cdots A_n) \end{aligned} \tag{1-3}$$

式 (1-3) 称为概率的**一般加法公式**.

【例 1.1】 设 A 和 B 是同一试验 E 的两个随机事件，证明

$$1 - P(\overline{A}) - P(\overline{B}) \leqslant P(AB) \leqslant P(A \cup B)$$

证明：因为 $AB \subseteq A \subseteq (A \cup B)$，所以

$$P(AB) \leqslant P(A \cup B)$$

由概率的性质 7、性质 3 及事件的对偶律，可得

$$P(\overline{A}) + P(\overline{B}) \geqslant P(\overline{A} \cup \overline{B}) = P(\overline{AB}) = 1 - P(AB)$$

因此

$$1 - P(\overline{A}) - P(\overline{B}) \leqslant P(AB)$$

□

1.3.3 古典概型

如果随机试验具有以下两个特点：

1. 试验的样本空间只包含有限个样本点；

2. 在试验中每个基本事件发生的可能性相同，

则称这种试验为**等可能概型**或**古典概型**（classic probability model）.

设试验 E 是古典概型,样本空间为 $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$,则基本事件 $\{\omega_1\}, \{\omega_2\}, \cdots, \{\omega_n\}$ 两两互不相容，且

$$\Omega = \{\omega_1\} \cup \{\omega_2\} \cup \cdots \cup \{\omega_n\}$$

由于 $P(\Omega)=1$ 及 $P(\{\omega_1\})=P(\{\omega_2\})=\cdots=P(\{\omega_n\})$，因此

$$P(\{\omega_1\})=P(\{\omega_2\})=\cdots=P(\{\omega_n\})=\frac{1}{n}$$

如果事件 A 包含 k 个基本事件，$A=\{\omega_{i_1}\}\cup\{\omega_{i_2}\}\cup\cdots\cup\{\omega_{i_k}\}$，其中 $i_1,i_2,\cdots,i_k$ 是 $1,2,\cdots,n$ 中某 k 个不同的数，则有

$$P(A)=P(\{\omega_{i_1}\})+P(\{\omega_{i_2}\})+\cdots+P(\{\omega_{i_k}\})=\frac{k}{n}$$

即

$$P(A)=\frac{A\,\text{包含的基本事件个数}}{\Omega\,\text{包含的基本事件总数}}$$

1.3.4 几何概型

如果随机试验是将一个点随机地投到某一区域 Ω 内，而这个随机点落在 Ω 中任意两个度量相等的子区域内的可能性是一样的，则称这样的试验属于**几何概型**（geometric probability model）.

注： Ω 可以是直线上的某一区间，也可以是平面或空间内的某一区域. 区域的度量是指直线上区间的长度，或者平面内区域的面积，或者空间内区域的体积.

对于任何有度量的子区域 $A\subseteq\Omega$，定义事件 $A=$“随机点落在区域 A 内” 的概率为

$$P(A)=\frac{A\,\text{的度量}}{\Omega\,\text{的度量}}$$

【例 1.2】(蒲丰投针问题) 在平面上画有等距离的平行线，平行线间的距离为 $2a\,(a>0)$. 向该平面任意投掷一枚长为 $2l\,(l<a)$ 的圆柱形的针，求此针与任一平行线相交的概率.

解：

针投在该平面上，设 x 为针的中点 M 到最近的一条平行线的距离，φ 为针与此直线的交角，如图 1.1 所示，则有 $0\leqslant x\leqslant a$, $0\leqslant\varphi\leqslant\pi$. 因此样本空间为

$$\Omega=\{(\varphi,x)\mid 0\leqslant\varphi\leqslant\pi,\,0\leqslant x\leqslant a\}$$

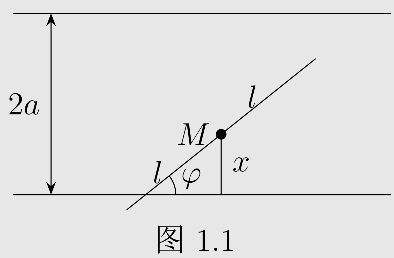


图 1.1

针与最近的一条平行线相交的充分必要条件是 $x \leqslant l\sin\varphi$. 设事件 $A =$ “针与最近的一条平行线相交”，则

$$A = \{(\varphi, x) \mid 0 \leqslant \varphi \leqslant \pi,\ 0 \leqslant x \leqslant l\sin\varphi\}$$

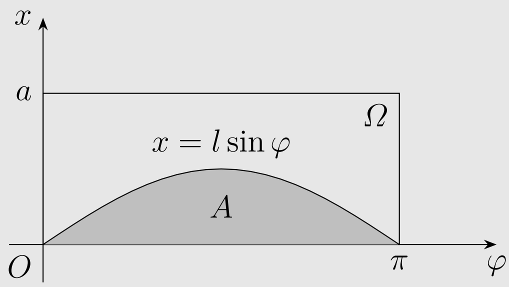


图 1.2

所求概率为

$$p = \frac{A\text{的面积}}{\Omega\text{的面积}} = \frac{\displaystyle\int_0^{\pi} l\sin\varphi\,\mathrm{d}\varphi}{\pi a} = \frac{2l}{\pi a} \tag{1-4}$$

蒲丰投针问题可以用来计算 π 的近似值. 如果投针 N 次，其中针与平行线相交 n 次，当 N 很大时，以频率 $\dfrac{n}{N}$ 作为概率 p 的近似值，代入式 (1-4) 可得

$$\pi \approx \frac{2lN}{an}$$

1.4 条件概率

1.4.1 条件概率与乘法公式

定义 1.3 设 A 和 B 是试验 E 的两个事件，且 $P(A) > 0$，称 $\dfrac{P(AB)}{P(A)}$ 为在事件 A 已经发生的条件下，事件 B 发生的**条件概率**（conditional probability），记为 $P(B \mid A)$，即

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

概率的**乘法公式**：对于任意两个事件 A 和 B，如果 $P(A) > 0$，则有

$$P(AB) = P(A)\,P(B \mid A) \tag{1-5}$$

同样可以在 $P(B) > 0$ 时，定义在事件 B 已经发生的条件下，事件 A 发生的条件概率为

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

在 $P(A)>0, P(B)>0$ 的条件下，有

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

条件概率具有如下性质：

性质 1（非负性） 对任意事件 B，有 $P(B\mid A)\geqslant 0$.

性质 2（规范性） 对于必然事件 Ω，有 $P(\Omega\mid A)=1$.

性质 3（可列可加性） 对于两两互不相容的事件 $B_1, B_2, \cdots$，有

$$P(\left(\bigcup_{i=1}^{\infty} B_i\right)\mid A)=\sum_{i=1}^{\infty} P(B_i\mid A)$$

可由条件概率的三个基本性质推导出其他性质，例如：

1. $P(\varnothing\mid A)=0$
2. $P(\overline{B}\mid A)=1-P(B\mid A)$
3. $P((B_1\cup B_2)\mid A)=P(B_1\mid A)+P(B_2\mid A)-P(B_1B_2\mid A)$

可以把乘法公式推广到有限个事件的交的情况：设 $A_1, A_2, \cdots, A_n$ 是同一试验的事件，且 $P(A_1A_2\cdots A_{n-1})>0$，则有

$$P(A_1A_2\cdots A_n)=P(A_1)\,P(A_2\mid A_1)\,P(A_3\mid A_1A_2)\cdots P(A_n\mid A_1A_2\cdots A_{n-1}) \tag{1-6}$$

1.4.2 全概率公式

设试验 E 的样本空间为 Ω，事件 $A_1, A_2, \cdots, A_n$ 两两互不相容，且 $\bigcup\limits_{i=1}^{n} A_i=\Omega$，则称 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个**分割**或**完全事件组**.

定理 1.1（全概率公式） 若事件 $A_1, A_2, \cdots, A_n$ 构成一个完全事件组，且 $P(A_i)>0\ (i=1,2,\cdots,n)$，则对任意事件 B，有

$$P(B)=\sum_{i=1}^{n} P(A_i)\,P(B\mid A_i) \tag{1-7}$$

证明：

$$B = B\Omega = B\left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n}(A_i B)$$

这里 $(A_iB)\cap(A_jB)=\varnothing\ (i\neq j,\ i,j=1,2,\cdots,n)$，由概率的有限可加性得

$$P(B) = P(\bigcup_{i=1}^{n}(A_iB)) = \sum_{i=1}^{n} P(A_iB)$$

由乘法公式得

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

□

1.4.3 贝叶斯公式

如果事件 B 是由于在两两互不相容的事件 $A_1, A_2, \cdots, A_n$ 中某一个发生的情况下而发生的，并且知道各个事件 A_i 发生的概率 $P(A_i)$ 以及在事件 A_i 发生的条件下事件 B 发生的条件概率 $P(B \mid A_i)$，则由全概率公式可得事件 B 发生的概率 $P(B)$. 我们把事件 $A_1, A_2, \cdots, A_n$ 看做是导致事件 B 发生的原因，$P(A_i)$ 称为**先验概率**，它反映出各种原因发生的可能性大小. 如果在试验中发生了事件 B，这一信息有助于探讨事件 B 发生的原因. 条件概率 $P(A_i \mid B)$ 称为**后验概率**，它使得我们在试验之后对各种原因发生的可能性大小有进一步的了解.

定理 1.2（贝叶斯公式） 对于任意事件 A, B，如果 $P(A)>0, P(B)>0$，则

$$P(A \mid B) = \frac{P(A)\,P(B \mid A)}{P(B)} \tag{1-8}$$

如果事件 $A_1, A_2, \cdots, A_n$ 构成一个完全事件组，且 $P(A_i)>0\ (i=1,2,\cdots,n)$，则对于任一事件 B，如果 $P(B)>0$，有

$$P(A_i \mid B) = \frac{P(A_i)\,P(B \mid A_i)}{\sum_{j=1}^{n} P(A_j)\,P(B \mid A_j)} \tag{1-9}$$

证明：如果 $P(A)>0, P(B)>0$，由乘法公式可得

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

由此得

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

对于完全事件组中的任一事件 A_i，有

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

利用全概率公式，得

$$P(A_i\mid B)=\frac{P(A_i)\,P(B\mid A_i)}{\sum\limits_{j=1}^{n}P(A_j)\,P(B\mid A_j)}$$

□

1.5 事件的独立性

定义 1.4 设 A 与 B 是同一试验 E 的两个事件，如果 $P(AB)=P(A)\,P(B)$，则称事件 A 与事件 B 是**相互独立**的.

对于同一试验 E 的两个事件 A 与 B，如果 $P(A)>0$，则 A 与 B 相互独立的充分必要条件是 $P(B\mid A)=P(B)$；如果 $P(B)>0$，则 A 与 B 相互独立的充分必要条件是 $P(A\mid B)=P(A)$.

结论 如果事件 A 与事件 B 相互独立，则事件 A 与事件 $\overline{B}$ 相互独立.

证明：$A=A(B\cup\overline{B})=(AB)\cup(A\overline{B})$，而 $(AB)(A\overline{B})=\varnothing$，所以

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

如果 A 与 B 相互独立，则 $P(AB)=P(A)\,P(B)$，代入上式可得

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

由此得

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

因此事件 A 与事件 $\overline{B}$ 是相互独立的. □

同理，如果事件 A 与事件 B 相互独立，则事件 $\overline{A}$ 与事件 B 相互独立，事件 $\overline{A}$ 与事件 $\overline{B}$ 相互独立.

定义 1.5 对于同一试验 E 的三个事件 A, B, C，如果满足

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

则称三个事件 A, B, C 是**两两相互独立**的.

定义 1.6 如果三个事件 A, B, C 是两两相互独立的，并且有 $P(ABC) = P(A)\,P(B)\,P(C)$，则称三个事件 A, B, C 是**相互独立**的.

定义 1.7 设 $A_1, A_2, \cdots, A_n$ 是同一试验 E 的 n 个事件，如果对于任意正整数 k 及这 n 个事件中的任意 $k\,(2 \leqslant k \leqslant n)$ 个事件 $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$，都有等式

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

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

若 $n\,(n \geqslant 2)$ 个事件相互独立，则其中任意 $k\,(2 \leqslant k \leqslant n)$ 个事件相互独立.

1.6 伯努利概型

如果将试验 E 重复执行 n 次，在每一次试验中，事件 A 或者发生，或者不发生. 假设每次试验的结果互不影响，即在每次试验中事件 A 发生的概率保持不变，不受其他各次试验结果的影响，则称这 n 次试验相互独立.

如果试验 E 只有两个可能的对立结果 A 和 $\overline{A}$，并且 $P(A)=p$，$P(\overline{A})=1-p$，其中 $0<p<1$. 将试验 E 独立地重复进行 n 次所构成的一个试验叫做 n **重伯努利试验**，简称为**伯努利试验**（Bernoulli experiment）或**伯努利概型**（Bernoulli probability model）.

n 重伯努利试验的基本事件可记为 $\omega=\omega_1\omega_2\cdots\omega_n$，其中 $\omega_i\,(1\leqslant i\leqslant n)$ 为 A 或者为 $\overline{A}$，即 ω 是从 A 及 $\overline{A}$ 中每次取 1 个，独立地重复取 n 次的一种排列，共有 2^n 个基本事件.

如果 ω 中有 k 个 A，则必有 $n-k$ 个 $\overline{A}$，由独立性可得这一基本事件的概率为 $p^k(1-p)^{n-k}$.

由于在 2^n 个基本事件中共有 C_n^k 个含 k 个 A 及 $n-k$ 个 $\overline{A}$，因此在 n 次独立重复试验中，事件 A 恰好发生 k 次的概率 $P_n(k)$ 为

$$P_n(k)=C_n^kp^k(1-p)^{n-k}\quad k=0,1,2,\cdots,n \tag{1-10}$$

由二项式定理可得

$$\sum_{k=0}^{n}P_n(k)=\sum_{k=0}^{n}C_n^kp^k(1-p)^{n-k}=[p+(1-p)]^n=1$$

由此可见，$C_n^kp^k(1-p)^{n-k}$ 是二项展开式中的一项，因此式 (1-10) 又称为**二项概率公式**.

2　随机变量及其分布

2.1　随机变量的分布函数

2.1.1　随机变量

定义 2.1　设随机试验 E 的样本空间为 $\Omega = \{\omega\}$. 如果对于每一个 $\omega \in \Omega$，都有一个实数 $X(\omega)$ 与之对应，则称 $X = X(\omega)$ 为**随机变量**（random variable）.

随机变量常用大写字母 X, Y, Z 等表示.

2.1.2　分布函数

定义 2.2　设 X 是一个随机变量，对于任意实数 x，令 $F(x) = P\{X \leqslant x\}$，称 $F(x)$ 为随机变量 X 的**分布函数**（cumulative distribution function）.

随机变量 X 的分布函数 $F(x)$ 是定义在 $(-\infty, +\infty)$ 上的函数，是随机事件 $\{X \leqslant x\}$ 发生的概率. 分布函数值 $F(a)$ 表示 X 落在区间 $(-\infty, a]$ 上的概率.

性质 1　对于任意实数 x，有 $0 \leqslant F(x) \leqslant 1$.

性质 2　对于任意实数 $x_1, x_2\,(x_1 < x_2)$，有 $P\{x_1 < X \leqslant x_2\} = F(x_2) - F(x_1)$.

证明： 对于任意实数 $x_1, x_2\,(x_1 < x_2)$，由于

$$\{x_1 < X \leqslant x_2\} = \{X \leqslant x_2\} - \{X \leqslant x_1\}$$

所以有

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

□

性质 3

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$

$$F(+\infty) = \lim_{x \to +\infty} F(x) = 1$$

性质 4 $F(x)$ 处处右连续，即 $F(x^+) = F(x)$.

性质 5 $F(x)$ 是一个单调不减函数.

证明： 对于任意实数 $x_1, x_2\,(x_1 < x_2)$，有

$$F(x_2) - F(x_1) = P\{x_1 < X \leqslant x_2\} \geqslant 0$$

因此 $F(x)$ 是单调不减函数. □

2.2 离散型随机变量及其概率分布

定义 2.3 如果一个随机变量 X 所有可能取到的不相同的值是有限个或可列无限多个，并且以确定的概率取这些不同的值，则称 X 为**离散型随机变量**（discrete random variable）.

定义 2.4 设离散型随机变量 X 所有可能取的值为 $x_k\,(k=1,2,\cdots)$，X 取各个可能值的概率，即事件 $\{X=x_k\}$ 的概率为

$$P\{X=x_k\}=p_k \quad k=1,2,\cdots \tag{2-1}$$

并且 p_k 满足以下两个条件：

1. 非负性：$p_k \geqslant 0$；
2. 归一性：$\sum\limits_{k=1}^{\infty} p_k = 1$，

则称式 (2-1) 为离散型随机变量 X 的**概率分布**（probability distribution）或**分布律**.

概率分布也可以用如下的表格来表示：

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

概率分布反映了离散型随机变量的统计规律性.

对于任意实数 x，随机事件 $\{X \leqslant x\}$ 可以表示成 $\bigcup\limits_{x_k \leqslant x}\{X=x_k\}$. 由于 $x_k\,(k=1,2,\cdots)$ 互不相同，根据概率的可加性，可得离散型随机变量 X 的分布函数为

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

2.3 连续型随机变量及其概率密度

定义 2.5 对于随机变量 X 的分布函数 $F(x)$，如果存在非负函数 $f(x)$，使得对任意的 x，都有 $F(x)=\displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$，则称随机变量 X 是**连续型随机变量**（continuous random variable），其中函数 $f(x)$ 叫做 X 的**概率密度函数**（probability density function），简称为**概率密度**（probability density），记作 $X \sim f(x)$.

由定义 2.5 可知，连续型随机变量的分布函数处处连续.

概率密度 $f(x)$ 的性质如下.

性质 1 $f(x) \geqslant 0$

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

满足性质 1 与性质 2 的函数 $f(x)$ 必然是某随机变量的概率密度.

性质 3 对于任意实数 $a, b\,(a < b)$，有

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

由以上性质可知，概率密度曲线总是位于 x 轴上方，并且介于它和 x 轴之间的面积等于 1；随机变量落在区间 $(a, b]$ 的概率 $P\{a < X \leqslant b\}$ 等于区间 $(a, b]$ 上曲线 $y = f(x)$ 之下的曲边梯形的面积.

性质 4 如果 $f(x)$ 在点 x 处连续，则有 $F'(x) = f(x)$.

由性质 4 可知，在 $f(x)$ 的连续点有

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

可见概率密度反映了随机变量在点 x 处概率分布的密集程度.$f(x)$ 的大小能反映出随机变量 X 在点 x 附近取值的可能性大小，即概率的大小. 因此，用概率密度描述连续型随机变量的分布比用分布函数更直观. 当不考虑高阶无穷小时，有

$$P\{x < X \leqslant x + \Delta x\} \approx f(x)\Delta x$$

结论 连续型随机变量取任意指定实数的概率均为零.

证明： 对于 X 的任意一个可取的值 x，设 $\Delta x > 0$，由于事件 $\{X = x\} \subseteq \{x - \Delta x < X \leqslant x\}$，因此有

$$0 \leqslant P\{X = x\} \leqslant P\{x - \Delta x < X \leqslant x\} = F(x) - F(x - \Delta x)$$

令 $\Delta x \to 0$，可得

$$P\{X = x\} = 0$$

因此，连续型随机变量取任意指定实数的概率均为零. □

据此，在计算连续型随机变量在某一区间取值的概率时，可以不区分该区间是开区间或闭区间或半开半闭区间，即有

$$P\{x_1 < X < x_2\} = P\{x_1 \leqslant X \leqslant x_2\} = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x$$

由上述结论可知，概率为 0 的事件未必是不可能事件.

2.4 常用的分布

2.4.1 (0-1) 分布

定义 2.6 如果离散型随机变量 X 只取 0 与 1 两个值，其概率分布为

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

或写成

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

则称随机变量 X 服从参数为 p 的 **(0-1) 分布**或**两点分布**（two-point distribution）.

服从两点分布的随机变量 X 的概率分布也可以写成

X	0	1
P	$1-p$	p

2.4.2 二项分布

在 n 重伯努利试验中，如果以 X 表示事件 A 出现的次数，则 X 是一个离散型随机变量，它的所有可能取值是 $0,1,2,\cdots,n$. 设 $P(A)=p\,(0<p<1)$，则由二项概率公式（式 (1-10)）可得

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

定义 2.7 如果随机变量 X 的概率分布为

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

则称随机变量 X 服从参数为 n,p 的**二项分布**（binomial distribution），记作 $X\sim B(n,p)$.

由定义 2.7 可得

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

特别地，当 $n=1$ 时，二项分布 $B(1,p)$ 的概率分布为

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

这就是 (0-1) 分布. 因此，(0-1) 分布是二项分布的特例.

2.4.3 泊松分布

定义 2.8 如果离散型随机变量 X 的所有可能取值为 $0,1,2,\cdots$，并且

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

其中 $\lambda>0$ 是常数，则称随机变量 X 服从参数为 λ 的**泊松分布**（poisson distribution），记作 $X\sim P(\lambda)$ 或 $X\sim\pi(\lambda)$.

易知

$$\frac{\lambda^k e^{-\lambda}}{k!} > 0$$

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

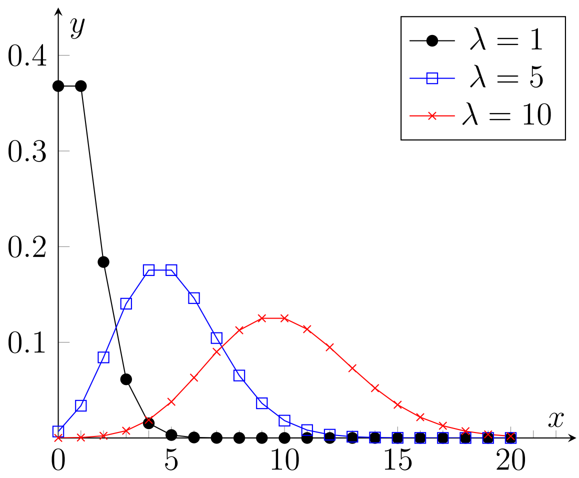


图 2.1　泊松分布

定理 2.1(泊松定理)　设 $\lambda > 0$ 是常数，n 为任意正整数，$np_n = \lambda$，则对任一固定的非负整数 k，有

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

证明：因为 $np_n = \lambda$，故 $p_n = \dfrac{\lambda}{n}$，从而对于任意固定的非负整数 k，有

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

对于固定的 k，当 $n\to\infty$ 时，有

$$1\times\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\to 1$$
$$\left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}$$
$$\left(1-\frac{\lambda}{n}\right)^{-k}\to 1$$

所以

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

□

由于 $np_n=\lambda$，意味着当 n 很大时 p_n 必定很小. 因此，泊松定理表明，当 n 很大而 p 很小时，有下面的近似公式

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

其中 $\lambda=np$. 也就是说，对于二项分布 $B(n,p)$，当 n 很大而 p 很小时，近似为泊松分布 $P(np)$.

2.4.4 几何分布

设试验 E 只有两个对立的结果 A 和 $\overline{A}$，并且 $P(A)=p$, $P(\overline{A})=1-p$，其中 $0<p<1$. 将试验 E 独立地重复进行下去，直到事件 A 发生为止. 如果用 X 表示所需要的试验次数，则 X 是一个随机变量，它可能取的值是 $1,2,3,\cdots$. X 的概率分布为

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

称随机变量 X 服从参数为 p 的**几何分布**（geometric distribution）.

易知

$$(1-p)^{k-1}p>0$$
$$\sum_{k=1}^{\infty}(1-p)^{k-1}p=p\sum_{i=0}^{\infty}(1-p)^i=p\cdot\frac{1}{1-(1-p)}=1$$

2.4.5 均匀分布

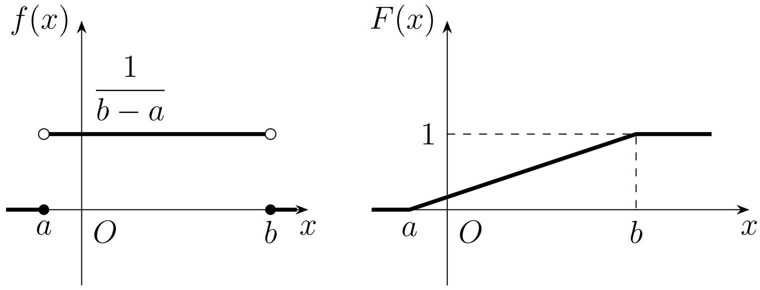
定义 2.9 如果连续型随机变量 X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\[2ex] 0 & \text{其他}\end{cases}$$

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

若 $X\sim U(a,b)$，则 X 的分布函数为

$$F(x)=\begin{cases}0 & x<a\\[1ex] \dfrac{x-a}{b-a} & a\leqslant x<b\\[2ex] 1 & x\geqslant b\end{cases}$$



(a) 均匀分布的概率密度 (b) 均匀分布的分布函数

图 2.2

对任意的两个数 $x_1,x_2\in(a,b)$，如果 $x_1<x_2$，则有

$$P\{x_1\leqslant X\leqslant x_2\}=\int_{x_1}^{x_2}\frac{\mathrm{d}x}{b-a}=\frac{x_2-x_1}{b-a}$$

这说明随机变量 X 位于区间 (a,b) 的任一子区间 $[x_1,x_2]$ 内的概率，只依赖于子区间 $[x_1,x_2]$ 的长度，而与子区间的位置无关.

2.4.6 指数分布

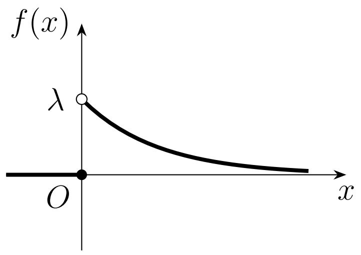
定义 2.10 如果连续型随机变量 X 的概率密度为

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

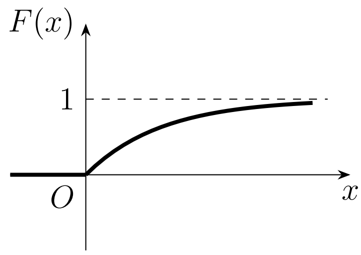
其中 $\lambda>0$ 是常数，则称 X 服从参数为 λ 的**指数分布**（exponential distribution）.

服从指数分布的随机变量 X 的分布函数为

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



(a) 指数分布的概率密度



(b) 指数分布的分布函数

图 2.3

2.4.7 正态分布

1 正态分布及其性质

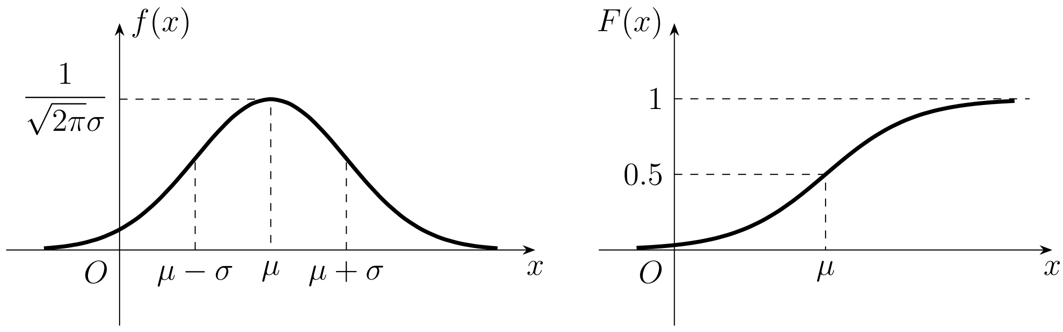
定义 2.11 如果连续型随机变量 X 的概率密度为

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

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

若 $X\sim N(\mu,\sigma^2)$，则 X 的分布函数为

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



(a) 正态分布的概率密度 (b) 正态分布的分布函数

图 2.4

概率密度曲线 $y=f(x)$ 关于直线 $x=\mu$ 对称，并在 $x=\mu$ 处取得最大值 $\dfrac{1}{\sqrt{2\pi}\sigma}$，在横坐标 $x=\mu\pm\sigma$ 处有拐点，以 x 轴为水平渐近线.

如果固定 σ，改变 μ 的值，则概率密度曲线沿着 x 轴平移，但形状不变，如图 2.5 所示.

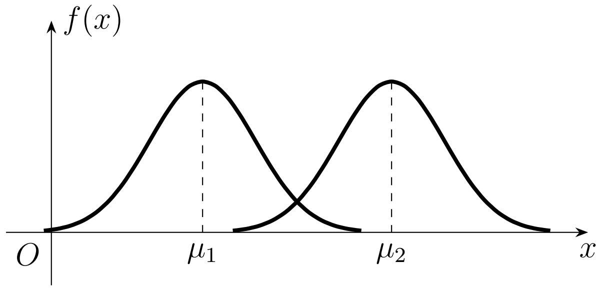


图 2.5

如果固定 μ，改变 σ 的值，则由 $y_{\max}=\dfrac{1}{\sqrt{2\pi}\sigma}$ 可知：σ 越小，概率密度曲线在 $x=\mu$ 附近越陡峭，X 落在 $x=\mu$ 附近的概率越大；σ 越大，概率密度曲线越平坦. 如图 2.6 所示.

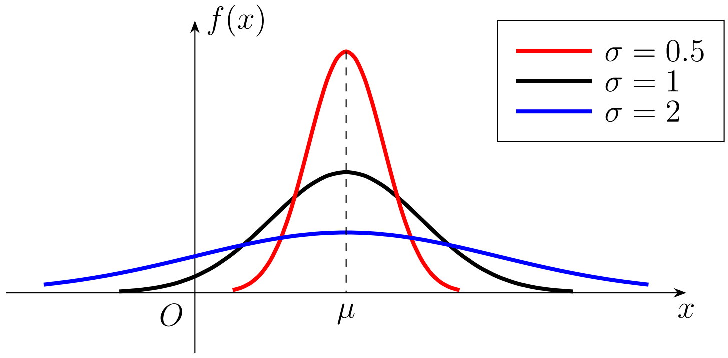


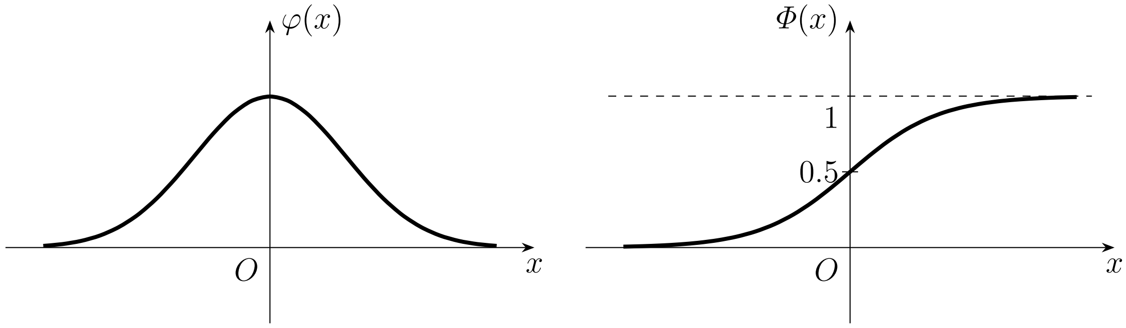
图 2.6

2 标准正态分布

定义 2.12 设 $X \sim N(\mu, \sigma^2)$，如果 $\mu = 0, \sigma = 1$，则称 X 服从**标准正态分布**（standard normal distribution），记作 $X \sim N(0, 1)$.

服从标准正态分布的随机变量 X 的概率密度记作 $\varphi(x)$，分布函数记作 $\Phi(x)$，即

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



(a) 标准正态分布的概率密度

(b) 标准正态分布的分布函数

图 2.7

性质 1

$$\Phi(-x) = 1 - \Phi(x) \tag{2-2}$$

证明： 由于 $\varphi(x)$ 是偶函数，所以 $\varphi(-x) = \varphi(x)$，则

$$\begin{aligned}
\Phi(-x) &= \int_{-\infty}^{-x} \varphi(t)\, \mathrm{d}t \\
&= \int_{-\infty}^{+\infty} \varphi(t)\, \mathrm{d}t - \int_{-x}^{+\infty} \varphi(t)\, \mathrm{d}t \\
&\xlongequal{u=-t} 1 - \int_{x}^{-\infty} \varphi(-u)\, \mathrm{d}(-u) \\
&= 1 - \int_{-\infty}^{x} \varphi(u)\, \mathrm{d}u \\
&= 1 - \Phi(x)
\end{aligned}$$

□

性质 2 若 $X \sim N(0,1)$，则对于任意正的实数 a，有

$$P\{|X| > a\} = 2[1 - \Phi(a)] \tag{2-3}$$

$$P\{|X| \leqslant a\} = 2\Phi(a) - 1 \tag{2-4}$$

证明：

$$P\{|X| > a\} = 2P\{X > a\} = 2(1 - P\{X \leqslant a\}) = 2[1 - \Phi(a)]$$

$$P\{|X| \leqslant a\} = 1 - P\{|X| > a\} = 1 - 2[1 - \Phi(a)] = 2\Phi(a) - 1$$

□

性质 3（正态分布的标准化） 若 $X \sim N(\mu, \sigma^2)$，则

$$F(x) = \Phi(\frac{x - \mu}{\sigma}) \tag{2-5}$$

证明：设 $X \sim N(\mu, \sigma^2)$，其分布函数为

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

令 $u = \dfrac{t - \mu}{\sigma}$，可得

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

即

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

□

由性质 3 可得，如果 $X \sim N(\mu, \sigma^2)$，则对于任意实数 $x_1, x_2\,(x_1 < x_2)$，有

$$P\{x_1 < X < x_2\} = F(x_2) - F(x_1) = \Phi(\frac{x_2 - \mu}{\sigma}) - \Phi(\frac{x_1 - \mu}{\sigma}) \tag{2-6}$$

结论（3σ 规则） 正态随机变量 X 以 99.74% 的概率落在 $(\mu-3\sigma,\mu+3\sigma)$ 区间内，而落在该区间以外的概率小于千分之三. 在一次试验中，X 落在 $(\mu-3\sigma,\mu+3\sigma)$ 区间以外这个事件几乎不会发生.

证明：由式 (2-2) 和式 (2-5) 有

$$\begin{aligned} P\{|X-\mu|<k\sigma\} &= P\{\mu-k\sigma<X\leqslant\mu+k\sigma\} \\ &= F(\mu+k\sigma)-F(\mu-k\sigma) \\ &= \Phi(k)-\Phi(-k) \\ &= 2\Phi(k)-1 \end{aligned}$$

所以

$$P\{|X-\mu|<\sigma\}=2\Phi(1)-1=0.6826$$
$$P\{|X-\mu|<2\sigma\}=2\Phi(2)-1=0.9544$$
$$P\{|X-\mu|<3\sigma\}=2\Phi(3)-1=0.9974$$

□

3 标准正态分布的上 α 分位点

定义 2.13 设 $X\sim N(0,1)$. 对于给定的 $\alpha\,(0<\alpha<1)$，如果 u_α 满足条件

$$P\{X\geqslant u_\alpha\}=\frac{1}{\sqrt{2\pi}}\int_{u_\alpha}^{+\infty}e^{-\frac{x^2}{2}}\,\mathrm{d}x=\alpha$$

则称点 u_α 为标准正态分布的**上 α 分位点**.

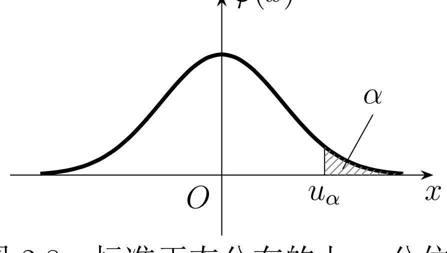


图 2.8 标准正态分布的上 α 分位点

性质 1 上 $1-\alpha$ 分位点 $u_{1-\alpha}$ 与上 α 分位点 u_α 关于原点对称，即

$$u_{1-\alpha} = -u_\alpha \tag{2-7}$$

性质 2

$$\Phi(u_\alpha) = 1-\alpha \tag{2-8}$$

证明：

$$\Phi(u_\alpha) = P\{X \leqslant u_\alpha\} = 1 - P\{X > u_\alpha\} = 1-\alpha$$

□

2.5 随机变量的函数的分布

2.5.1 离散型随机变量的函数的分布

设离散型随机变量 X 的概率分布为

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

$y = g(x)$ 是连续函数，则对于 X 的函数 $Y = g(X)$，有

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

如果数值 $g(x_k)(k = 1, 2, \cdots)$ 中有相等的，就把 Y 取这些相等的数值的概率相加，作为 $Y = g(X)$ 取该值的概率，便可得到 $Y = g(X)$ 的概率分布.

【例 2.1】 设随机变量 X 的概率密度为

X	-2	-1	0	1	2
P	0.2	0.1	0.3	0.2	0.2

求 $Y = (X+1)^2$ 的概率分布.

解：随机变量 $Y=(X+1)^2$ 的所有可能取值为 0,1,4,9，且 Y 取每个值的概率为

$$P\{Y=0\}=P\{(X+1)^2=0\}=P\{X=-1\}=0.1$$

$$P\{Y=1\}=P\{(X+1)^2=1\}=P\{X=0\}+P\{X=-2\}=0.5$$

$$P\{Y=4\}=P\{(X+1)^2=4\}=P\{X=1\}=0.2$$

$$P\{Y=9\}=P\{(X+1)^2=9\}=P\{X=2\}=0.2$$

所以 $Y=(X+1)^2$ 的概率分布为

Y	0	1	4	9
P	0.1	0.5	0.2	0.2

2.5.2 连续型随机变量的函数的分布

求随机变量 $Y=g(X)$ 的分布函数的关键一步是从 $g(X)\leqslant y$ 中解出 X 应满足的不等式.

【例 2.2】 设随机变量 X 在区间 $(-1,3)$ 上服从均匀分布，记

$$Y=\begin{cases}-1 & X<0\\ 0 & X=0\\ 1 & X>0\end{cases}$$

求随机变量 Y 的概率分布.

解： X 的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{4} & -1<x<3\\ 0 & \text{其他}\end{cases}$$

Y 是一个离散型随机变量，可取值为 -1,0,1，且

$$P\{Y=-1\}=P\{X<0\}=\int_{-1}^{0}\frac{1}{4}\mathrm{d}x=\frac{1}{4}$$

$$P\{Y=0\}=P\{X=0\}=0$$

$$P\{Y=1\}=P\{X>0\}=\int_{0}^{3}\frac{1}{4}\mathrm{d}x=\frac{3}{4}$$

因此，Y 的概率分布为

Y	-1	0	1
P	$\frac{1}{4}$	0	$\frac{3}{4}$

【例 2.3】 设随机变量 X 服从标准柯西分布，即 X 的概率密度为

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

求随机变量 $Y = \arctan X$ 的概率密度.

解： Y 的分布函数为

$$F_Y(y) = P\{Y \leqslant y\} = P\{\arctan X \leqslant y\}$$

当 $|y| < \frac{\pi}{2}$ 时，$\{\arctan X \leqslant y\} = \{X \leqslant \tan y\}$，于是得 Y 的分布函数 $F_Y(y)$ 与 X 的分布函数 $F_X(x)$ 之间的关系如下：

$$F_Y(y) = P\{\arctan X \leqslant y\} = P\{X \leqslant \tan y\} = F_X(\tan y)$$

因此 Y 的概率密度为

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) \\ &= \frac{\mathrm{d}}{\mathrm{d}y} F_X(\tan y) \\ &= f_X(\tan y) \cdot \sec^2 y \\ &= \frac{\sec^2 y}{\pi(1+\tan^2 y)} \\ &= \frac{1}{\pi} \end{aligned}$$

当 $y \geqslant \frac{\pi}{2}$ 时，事件 $\{\arctan X \leqslant y\}$ 是必然事件，所以 $F_Y(y) = 1$，从而 $f_Y(y) = 0$.

当 $y \leqslant -\frac{\pi}{2}$ 时，事件 $\{\arctan X \leqslant y\}$ 是不可能事件，所以 $F_Y(y) = 0$，从而 $f_Y(y) = 0$.

综上所述，Y 的概率密度为

$$f_Y(y) = \begin{cases} \frac{1}{\pi} & |y| < \frac{\pi}{2} \\ 0 & |y| \geqslant \frac{\pi}{2} \end{cases}$$

定理 2.2 设随机变量 X 的概率密度为 $f_X(x), -\infty < x < +\infty$，函数 $g(x)$ 是处处可导的严格单调函数，则随机变量 $Y = g(X)$ 的概率密度为

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

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

证明： 当 $g(x)$ 处处可导且严格单调增加时，它的反函数 $h(y)$ 在区间 (α, β) 内也处处可导且严格单调增加，所以当 $y \leqslant \alpha$ 时，$F_Y(y) = 0$；当 $y \geqslant \beta$ 时，$F_Y(y) = 1$；当 $\alpha < y < \beta$ 时，有

$$F_Y(y) = P\{Y \leqslant y\} = P\{g(X) \leqslant y\} = P\{X \leqslant h(y)\} = F_X(h(y))$$

于是 $Y = g(X)$ 的概率密度为

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

当 $g(x)$ 处处可导且严格单调减少时，它的反函数 $h(y)$ 在区间 (α, β) 内也处处可导且严格单调减少，所以当 $y \leqslant \alpha$ 时，$F_Y(y) = 0$；当 $y \geqslant \beta$ 时，$F_Y(y) = 1$；当 $\alpha < y < \beta$ 时，有

$$\begin{aligned} F_Y(y) &= P\{Y \leqslant y\} \\ &= P\{g(X) \leqslant y\} \\ &= P\{X \geqslant h(y)\} \\ &= 1 - P\{X < h(y)\} \\ &= 1 - F_X(h(y)) \end{aligned}$$

于是 $Y = g(X)$ 的概率密度为

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

综上所述，$Y = g(X)$ 的概率密度为

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

□

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

证明： 因为 $X \sim N(\mu, \sigma^2)$，所以 X 的概率密度为

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

由 $y = ax + b$ 解得

$$x = h(y) = \frac{y-b}{a}$$
$$h'(y) = \frac{1}{a}$$

由定理 2.2 得 $Y = aX + b$ 的概率密度为

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

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

特别地，当 $a = \dfrac{1}{\sigma}, b = -\dfrac{\mu}{\sigma}$ 时，有 $Y = \dfrac{X-\mu}{\sigma} \sim N(0, 1)$.

3　二维随机变量及其分布

3.1　二维随机变量

3.1.1　二维随机变量及其分布函数

定义 3.1　设随机试验 E 的样本空间为 Ω，X 和 Y 是定义在 Ω 上的两个随机变量，由它们构成的向量 (X, Y) 称为**二维随机变量**或**二维随机向量**.

定义 3.2　设 (X, Y) 是二维随机变量. 对于任意实数 x 和 y，记事件 $\{X \leqslant x\}$ 与 $\{Y \leqslant y\}$ 的交事件为 $\{X \leqslant x, Y \leqslant y\}$，称二元函数

$$F(x, y) = P\{X \leqslant x, Y \leqslant y\},\ (x, y) \in \mathbf{R}^2$$

为**二维随机变量 (X, Y) 的分布函数**，或称为随机变量 X 和 Y 的**联合分布函数**.

如果将二维随机变量 (X, Y) 看做 xOy 平面上随机点的坐标，则分布函数 $F(x, y)$ 在点 (x, y) 处的函数值就是随机点落在以 (x, y) 为顶点且位于该点左下方的无界域内的概率，如图 3.1 所示.

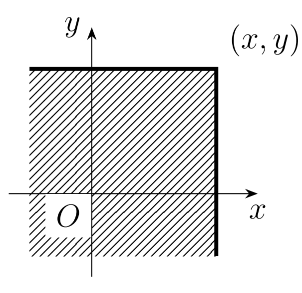


图 3.1

分布函数 $F(x,y)$ 具有如下性质：

性质 1 $0 \leqslant F(x,y) \leqslant 1$，且

$$F(-\infty,-\infty)=\lim_{\substack{x\to-\infty\\ y\to-\infty}} F(x,y)=0$$

$$F(+\infty,+\infty)=\lim_{\substack{x\to+\infty\\ y\to+\infty}} F(x,y)=1$$

对于任意固定的 x，有 $F(x,-\infty)=\lim\limits_{y\to-\infty} F(x,y)=0.$
对于任意固定的 y，有 $F(-\infty,y)=\lim\limits_{x\to-\infty} F(x,y)=0.$

性质 2 $F(x,y)$ 对于每个变量都是单调不减函数. 即对于任意固定的 y，当 $x_1 < x_2$ 时，有 $F(x_1,y) \leqslant F(x_2,y)$；对于任意固定的 x，当 $y_1 < y_2$ 时，有 $F(x,y_1) \leqslant F(x,y_2)$.

性质 3 $F(x,y)$ 关于 x 右连续，关于 y 右连续，即

$$F(x^+,y)=F(x,y)$$

$$F(x,y^+)=F(x,y)$$

性质 4 对于任意的 $x_1 < x_2,\ y_1 < y_2$，有

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

3.1.2 边缘分布

定义 3.3 设二维随机变量 (X,Y) 的分布函数为 $F(x,y)$，记随机变量 X 的分布函数为 $F_X(x)$，随机变量 Y 的分布函数为 $F_Y(y)$，分别称为二维随机变量 (X,Y) 关于 X 和关于 Y 的**边缘分布函数**.

$$F_X(x) = P\{X \leqslant x\} = P\{X \leqslant x, Y < +\infty\} = F(x, +\infty), \quad x \in \mathbf{R}$$
$$F_Y(y) = P\{Y \leqslant y\} = P\{X < +\infty, Y \leqslant y\} = F(+\infty, y), \quad y \in \mathbf{R}$$

3.1.3 随机变量的独立性

定义 3.4 设二维随机变量 (X, Y) 的分布函数及其关于 X 和关于 Y 的边缘分布函数分别为 $F(x, y), F_X(x), F_Y(y)$，如果对于任意实数 x, y，都有 $F(x, y) = F_X(x)\, F_Y(y)$，则称随机变量 X 与 Y 是**相互独立**的.

3.2 二维离散型随机变量

3.2.1 二维离散型随机变量及其概率分布

定义 3.5 如果二维随机变量 (X, Y) 所有可能取的值是有限对或可列无限对，则称 (X, Y) 是**二维离散型随机变量**.

设 (X, Y) 所有可能取的值为 $(x_i, y_j)(i, j = 1, 2, \cdots)$，记事件 $\{X = x_i\}$ 与 $\{Y = y_j\}$ 的交事件为 $\{X = x_i, Y = y_j\}$.

定义 3.6 设 (X, Y) 所有可能取的值为 $(x_i, y_j)(i, j = 1, 2, \cdots)$，如果

$$P\{X = x_i, Y = y_j\} = p_{ij},\ i, j = 1, 2, \cdots \tag{3-1}$$

且有

$$p_{ij} \geqslant 0$$
$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$$

则称式 (3-1) 为二维离散型随机变量 (X, Y) 的**概率分布**，或称为随机变量 X 和随机变量 Y 的**联合概率分布或联合分布律**.

(X,Y) 的概率分布可以用如下的表格表示：

$X \backslash Y$	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}	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	

如果二维随机变量 (X,Y) 的概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}\,(i,j=1,2,\cdots)$，则 (X,Y) 的分布函数为

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

3.2.2 边缘概率分布

设二维离散型随机变量 (X,Y) 的概率分布为

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

则有

$$P\{X=x_i\}=P\{X=x_i,\bigcup_{j=1}^{\infty}\{Y=y_j\}\}=P\{\bigcup_{j=1}^{\infty}\{X=x_i,Y=y_j\}\}$$

由于事件 $\{X=x_i,Y=y_j\}(j=1,2,\cdots)$ 是互不相容的，因此

$$P\{X=x_i\}=\sum_{j=1}^{\infty}P\{X=x_i,Y=y_j\}=\sum_{j=1}^{\infty}p_{ij}$$

记 $\sum\limits_{j=1}^{\infty}p_{ij}=p_{i\cdot}$，则有

$$P\{X=x_i\}=p_{i\cdot},\ i=1,2,\cdots$$

同理可得

$$P\{Y=y_j\}=\sum_{i=1}^{\infty}P\{X=x_i,Y=y_j\}=\sum_{i=1}^{\infty}p_{ij}=p_{\cdot j},\ j=1,2,\cdots$$

定义 3.7 设二维离散型随机变量 (X,Y) 的概率分布为

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

随机变量 X 和 Y 的概率分布

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

分别称为 (X,Y) 关于 X 和关于 Y 的**边缘概率分布**或**边缘分布律**.

3.2.3 随机变量的独立性

设二维离散型随机变量 (X,Y) 的概率分布为

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

(X,Y) 关于 X 和关于 Y 的边缘概率分布依次为

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

则随机变量 X 和 Y 相互独立的充分必要条件是：对任意的 i,j，都有

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

即

$$p_{ij}=p_{i\cdot}\,p_{\cdot j},\ i,j=1,2,\cdots$$

3.3 二维连续型随机变量

3.3.1 二维连续型随机变量及其概率密度

定义 3.8 设二维随机变量 (X,Y) 的分布函数为 $F(x,y)$，如果存在非负的二元函数 $f(x,y)$，对于任意实数 x,y，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**，$f(x,y)$ 称为二维连续型随机变量 (X,Y) 的**概率密度**，或称为随机变量 X 和 Y 的**联合概率密度**.

概率密度 $f(x,y)$ 具有下列性质：

性质 1 $f(x,y)\geqslant 0$

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

性质 3 如果 $f(x,y)$ 在点 (x,y) 处连续，则

$$f(x,y)=\frac{\partial^2F(x,y)}{\partial x\partial y}$$

性质 4 设 G 是 xOy 平面上的一个区域，则

$$P\{(X,Y)\in G\}=\iint\limits_{G}f(x,y)\,\mathrm{d}x\mathrm{d}y$$

$z=f(x,y)$ 表示空间 $Oxyz$ 中的一张曲面. 性质 1 和性质 2 表明，曲面 $z=f(x,y)$ 位于 xOy 平面上方，介于它和 xOy 平面之间的体积为 1. 性质 4 表示，随机点 (X,Y) 落在区域 G 内的概率 $P\{(X,Y)\in G\}$ 等于以 G 为底、以曲面 $z=f(x,y)$ 为顶的曲顶柱体体积的数值.

3.3.2 边缘概率密度

定义 3.9 设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y),(x,y)\in\mathbf{R}^2$，将一元函数

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

分别称为二维随机变量 (X,Y) 关于 X 和关于 Y 的**边缘概率密度**.

3.3.3 随机变量的独立性

设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y),(x,y)\in\mathbf{R}^2$，$(X,Y)$ 关于 X 和关于 Y 的边缘概率密度分别为 $f_X(x)$ 和 $f_Y(y)$，则随机变量 X 和 Y 相互独立的充分必要条件是：对任意 $x,y\in\mathbf{R}$，都有 $f(x,y)=f_X(x)\,f_Y(y)$.

3.3.4 二维均匀分布

设 D 是 xOy 平面上的有界区域，其面积为 A. 如果二维随机变量 (X,Y) 具有概率密度

$$f(x,y)=\begin{cases}\dfrac{1}{A} & (x,y)\in D\\[2ex] 0 & \text{其他}\end{cases}$$

则称 (X,Y) 在区域 D 上服从**均匀分布**.

3.3.5 二维正态分布

设二维随机变量 (X,Y) 的概率密度为

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

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 都是常数，且 $\sigma_1>0,\sigma_2>0,-1<\rho<1$，则称 (X,Y) 服从参数为 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 的**二维正态分布**，记作 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$.

结论 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 (X,Y) 关于 X 和关于 Y 的边缘概率密度分别为

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, -\infty < x < +\infty \\
f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, -\infty < y < +\infty
\end{aligned}
$$

证明：

$$
\begin{aligned}
&\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right] \\
&=\frac{-1}{2(1-\rho^2)}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2}+\left[\frac{(y-\mu_2)^2}{\sigma_2^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\rho^2\frac{(x-\mu_1)^2}{\sigma_1^2}\right]-\rho^2\frac{(x-\mu_1)^2}{\sigma_1^2}\right\} \\
&=\frac{-1}{2(1-\rho^2)}\left\{\frac{(1-\rho^2)(x-\mu_1)^2}{\sigma_1^2}+\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \\
&=-\frac{(x-\mu_1)^2}{2\sigma_1^2}-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2
\end{aligned}
$$

因此

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty} e^{\frac{-1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2}\,\mathrm{d}y
\end{aligned}
$$

对于任意给定的实数 x，令 $t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)$，则

$$
\mathrm{d}t=\frac{1}{\sigma_2\sqrt{1-\rho^2}}\mathrm{d}y
$$

因此

$$
\begin{aligned}
f_X(x) &= \frac{1}{2\pi\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\frac{1}{\sigma_2\sqrt{1-\rho^2}}\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,\mathrm{d}t
\end{aligned}
$$

因为

$$
\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,\mathrm{d}t=1
$$

所以

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

同理可得

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

□

由此可知，若 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则 (X,Y) 关于 X 和关于 Y 的边缘分布都是一维正态分布，且有 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$. (X,Y) 的分布与参数 ρ 有关，对于不同的 ρ，有不同的二维正态分布，但 (X,Y) 关于 X 和关于 Y 的边缘分布都与 ρ 无关.

上述结论还表明，仅仅根据关于 X 和关于 Y 的边缘分布，一般是不能确定随机变量 X 和 Y 的联合分布的.

结论 设 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则 X 与 Y 相互独立的充分必要条件是 $\rho = 0$.

证明：

$$f_X(x)\, f_Y(y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(y-\mu_2)^2}{2\sigma_2^2}\right]}$$

先证充分性. 如果 $\rho = 0$，则对于任意实数 x 和 y，都有 $f(x,y) = f_X(x)\, f_Y(y)$，因此 X 和 Y 相互独立. 充分性得证.

再证必要性. 如果 X 和 Y 相互独立，由于 $f(x,y), f_X(x), f_Y(y)$ 都是连续函数，因此对于任意实数 x 和 y，都有 $f(x,y) = f_X(x)\, f_Y(y)$. 如果取 $x = \mu_1, y = \mu_2$，则有

$$\begin{cases} f(\mu_1, \mu_2) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ f_X(\mu_1)\, f_Y(\mu_2) = \dfrac{1}{2\pi\sigma_1\sigma_2} \\ f(\mu_1, \mu_2) = f_X(\mu_1)\, f_Y(\mu_2) \end{cases}$$

从而 $\rho = 0$. 必要性得证. □

3.4 条件分布

3.4.1 离散型随机变量的条件分布

定义 3.10 设 (X,Y) 是二维离散型随机变量，对于固定的 j，如果 $P\{Y=y_j\}=p_{\cdot j}>0$，则称

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

为在条件 $Y=y_j$ 下随机变量 X 的**条件概率分布**.

对于固定的 i，如果 $P\{X=x_i\}=p_{i\cdot}>0$，则称

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

为在条件 $X=x_i$ 下随机变量 Y 的**条件概率分布**.

3.4.2 连续型随机变量的条件分布

定义 3.11 设 (X,Y) 是二维连续型随机变量，对于给定的实数 y 及任意给定的正数 ε，都有 $P\{y-\varepsilon<Y\leqslant y+\varepsilon\}>0$. 如果对于任意实数 x，极限

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

存在，则称此极限值为在条件 $Y=y$ 下 X 的**条件分布函数**，记作 $F_{X|Y}(x \mid y)$. 如果非负函数 $f_{X|Y}(x \mid y)$ 使得

$$F_{X|Y}(x \mid y)=\int_{-\infty}^{x}f_{X|Y}(u \mid y)\,\mathrm{d}u,\ x\in\mathbf{R}$$

成立，则称 $f_{X|Y}(x \mid y)$ 为在条件 $Y=y$ 下 X 的**条件概率密度**.

设二维连续型随机变量 (X,Y) 的分布函数为 $F(x,y)$，概率密度为 $f(x,y)$. 如果在点 (x,y) 处 $f(x,y)$ 连续，(X,Y) 关于 Y 的边缘概率密度 $f_Y(y)$ 连续，且 $f_Y(y)>0$，Y 的分布

函数为 $F_Y(y)$，则有

$$\begin{aligned}
F_{X|Y}(x \mid y) &= \lim_{\varepsilon\to 0^+} \frac{P\{X \leqslant x, y-\varepsilon < Y \leqslant y+\varepsilon\}}{P\{y-\varepsilon < Y \leqslant y+\varepsilon\}} \\
&= \lim_{\varepsilon\to 0^+} \frac{F(x,y+\varepsilon)-F(x,y-\varepsilon)}{F_Y(y+\varepsilon)-F_Y(y-\varepsilon)} \\
&= \lim_{\varepsilon\to 0^+} \frac{\dfrac{F(x,y+\varepsilon)-F(x,y)}{\varepsilon}+\dfrac{F(x,y-\varepsilon)-F(x,y)}{-\varepsilon}}{\dfrac{F_Y(y+\varepsilon)-F_Y(y)}{\varepsilon}+\dfrac{F_Y(y-\varepsilon)-F_Y(y)}{-\varepsilon}} \\
&= \frac{\dfrac{\partial F(x,y)}{\partial y}}{\dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}y}} \\
&= \frac{\displaystyle\int_{-\infty}^{x} f(u,y)\,\mathrm{d}u}{f_Y(y)} \\
&= \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}\mathrm{d}u
\end{aligned}$$

由此可得在条件 $Y=y$ 下 X 的条件概率密度为

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)},\ x \in \mathbf{R}$$

类似地，可以定义在条件 $X=x$ 下 Y 的条件分布函数 $F_{Y|X}(y \mid x)$ 和在条件 $X=x$ 下 Y 的条件概率密度 $f_{Y|X}(y \mid x) = \dfrac{f(x,y)}{f_X(x)},\ y \in \mathbf{R}$.

3.5 二维随机变量的函数的分布

设 (X,Y) 为二维随机变量，$g(x,y)$ 为二元函数，则一维随机变量 $Z=g(X,Y)$ 是二维随机变量 (X,Y) 的函数.

3.5.1 二维离散型随机变量的函数的分布

【例 3.1】 设 (X,Y) 的概率分布为

$X \backslash Y$	-1	1	2
0	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
2	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{10}$

求 $X+Y$ 的概率分布.

解： $X+Y$ 所有可能取的值为 $-1,1,2,3,4$，且

$$P\{X+Y=-1\}=P\{X=0,Y=-1\}=\frac{1}{10}$$

$$P\{X+Y=1\}=P\{X=0,Y=1\}+P\{X=2,Y=-1\}=\frac{2}{10}+\frac{3}{10}=\frac{1}{2}$$

$$P\{X+Y=2\}=P\{X=0,Y=2\}=\frac{1}{10}$$

$$P\{X+Y=3\}=P\{X=2,Y=1\}=\frac{1}{10}$$

$$P\{X+Y=4\}=P\{X=2,Y=2\}=\frac{2}{10}=\frac{1}{5}$$

即 $X+Y$ 的概率分布为

$X+Y$	-1	1	2	3	4
P	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$

结论 设随机变量 X 和 Y 相互独立，且 $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$，则 $X+Y\sim P(\lambda_1+\lambda_2)$.

证明：因为 $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$，所以

$$P\{X=i\}=\frac{\lambda_1^i e^{-\lambda_1}}{i!},\ i=0,1,2,\cdots$$

$$P\{Y=j\}=\frac{\lambda_2^j e^{-\lambda_2}}{j!},\ j=0,1,2,\cdots$$

$X+Y$ 所有可能取的值为 $0,1,2,\cdots$. 由于 X 和 Y 相互独立，因此对于任意的非负整数 k，有

$$
\begin{aligned}
P\{X+Y=k\} &= P(\bigcup_{l=0}^{k}\{X=l,Y=k-l\}) \\
&= \sum_{l=0}^{k}(P\{X=l\}\cdot P\{Y=k-l\}) \\
&= \sum_{l=0}^{k}\left[\frac{\lambda_1^l e^{-\lambda_1}}{l!}\frac{\lambda_2^{k-l}e^{-\lambda_2}}{(k-l)!}\right] \\
&= \sum_{l=0}^{k}\left[\frac{k!}{l!(k-l)!}\lambda_1^l\lambda_2^{k-l}\frac{e^{-(\lambda_1+\lambda_2)}}{k!}\right] \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{l!(k-l)!}\lambda_1^l\lambda_2^{k-l} \\
&= \frac{(\lambda_1+\lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!}
\end{aligned}
$$

即 $X+Y\sim P(\lambda_1+\lambda_2)$. □

3.5.2 二维连续型随机变量的函数的分布

1 $Z=X+Y$ 的概率密度

结论 设 (X,Y) 为二维连续型随机变量，其概率密度为 $f(x,y)$，则 $Z=X+Y$ 的概率密度为

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

证明：首先求 Z 的分布函数

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

对固定的 z 和 x，作变量代换 $y=u-x$，得

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

因此

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

于是，随机变量 Z 的概率密度为

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

同理可得

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

□

如果 X 和 Y 相互独立，设 $f_X(x)$ 和 $f_Y(y)$ 分别是二维随机变量 (X,Y) 关于 X 和关于 Y 的边缘概率密度，则有

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

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

上式称为 f_X 和 f_Y 的**卷积公式**，记作 $f_X * f_Y$，即

$$f_Z(z) = f_X * f_Y = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\,\mathrm{d}y \tag{3-2}$$

结论 如果随机变量 X 和 Y 相互独立，并且 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，则 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

2 $M=\max(X,Y)$ 及 $N=\min(X,Y)$ 的分布函数

设 (X,Y) 为二维连续型随机变量，其概率密度为 $f(x,y)$. 随机变量 X 和 Y 相互独立，分布函数分别为 $F_X(x)$ 和 $F_Y(y)$，$M=\max(X,Y)$ 和 $N=\min(X,Y)$ 的分布函数分别为 $F_{\max}(z)$ 和 $F_{\min}(z)$.

由于事件 $\{M \leqslant z\} = \{X \leqslant z, Y \leqslant z\}$，而 X 和 Y 相互独立，所以事件 $\{X \leqslant z\}$ 与事件 $\{Y \leqslant z\}$ 相互独立，由此可得

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

由于事件 $\{N > z\} = \{X > z, Y > z\}$，而 X 和 Y 相互独立，所以事件 $\{X > z\}$ 与事件 $\{Y > z\}$ 相互独立，由此可得

$$\begin{aligned} F_{\min}(z) &= P\{N \leqslant z\} \\ &= 1 - P\{N > z\} \\ &= 1 - P\{X > z, Y > z\} \\ &= 1 - P\{X > z\}\, P\{Y > z\} \\ &= 1 - (1 - P\{X \leqslant z\})(1 - P\{Y \leqslant z\}) \\ &= 1 - [1 - F_X(z)][1 - F_Y(z)] \end{aligned}$$

3 瑞利分布

设随机变量 X 和 Y 相互独立，并且都服从正态分布 $N(0, \sigma^2)$，则随机变量 $Z = \sqrt{X^2 + Y^2}$ 服从参数为 σ $(\sigma > 0)$ 的**瑞利分布**（Rayleigh distribution）.

由于随机变量 X 和 Y 相互独立，因此二维随机变量 (X, Y) 的概率密度为

$$f(x, y) = f_X(x)\, f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

当 $z < 0$ 时，有

$$F_Z(z) = P\{Z \leqslant z\} = P\{\sqrt{X^2 + Y^2} \leqslant z\} = 0$$

当 $z \geqslant 0$ 时，有

$$
\begin{aligned}
F_Z(z) &= P\{Z \leqslant z\} = P\{\sqrt{X^2+Y^2} \leqslant z\} \\
&= \iint\limits_{\sqrt{x^2+y^2}\leqslant z} f(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \iint\limits_{\sqrt{x^2+y^2}\leqslant z} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}\,\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta \int_0^z \frac{1}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} r\,\mathrm{d}r \\
&= -e^{-\frac{r^2}{2\sigma^2}}\Big|_0^z \\
&= 1 - e^{-\frac{z^2}{2\sigma^2}}
\end{aligned}
$$

因此，随机变量 Z 的分布函数为

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

由此可得 Z 的概率密度为

$$
f_Z(z) = F_Z'(z) = \begin{cases} \dfrac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} & z \geqslant 0 \\ 0 & z < 0 \end{cases}
$$

综上所述，如果随机变量 X 服从参数为 σ $(\sigma > 0)$ 的瑞利分布，则

$$
F(X) = \begin{cases} 1 - e^{-\frac{x^2}{2\sigma^2}} & x \geqslant 0 \\ 0 & x < 0 \end{cases}
$$

$$
f(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & x \geqslant 0 \\ 0 & x < 0 \end{cases}
$$

3.6 n 维随机变量

定义 3.12 设 E 是一个随机试验，其样本空间为 Ω. 对于定义在 Ω 上的 n 个随机变量 $X_1, X_2, \cdots, X_n$，称由它们构成的向量 $(X_1, X_2, \cdots, X_n)$ 为 n **维随机向量**或 n **维随机变量**.

定义 3.13 对于任意 n 个实数 $x_1, x_2, \cdots, x_n$，称 n 元函数

$$F(x_1, x_2, \cdots, x_n) = P\{X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n\}$$

为 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的**分布函数**，或称为随机变量 $X_1, X_2, \cdots, X_n$ 的**联合分布函数**.

定义 3.14 如果 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 所有可能取的值是有限个或可列无限个 n 元数组，则称之为 n 维**离散型随机变量**，其概率分布（也叫做 $X_1, X_2, \cdots, X_n$ 的**联合概率分布**）为

$$P\{X_1 = x_{i_1}, X_2 = x_{i_2}, \cdots, X_n = x_{i_n}\} = p_{i_1 i_2 \cdots i_n},\ i_1, i_2, \cdots, i_n = 1, 2, \cdots$$

定义 3.15 如果存在非负的 n 元函数 $f(x_1, x_2, \cdots, x_n)$，使得对于任意的 n 个实数 $x_1, x_2, \cdots, x_n$，都有

$$F(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \cdots, t_n)\, \mathrm{d}t_n \cdots \mathrm{d}t_2\, \mathrm{d}t_1$$

则称 $(X_1, X_2, \cdots, X_n)$ 为 n 维连续型随机变量，称 $f(x_1, x_2, \cdots, x_n)$ 为 $(X_1, X_2, \cdots, X_n)$ 的**概率密度**或 $X_1, X_2, \cdots, X_n$ 的**联合概率密度**.

结论 如果已知 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 的分布函数 $F(x_1, x_2, \cdots, x_n)$，则可确定 $(X_1, X_2, \cdots, X_n)$ 的 $k\,(1 \leqslant k < n)$ 维边缘分布函数：在 $F(x_1, x_2, \cdots, x_n)$ 中保留相应位置的 k 个变量，而让其他变量趋向于 $+\infty$，其极限即为所求.

例如，$(X_1, X_2, \cdots, X_n)$ 关于 X_1 的边缘分布函数为

$$F_{X_1}(x_1) = F(x_1, +\infty, +\infty, \cdots, +\infty),\ x_1 \in \mathbf{R}$$

而 $(X_1, X_2, \cdots, X_n)$ 关于 (X_1, X_2, X_3) 的边缘分布函数为

$$F_{X_1 X_2 X_3}(x_1, x_2, x_3) = F(x_1, x_2, x_3, +\infty, +\infty, \cdots, +\infty),\ x_1, x_2, x_3 \in \mathbf{R}$$

如果 n 维连续型随机变量 $(X_1, X_2, \cdots, X_n)$ 具有概率密度 $f(x_1, x_2, \cdots, x_n)$，则 $(X_1, X_2, \cdots, X_n)$ 关于 X_1 的边缘概率密度为

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1, x_2, \cdots, x_n)\,\mathrm{d}x_2\mathrm{d}x_3\cdots\mathrm{d}x_n,\ x_1 \in \mathbf{R}$$

而 $(X_1, X_2, \cdots, X_n)$ 关于 (X_1, X_2, X_3) 的边缘概率密度为

$$f_{X_1X_2X_3}(x_1, x_2, x_3) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(x_1, x_2, \cdots, x_n)\,\mathrm{d}x_4\mathrm{d}x_5\cdots\mathrm{d}x_n,\ x_1, x_2, x_3 \in \mathbf{R}$$

设 n 维离散型随机变量 $(X_1, X_2, \cdots, X_n)$ 的概率分布为

$$P\{X_1 = x_{i_1}, X_2 = x_{i_2}, \cdots, X_n = x_{i_n}\} = p_{i_1i_2\cdots i_n},\ i_1, i_2, \cdots, i_n = 1, 2, \cdots$$

则 $(X_1, X_2, \cdots, X_n)$ 关于 X_1 的边缘概率分布（或边缘分布律）为

$$P\{X_1 = x_{i_1}\} = \sum_{i_2=1}^{\infty}\sum_{i_3=1}^{\infty}\cdots\sum_{i_n=1}^{\infty} p_{i_1i_2\cdots i_n},\ i_1 = 1, 2, \cdots$$

定义 3.16 如果对于任意 n 个实数 $x_1, x_2, \cdots, x_n$，有

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

则称随机变量 $X_1, X_2, \cdots, X_n$ 是**相互独立**的.

结论 如果 $(X_1, X_2, \cdots, X_n)$ 是 n 维离散型随机变量，则 $X_1, X_2, \cdots, X_n$ 相互独立的充分必要条件是：对于 $(X_1, X_2, \cdots, X_n)$ 的任意一组可能取的值 $x_{i_1}, x_{i_2}, \cdots, x_{i_n}$，有

$$\begin{aligned}
&P\{X_1 = x_{i_1}, X_2 = x_{i_2}, \cdots, X_n = x_{i_n}\}\\
=\ &P\{X_1 = x_{i_1}\}\,P\{X_2 = x_{i_2}\}\cdots P\{X_n = x_{i_n}\}\\
=\ &\prod_{j=1}^{n} P\{X_j = x_{i_j}\}
\end{aligned}$$

结论 如果 $(X_1, X_2, \cdots, X_n)$ 是 n 维连续型随机变量，则 $X_1, X_2, \cdots, X_n$ 相互独立的充分必要条件是：对于任意 n 个实数 $x_1, x_2, \cdots, x_n$，有

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

结论 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，且均服从 (0-1) 分布，其概率分布为

$$P\{X_k = 0\} = 1 - p,\ P\{X_k = 1\} = p,\ 0 < p < 1,\ k = 1, 2, \cdots, n$$

则随机变量 $X = X_1 + X_2 + \cdots + X_n$ 服从二项分布 $B(n, p)$.

结论 如果 n 维随机变量 $X_1, X_2, \cdots, X_n$ 相互独立，并且 $X_i \sim N(\mu_i, \sigma_i^2)\,(i = 1, 2, \cdots, n)$，则它们的和 $X = X_1 + X_2 + \cdots + X_n$ 仍然服从正态分布，且有

$$X \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

它们的线性函数 $\sum\limits_{i=1}^{n} c_i X_i$ 仍然服从正态分布，且有 $\sum\limits_{i=1}^{n} c_i X_i \sim N(\sum\limits_{i=1}^{n} c_i \mu_i, \sum\limits_{i=1}^{n} c_i^2 \sigma_i^2)$，其中 $c_1, c_2, \cdots, c_n$ 是不全为零的常数.

定义 3.17 如果对于任意 $m + n$ 个实数 $x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n$，有

$$F(x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n) = F_1(x_1, x_2, \cdots, x_m) F_2(y_1, y_2, \cdots, y_n)$$

其中 F, F_1 和 F_2 分别是 $m + n$ 维随机变量 $(X_1, X_2, \cdots, X_m, Y_1, Y_2, \cdots, Y_n)$，$m$ 维随机变量 $(X_1, X_2, \cdots, X_m)$ 和 n 维随机变量 $(Y_1, Y_2, \cdots, Y_n)$ 的分布函数，则称 m 维随机变量 $(X_1, X_2, \cdots, X_m)$ 和 n 维随机变量 $(Y_1, Y_2, \cdots, Y_n)$ 是相互独立的.

结论 设 $(X_1, X_2, \cdots, X_m)$ 和 $(Y_1, Y_2, \cdots, Y_n)$ 相互独立，则 $X_i\,(i = 1, 2, \cdots, m)$ 和 $Y_j\,(j = 1, 2, \cdots, n)$ 相互独立. 如果 h, g 是连续函数，则随机变量 $h(X_1, X_2, \cdots, X_m)$ 和 $g(Y_1, Y_2, \cdots, Y_n)$ 相互独立.

结论 设 $X_1, X_2, \cdots, X_n$ 是 n 个相互独立的随机变量，它们的分布函数分别为 $F_{X_1}(x_1), F_{X_2}(x_2), \cdots, F_{X_n}(x_n)$，则随机变量 $M = \max(X_1, X_2, \cdots, X_n)$ 的分布函数为

$$F_{\max}(z) = F_{X_1}(z) F_{X_2}(z) \cdots F_{X_n}(z) = \prod_{i=1}^{n} F_{X_i}(z),\ z \in \mathbf{R}$$

随机变量 $N=\min(X_1,X_2,\cdots,X_n)$ 的分布函数为

$$F_{\min}(z)=1-[1-F_{X_1}(z)][1-F_{X_2}(z)]\cdots[1-F_{X_n}(z)]=1-\prod_{i=1}^{n}[1-F_{X_i}(z)],\ z\in\mathbf{R}$$

当随机变量 $X_1,X_2,\cdots,X_n$ 相互独立且具有相同的分布函数 $F(x)$ 时，有

$$F_{\max}(z)=[F(z)]^n,\ z\in\mathbf{R}$$
$$F_{\min}(z)=1-[1-F(z)]^n,\ z\in\mathbf{R}$$

4 随机变量的数字特征

4.1 数学期望

4.1.1 数学期望的概念

定义 4.1 设离散型随机变量 X 的概率分布为 $P\{X=x_k\}=p_k$，$k=1,2,\cdots$，如果无穷级数 $\sum\limits_{k=1}^{\infty} x_k p_k$ 绝对收敛，则称无穷级数 $\sum\limits_{k=1}^{\infty} x_k p_k$ 的和为离散型随机变量 X 的**数学期望**（mathematic expectation）或**均值**，记作 $E(X)$ 或 EX，即

$$E(X)=\sum_{k=1}^{\infty} x_k p_k$$

设连续型随机变量 X 的概率密度为 $f(x)$，如果反常积分 $\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x$ 绝对收敛，则称反常积分 $\int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x$ 的值为连续型随机变量 X 的**数学期望**或**均值**，记作 $E(X)$ 或 EX，即

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

任给一个随机变量，其数学期望未必存在.

结论 若随机变量 X 服从参数为 p 的 (0-1) 分布，则 $E(X)=p$.

证明： X 的分布律为

X	0	1
P	$1-p$	p

因此

$$E(X)=0\times(1-p)+1\times p=p$$

□

结论 若随机变量 $X\sim B(n,p)$，则 $E(X)=np$.

证明： X 的概率分布为 $P\{X=k\}=C_n^kp^k(1-p)^{n-k}$, $k=0,1,2,\cdots,n$，因此

$$\begin{aligned}E(X)&=\sum_{k=0}^{n}kC_n^kp^k(1-p)^{n-k}\\&=\sum_{k=0}^{n}\frac{kn!}{k!(n-k)!}p^k(1-p)^{n-k}\\&=\sum_{k=1}^{n}\frac{np(n-1)!\,p^{k-1}(1-p)^{(n-1)-(k-1)}}{(k-1)!\,[(n-1)-(k-1)]!}\\&=np[p+(1-p)]^{n-1}\\&=np\end{aligned}$$

□

结论 若随机变量 $X\sim P(\lambda)$，则 $E(X)=\lambda$.

证明： X 的概率分布为 $P\{X=k\}=\dfrac{\lambda^ke^{-\lambda}}{k!}$, $k=0,1,2,\cdots$，因此

$$\begin{aligned}E(X)&=\sum_{k=0}^{\infty}k\frac{\lambda^ke^{-\lambda}}{k!}\\&=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\\&=\lambda e^{-\lambda}e^{\lambda}\\&=\lambda\end{aligned}$$

□

结论 若随机变量 X 服从参数为 p 的几何分布，则 $E(X)=\dfrac{1}{p}$.

证明： X 的概率分布为 $P\{X=k\}=(1-p)^{k-1}p,\ k=1,2,\cdots$，则

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&\xlongequal{q=1-p} p\sum_{k=1}^{\infty} kq^{k-1} \\
&= p\left(\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty} q^k\right) \\
&= p\left(\frac{\mathrm{d}}{\mathrm{d}q}\frac{q}{1-q}\right) \\
&= p\frac{1}{(1-q)^2} \\
&= p\frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

□

结论 若随机变量 X 在区间 $[a,b]$ 上服从均匀分布，则 $E(X)=\dfrac{a+b}{2}$.

证明： X 的概率密度为

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

因此

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x=\int_a^b \frac{x}{b-a}\mathrm{d}x=\frac{a+b}{2}$$

□

结论 若随机变量 X 服从参数为 $\lambda\,(\lambda > 0)$ 的指数分布，则 $E(X) = \dfrac{1}{\lambda}$.

证明： X 的概率密度为

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

因此

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x \\ &= \int_{0}^{+\infty} x\lambda e^{-\lambda x}\,\mathrm{d}x \\ &= -xe^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\,\mathrm{d}x \\ &= -\frac{1}{\lambda}e^{-\lambda x}\Bigg|_0^{+\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

□

结论 若随机变量 $X \sim N(\mu, \sigma^2)$，则 $E(X) = \mu$.

证明： X 的概率密度为

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

因此

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

□

4.1.2 随机变量函数的数学期望

定理 4.1 设随机变量 Y 是随机变量 X 的函数，$Y = g(X)$，其中 g 是一元连续函数.

若 X 是离散型随机变量，其概率分布为 $P\{X = x_k\} = p_k,\ k = 1, 2, \cdots$，如果无穷级数 $\sum\limits_{k=1}^{\infty} g(x_k)p_k$ 绝对收敛，则随机变量 Y 的数学期望为

$$E(Y) = E[g(X)] = \sum_{k=1}^{\infty} g(x_k)\, p_k$$

若 X 是连续型随机变量，其概率密度为 $f(x)$，如果反常积分 $\int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x$ 绝对收敛，则随机变量 Y 的数学期望为

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

定理 4.2 设随机变量 Z 是随机变量 X 和 Y 的函数，$Z = g(X, Y)$，其中 g 是二元连续函数.

若 (X, Y) 是二维离散型随机变量，其概率分布为 $P\{X = x_i, Y = y_j\} = p_{ij},\ i, j = 1, 2, \cdots$，如果无穷级数 $\sum\limits_{j=1}^{\infty}\sum\limits_{i=1}^{\infty} g(x_i, y_j)\, p_{ij}$ 绝对收敛，则随机变量 Z 的数学期望为

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

若 (X, Y) 是二维连续型随机变量，其概率密度为 $f(x, y)$，如果反常积分 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y)f(x, y)\,\mathrm{d}x\mathrm{d}y$ 绝对收敛，则随机变量 Z 的数学期望为

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

4.1.3 数学期望的性质

设 C 为常数，X 和 Y 是随机变量，且 $E(X)$ 和 $E(Y)$ 都存在.

性质 1 $E(C)=C$

证明： 如果随机变量 X 恒取常数 C，则有 $P\{X=C\}=1$，从而有 $E(C)=C\times 1=C$. □

性质 2 $E(CX)=CE(X)$

证明： 若离散型随机变量 X 的概率分布为 $P\{X=x_i\}=p_i$, $i=1,2,\cdots$，则

$$E(CX)=\sum_{i=1}^{\infty}Cx_ip_i=C\sum_{i=1}^{\infty}x_ip_i=CE(X)$$

若连续型随机变量 X 的概率密度为 $f(x)$，则

$$E(CX)=\int_{-\infty}^{+\infty}Cxf(x)\,\mathrm{d}x=C\int_{-\infty}^{+\infty}xf(x)\,\mathrm{d}x=CE(X)$$

□

性质 3 $E(X+Y)=E(X)+E(Y)$

证明： 设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$，(X,Y) 关于 X 和关于 Y 的边缘概率密度分别为 $f_X(x)$ 和 $f_Y(y)$，则有

$$\begin{aligned}
E(X+Y)&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\,\mathrm{d}x\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\,\mathrm{d}x\mathrm{d}y+\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\,\mathrm{d}x\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}x\left[\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}y\right]\mathrm{d}x+\int_{-\infty}^{+\infty}y\left[\int_{-\infty}^{+\infty}f(x,y)\,\mathrm{d}x\right]\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}xf_X(x)\,\mathrm{d}x+\int_{-\infty}^{+\infty}yf_Y(y)\,\mathrm{d}y\\
&=E(X)+E(Y)
\end{aligned}$$

设二维离散型随机变量 (X,Y) 的概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}$, $i,j=1,2,\cdots$，(X,Y) 关于 X 的边缘概率分布为 $P\{X=x_i\}=p_{i\cdot}$, $i=1,2,\cdots$，(X,Y) 关于 Y 的边缘概率

分布为 $P\{Y=y_j\}=p_{\cdot j}$，$j=1,2,\cdots$，则有

$$
\begin{aligned}
E(X+Y) &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}(x_i+y_j)\,p_{ij} \\
&= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_ip_{ij}+\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}y_jp_{ij} \\
&= \sum_{i=1}^{\infty}\left(x_i\sum_{j=1}^{\infty}p_{ij}\right)+\sum_{j=1}^{\infty}\left(y_j\sum_{i=1}^{\infty}p_{ij}\right) \\
&= \sum_{i=1}^{\infty}x_ip_{i\cdot}+\sum_{j=1}^{\infty}y_jp_{\cdot j} \\
&= E(X)+E(Y)
\end{aligned}
$$

□

综合性质 1、2、3，有

$$E(aX+bY+c)=aE(X)+bE(Y)+c\quad (a,b,c\text{ 均为常数})$$

性质 4 若随机变量 X 和 Y 相互独立，则有 $E(XY)=E(X)\,E(Y)$.

证明： 设二维连续型随机变量 (X,Y) 的概率密度为 $f(x,y)$，(X,Y) 关于 X 和关于 Y 的边缘概率密度分别为 $f_X(x)$ 和 $f_Y(y)$. 因为 X 和 Y 相互独立，所以 $f(x,y)=f_X(x)\,f_Y(y)$，从而有

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_X(x)\,f_Y(y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}xf_X(x)\,\mathrm{d}x\int_{-\infty}^{+\infty}yf_Y(y)\,\mathrm{d}y \\
&= E(X)\,E(Y)
\end{aligned}
$$

设二维离散型随机变量 (X,Y) 的概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}$，$i,j=1,2,\cdots$，(X,Y) 关于 X 的边缘概率分布为 $P\{X=x_i\}=p_{i\cdot}$，$i=1,2,\cdots$，(X,Y) 关于 Y 的边缘概率

分布为 $P\{Y=y_j\}=p_{\cdot j}$，$j=1,2,\cdots$. 因为 X 和 Y 相互独立，所以 $p_{ij}=p_{i\cdot}\,p_{\cdot j}$，从而有

$$\begin{aligned}E(XY)&=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_iy_jp_{ij}\\&=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}x_iy_jp_{i\cdot}p_{\cdot j}\\&=\sum_{i=1}^{\infty}x_ip_{i\cdot}\sum_{j=1}^{\infty}y_jp_{\cdot j}\\&=E(X)\,E(Y)\end{aligned}$$

□

性质 5（柯西-施瓦茨不等式）　对于两个随机变量 X 和 Y，设 $E(X^2)$ 和 $E(Y^2)$ 都存在，则

$$[E(XY)]^2\leqslant E(X^2)E(Y^2)$$

证明：对于任意实数 t，令 $g(t)=E[(X+tY)^2]$，则由数学期望的性质有

$$g(t)=E[(X+tY)^2]=E(X^2+2tXY+t^2Y^2)=E(X^2)+2tE(XY)+t^2E(Y^2)$$

由于 $g(t)\geqslant 0$，所以有

$$\varDelta=4[E(XY)]^2-4E(X^2)E(Y^2)\leqslant 0$$

从而

$$[E(XY)]^2\leqslant E(X^2)E(Y^2)$$

□

4.2 方差

4.2.1 方差的概念

定义 4.2 设 X 是一个随机变量，如果 $E([X-E(X)]^2)$ 存在，则称之为随机变量 X 的**方差**（variance），记作 $D(X)$ 或 DX，即

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

称 $\sqrt{D(X)}$ 为随机变量 X 的**标准差**（standard deviation）或**均方差**，记作 $\sigma(X)$，即

$$\sigma(X)=\sqrt{D(X)}$$

随机变量 X 的方差反映了 X 与其数学期望 $E(X)$ 的偏离程度. 如果 X 取值集中在 $E(X)$ 附近，则 $D(X)$ 较小；如果 X 取值比较分散，则 $D(X)$ 较大.

如果 X 是离散型随机变量，概率分布为 $P\{X=x_k\}=p_k$, $k=1,2,\cdots$，则由定义4.2，有

$$D(X)=E([X-E(X)]^2)=\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k$$

如果 X 是连续型随机变量，其概率密度为 $f(x)$，则由定义4.2，有

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

结论（方差的计算公式） $D(X)=E(X^2)-[E(X)]^2$

证明：

$$\begin{aligned}D(X)&=E([X-E(X)]^2)\\&=E(X^2-2XE(X)+[E(X)]^2)\\&=E(X^2)-2[E(X)]^2+[E(X)]^2\\&=E(X^2)-[E(X)]^2\end{aligned}$$

□

结论 由于 $D(X)=E(X^2)-[E(X)]^2\geqslant 0$，因此 $E(X^2)\geqslant[E(X)]^2$.

4.2.2 方差的性质

设 C 是常数，随机变量 X 和 Y 的方差 $D(X)$ 和 $D(Y)$ 都存在.

性质 1 $D(C)=0$

证明：

$$D(C)=E([C-E(C)]^2)=E([C-C]^2)=E(0)=0$$

□

性质 2 $D(CX)=C^2D(X)$

证明：

$$\begin{aligned}D(CX)&=E([CX-E(CX)]^2)\\&=E([CX-CE(X)]^2)\\&=E(C^2[X-E(X)]^2)\\&=C^2E([X-E(X)]^2)\\&=C^2D(X)\end{aligned}$$

□

性质 3 $D(X+C)=D(X)$

证明：

$$\begin{aligned}D(X+C)&=E([(X+C)-E(X+C)]^2)\\&=E([X+C-E(X)-C]^2)\\&=E([X-E(X)]^2)\\&=D(X)\end{aligned}$$

□

性质 4 如果随机变量 X 和 Y 相互独立，则有 $D(X \pm Y) = D(X) + D(Y)$.

证明：

$$\begin{aligned} D(X \pm Y) &= E([X \pm Y - E(X \pm Y)]^2) \\ &= E([(X - E(X)) \pm (Y - E(Y))]^2) \\ &= E([X - E(X)]^2) \pm 2E([X - E(X)][Y - E(Y)]) + E([Y - E(Y)]^2) \end{aligned}$$

因为 X 与 Y 相互独立，故 $X - E(X)$ 与 $Y - E(Y)$ 也相互独立，又 $E(X - E(X)) = 0$，再由数学期望的性质 4 及方差的定义，得

$$\begin{aligned} D(X \pm Y) &= D(X) \pm 2E(X - E(X))\, E(Y - E(Y)) + D(Y) \\ &= D(X) + D(Y) \end{aligned}$$

□

性质 5 随机变量 X 的方差 $D(X) = 0$ 的充分必要条件是 X 以概率 1 取常数 C，即 $P\{X = C\} = 1$，其中 $C = E(X)$.

结论 对于任意常数 C，有 $E[(X - C)^2] \geqslant E([X - E(X)]^2) = D(X)$，当且仅当 $C = E(X)$ 时等号成立，此时 $E[(X - C)^2]$ 取得最小值.

4.2.3 常见概率分布的方差

结论 若随机变量 X 服从参数为 p 的 (0-1) 分布，则 $D(X) = p(1 - p)$.

证明： X 服从参数为 p 的 (0-1) 分布，则 $E(X) = p$，所以

$$\begin{aligned} D(X) &= E([X - E(X)]^2) \\ &= (0 - p)^2(1 - p) + (1 - p)^2 p \\ &= p(1 - p) \end{aligned}$$

□

结论 若随机变量 $X \sim B(n,p)$，则 $D(X) = np(1-p)$.

证明： 根据二项分布的意义可知，p 为 n 重伯努利试验中每次试验成功的概率. 引入随机变量

$$X_k = \begin{cases} 1 & \text{第 } k \text{ 次试验成功} \\ 0 & \text{第 } k \text{ 次试验不成功} \end{cases} \qquad k = 1, 2, \cdots, n$$

则有

$$X = X_1 + X_2 + \cdots + X_n$$

由于 X_k 只依赖于第 k 次试验，而各次试验相互独立，于是 $X_1, X_2, \cdots, X_n$ 相互独立，且均服从参数为 p 的 (0-1) 分布. 由于 $D(X_i) = p(1-p)$, $i = 1, 2, \cdots, n$，所以根据方差的性质 4，有

$$D(X) = \sum_{i=1}^{n} D(X_i) = np(1-p)$$

□

结论 若随机变量 $X \sim P(\lambda)$，则 $D(X) = \lambda$.

证明： 若随机变量 $X \sim P(\lambda)$，则

$$E(X^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}$$

记 $i = k - 1$，则

$$\begin{aligned} E(X^2) &= \lambda \sum_{i=0}^{\infty} (i+1) \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \lambda \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} + \lambda \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \\ &= \lambda^2 + \lambda \end{aligned}$$

由于 $E(X) = \lambda$，因此

$$D(X) = E(X^2) - [E(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

□

结论 若随机变量 X 服从参数为 p 的几何分布，则 $D(X)=\dfrac{1-p}{p^2}$.

证明：随机变量 X 服从参数为 p 的几何分布，则 $E(X)=\dfrac{1}{p}$，所以

$$\begin{aligned}
D(X) &= E(X^2)-[E(X)]^2 \\
&= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p-\frac{1}{p^2} \\
&= p\left[\sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1}-\sum_{k=1}^{\infty}k(1-p)^{k-1}\right]-\frac{1}{p^2} \\
&\xlongequal{q=1-p} p\left[\sum_{k=1}^{\infty}(k+1)kq^{k-1}-\sum_{k=1}^{\infty}kq^{k-1}\right]-\frac{1}{p^2} \\
&= p\left(\frac{\mathrm{d}^2}{\mathrm{d}q^2}\sum_{k=1}^{\infty}q^{k+1}-\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty}q^k\right)-\frac{1}{p^2} \\
&= p\left(\frac{\mathrm{d}^2}{\mathrm{d}q^2}\frac{q^2}{1-q}-\frac{\mathrm{d}}{\mathrm{d}q}\frac{q}{1-q}\right)-\frac{1}{p^2} \\
&= p\left[\frac{2}{(1-q)^3}-\frac{1}{(1-q)^2}\right]-\frac{1}{p^2} \\
&= p\left(\frac{2}{p^3}-\frac{1}{p^2}\right)-\frac{1}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}$$

□

结论 若随机变量 X 在区间 $[a,b]$ 上服从均匀分布，则 $D(X)=\dfrac{(b-a)^2}{12}$.

证明：若随机变量 X 在区间 $[a,b]$ 上服从均匀分布，则

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\,\mathrm{d}x=\int_a^b x^2\frac{1}{b-a}\mathrm{d}x=\left.\frac{x^3}{3(b-a)}\right|_a^b=\frac{b^2+ab+a^2}{3}$$

由于 $E(X)=\dfrac{a+b}{2}$，因此

$$D(X)=E(X^2)-[E(X)]^2=\frac{b^2+ab+a^2}{3}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}$$

□

结论 若随机变量 X 服从参数为 $\lambda\,(\lambda > 0)$ 的指数分布，则 $D(X) = \dfrac{1}{\lambda^2}$.

证明：由于 $E(X) = \dfrac{1}{\lambda}$，则

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x \\ &= \int_{0}^{+\infty} x^2 \lambda e^{-\lambda x}\,\mathrm{d}x \\ &= -x^2 e^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} 2x e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{2}{\lambda}\int_0^{+\infty} x\lambda e^{-\lambda x}\,\mathrm{d}x \\ &= \frac{2}{\lambda}E(X) \\ &= \frac{2}{\lambda^2} \end{aligned}$$

因此 $D(X) = E(X^2) - [E(X)]^2 = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$. □

结论 若随机变量 $X \sim N(\mu, \sigma^2)$，则 $D(X) = \sigma^2$.

证明：随机变量 $X \sim N(\mu, \sigma^2)$，则 $E(X) = \mu$，因此

$$\begin{aligned} D(X) &= E([X - E(X)]^2) \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x \\ &\xlongequal{t=\frac{x-\mu}{\sigma}} \sigma^2 \int_{-\infty}^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t \\ &= -\sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\,\mathrm{d}e^{-\frac{t^2}{2}} \\ &= -\frac{\sigma^2}{\sqrt{2\pi}} \left(te^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\,\mathrm{d}t \right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}}\,\mathrm{d}t \\ &= \sigma^2 \end{aligned}$$

□

分布	符号	概率分布/概率密度	数学期望	方差
(0-1) 分布	$B(1,p)$	$P\{X=k\}=p^k(1-p)^{1-k},\ k=0,1,\ 0<p<1$	p	$p(1-p)$
二项分布	$B(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k},\ k=0,1,\cdots,n$	np	$np(1-p)$
泊松分布	$P(\lambda)$	$P\{X=k\}=\dfrac{\lambda^k e^{-\lambda}}{k!},\ k=0,1,2\cdots$	λ	λ
几何分布	-	$P\{X=k\}=(1-p)^{k-1}p,\ k=1,2,\cdots$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
均匀分布	$U(a,b)$	$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其他}\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
指数分布	-	$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\leqslant 0\end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
正态分布	$N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<+\infty$	μ	σ^2

4.2.4 随机变量的标准化

设随机变量 X 具有数学期望 $E(X)=\mu$ 及方差 $D(X)=\sigma^2>0$，则称 $X^*=\dfrac{X-\mu}{\sigma}$ 为 X 的**标准化随机变量**.

随机变量 X 的标准化随机变量 X^* 满足 $E(X^*)=0$，$D(X^*)=1$.

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

4.3 协方差与相关系数

4.3.1 协方差

在方差性质 4 的证明中可知，如果两个随机变量 X 和 Y 相互独立，则有

$$E([X-E(X)][Y-E(Y)])=0$$

这表明，当 $E([X-E(X)][Y-E(Y)])\neq 0$ 时，X 与 Y 不相互独立，因此可以用这个量来描述 X 和 Y 之间的关系.

定义 4.3 设随机变量 X 和 Y 的数学期望 $E(X)$ 和 $E(Y)$ 都存在，如果 $E([X-E(X)][Y-E(Y)])$ 存在，则称之为随机变量 X 和 Y 的**协方差**（covariance），记作 $\mathrm{Cov}(X,Y)$，即

$$\mathrm{Cov}(X,Y)=E([X-E(X)][Y-E(Y)])=E(XY)-E(X)E(Y)$$

性质 1 $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$

性质 2 对于常数 a 和 b，有 $\mathrm{Cov}(aX,bY)=ab\,\mathrm{Cov}(X,Y)$.

证明：

$$\begin{aligned}\mathrm{Cov}(aX,bY)&=E(aX\cdot bY)-E(aX)E(bY)\\&=abE(XY)-abE(X)E(Y)\\&=ab[E(XY)-E(X)E(Y)]\\&=ab\,\mathrm{Cov}(X,Y)\end{aligned}$$

□

性质 3 对于随机变量 X,Y 和 Z，有

$$\mathrm{Cov}(X+Y,Z)=\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)$$

证明：

$$\begin{aligned}\mathrm{Cov}(X+Y,Z)&=E((X+Y)Z)-E(X+Y)E(Z)\\&=E(XZ)+E(YZ)-E(X)E(Z)-E(Y)E(Z)\\&=[E(XZ)-E(X)E(Z)]+[E(YZ)-E(Y)E(Z)]\\&=\mathrm{Cov}(X,Z)+\mathrm{Cov}(Y,Z)\end{aligned}$$

□

性质 4 对任意随机变量 X, Y，有

$$D(X \pm Y) = D(X) + D(Y) \pm 2\,\mathrm{Cov}(X, Y)$$

4.3.2 相关系数

定义 4.4 设随机变量 X 和 Y 的方差都存在且不等于零，协方差 $\mathrm{Cov}(X, Y)$ 存在，称 $\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 为随机变量 X 和 Y 的**相关系数**（correlation coefficient），记作 ρ_{XY}，即

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

当 $\rho_{XY} = 0$ 时，称 X 与 Y **不相关**.

性质 1 $|\rho_{XY}| \leqslant 1$

证明： 由柯西-施瓦茨不等式可得

$$\begin{aligned}[\mathrm{Cov}(X,Y)]^2 &= [E([X - E(X)][Y - E(Y)])]^2 \\ &\leqslant E([X - E(X)]^2)E([Y - E(Y)]^2) \\ &= D(X)D(Y)\end{aligned}$$

因此 $|\mathrm{Cov}(X,Y)| \leqslant \sqrt{D(X)}\sqrt{D(Y)}$，从而有

$$|\rho_{XY}| = \left|\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}\right| \leqslant 1$$

□

性质 2 如果随机变量 X 和 Y 相互独立，则 $\rho_{XY} = 0$.

证明： 若 X 和 Y 相互独立，则 $\mathrm{Cov}(X,Y) = 0$，进而 $\rho_{XY} = 0$. □

性质 3 $|\rho_{XY}|=1$ 的充分必要条件是：存在常数 a,b，使得 $P\{Y=a+bX\}=1$.

证明：设 $D(X)=\sigma_X^2>0$，$D(Y)=\sigma_Y^2>0$. 对于任意实数 b，有

$$\begin{aligned}D(Y-bX)&=E([Y-bX-E(Y-bX)]^2)\\&=E(\{[Y-E(Y)]-b[X-E(X)]\}^2)\\&=E([Y-E(Y)]^2)-2bE([Y-E(Y)][X-E(X)])+b^2E([X-E(X)]^2)\\&=\sigma_Y^2-2b\operatorname{Cov}(X,Y)+b^2\sigma_X^2\end{aligned}$$

取 $b=\dfrac{\operatorname{Cov}(X,Y)}{\sigma_X^2}$，则有

$$\begin{aligned}D(Y-bX)&=\sigma_Y^2-\frac{2[\operatorname{Cov}(X,Y)]^2}{\sigma_X^2}+\frac{[\operatorname{Cov}(X,Y)]^2}{\sigma_X^2}\\&=\sigma_Y^2-\frac{[\operatorname{Cov}(X,Y)]^2}{\sigma_X^2}\\&=\sigma_Y^2\left\{1-\frac{[\operatorname{Cov}(X,Y)]^2}{\sigma_X^2\sigma_Y^2}\right\}\\&=\sigma_Y^2(1-\rho_{XY}^2)\end{aligned}$$

由此得 $|\rho_{XY}|=1$ 的充分必要条件是 $D(Y-bX)=0$. 根据方差的性质 5，$D(Y-bX)=0$ 的充分必要条件是 $Y-bX$ 以概率 1 取常数 $a=E(Y-bX)$，即

$$P\{Y-bX=a\}=1$$

亦即

$$P\{Y=a+bX\}=1$$

□

相关系数表示随机变量 X 和 Y 线性相关的程度. 当 $|\rho_{XY}|=1$ 时，X 与 Y 之间以概率 1 存在线性关系；当 $|\rho_{XY}|$ 较大时，称 X 与 Y 线性相关的程度较好；当 $|\rho_{XY}|$ 较小时，称 X 与 Y 线性相关的程度较差. 当 $\rho_{XY}>0$ 时，称 X 与 Y 正相关，这时随着 X 的增加，Y 的值也有增加的趋势；当 $\rho_{XY}<0$ 时，称 X 与 Y 负相关，这时随着 X 的增加，Y 的值有减小的趋势.

如果 X 与 Y 相互独立，则 $\rho_{XY}=0$，即 X 与 Y 不相关. 反之，如果 X 与 Y 不相关，

则 X 和 Y 之间不存在线性关系，但 X 与 Y 未必独立，二者可能存在其他关系.

对于随机变量 X 和 Y，下列命题是等价的：

1. $\mathrm{Cov}(X,Y)=0$.
2. X 与 Y 不相关.
3. $E(XY)=E(X)E(Y)$.
4. $D(X+Y)=D(X)+D(Y)$.

结论 若二维随机变量 (X,Y) 服从二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X 与 Y 的相关系数 $\rho_{XY}=\rho$.

证明： (X,Y) 的概率密度为

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

(X,Y) 关于 X 和关于 Y 的边缘概率密度分别为

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

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

因此 $E(X)=\mu_1,E(Y)=\mu_2,D(X)=\sigma_1^2,D(Y)=\sigma_2^2$.

$$\begin{aligned}\mathrm{Cov}(X,Y)&=E([X-E(X)][Y-E(Y)])\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)\,\mathrm{d}x\mathrm{d}y\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)e^{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]}\mathrm{d}x\mathrm{d}y\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)^2}\mathrm{d}x\mathrm{d}y\end{aligned}$$

令 $t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right)$，$u=\frac{x-\mu_1}{\sigma_1}$，则有

$$\begin{aligned}\mathrm{Cov}(X,Y)&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\sigma_1\sigma_2(\sqrt{1-\rho^2}tu+\rho u^2)e^{-\frac{t^2}{2}}e^{-\frac{u^2}{2}}\mathrm{d}t\mathrm{d}u\\&=\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\int_{-\infty}^{+\infty}te^{-\frac{t^2}{2}}\mathrm{d}t\int_{-\infty}^{+\infty}ue^{-\frac{u^2}{2}}\mathrm{d}u+\\&\quad\rho\sigma_1\sigma_2\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}u^2e^{-\frac{u^2}{2}}\mathrm{d}u\end{aligned}$$

而

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}}\mathrm{d}t = -e^{-\frac{t^2}{2}}\Big|_{-\infty}^{+\infty} = 0 \\
&\int_{-\infty}^{+\infty} ue^{-\frac{u^2}{2}}\mathrm{d}u = 0 \\
&\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t = 1 \\
&\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}u^2e^{-\frac{u^2}{2}}\mathrm{d}u = 1
\end{aligned}
$$

因此

$$\mathrm{Cov}(X,Y) = \rho\sigma_1\sigma_2$$

所以

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho$$

□

若 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 X 与 Y 相互独立的充分必要条件是 $\rho=0$. 由于 $\rho_{XY}=\rho$，所以当 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 时，X 与 Y 相互独立的充分必要条件是 X 与 Y 不相关.

由于二维正态随机变量 (X,Y) 的概率密度中的参数 ρ 就是 X 与 Y 的相关系数，因此二维正态随机变量 (X,Y) 的分布完全由 X 和 Y 的数学期望、方差以及 X 与 Y 的相关系数确定.

4.4 矩

4.4.1 矩的概念

定义 4.5 设 X 和 Y 是随机变量. 如果 $E(X^k)$, $k=1,2,\cdots$ 存在，则称之为随机变量 X 的 k **阶原点矩**，记作 $E(X^k)=\mu_k\,(k=1,2,\cdots)$.

如果 $E([X-E(X)]^k)$, $k=1,2,\cdots$ 存在，则称之为随机变量 X 的 k **阶中心矩**.

如果 $E(X^kY^l)$, $k,l=1,2,\cdots$ 存在，则称之为随机变量 X 和 Y 的 $k+l$ **阶混合原点矩**.

如果 $E([X-E(X)]^k[Y-E(Y)]^l),\ k,l=1,2,\cdots$ 存在，则称之为随机变量 X 和 Y 的 $k+l$ **阶混合中心矩**.

随机变量 X 的数学期望 $E(X)$ 是 X 的一阶原点矩，方差 $D(X)$ 是 X 的二阶中心矩，随机变量 X 和 Y 的协方差 $\mathrm{Cov}(X,Y)$ 是 X 和 Y 的二阶混合中心矩.

4.4.2 协方差矩阵

定义 4.6 设二维随机变量 (X_1,X_2) 关于 X_1 和 X_2 的二阶中心距和二阶混合中心距

$$c_{ij}=E([X_i-E(X_i)][X_j-E(X_j)]),\ i,j=1,2$$

都存在，则称矩阵

$$\boldsymbol{C}=\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

为二维随机变量 (X_1,X_2) 的**协方差矩阵**.

设 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 关于 $X_1,X_2,\cdots,X_n$ 的二阶中心距和二阶混合中心距

$$c_{ij}=E([X_i-E(X_i)][X_j-E(X_j)]),\ i,j=1,2,\cdots,n$$

都存在，则称矩阵

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

为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的**协方差矩阵**.

由于 $c_{ij}=c_{ji}(i\neq j,\ i,j=1,2,\cdots,n)$，所以 $\boldsymbol{C}$ 是对称矩阵.

4.4.3 n 维正态分布

设二维随机变量 $(X_1,X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则

$$E(X_1)=\mu_1,\ E(X_2)=\mu_2,\ D(X_1)=\sigma_1^2,\ D(X_2)=\sigma_2^2$$

又由于 $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1) = \rho\sigma_1\sigma_2$，从而有

$$c_{11} = \sigma_1^2,\ c_{12} = c_{21} = \rho\sigma_1\sigma_2,\ c_{22} = \sigma_2^2$$

所以 (X_1, X_2) 的协方差矩阵为

$$\boldsymbol{C} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

其行列式 $|\boldsymbol{C}| = \sigma_1^2\sigma_2^2(1-\rho^2)$，$\boldsymbol{C}$ 的逆矩阵为

$$\boldsymbol{C}^{-1} = \frac{1}{|\boldsymbol{C}|}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

令

$$\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},\ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

则

$$\begin{aligned}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) &= \frac{1}{|\boldsymbol{C}|}\begin{bmatrix} x_1-\mu_1 & x_2-\mu_2 \end{bmatrix}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \\ &= \frac{1}{1-\rho^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\end{aligned}$$

因此二维正态随机变量 (X_1, X_2) 的概率密度可以写成

$$f(x_1, x_2) = \frac{1}{(2\pi)^{\frac{2}{2}}|\boldsymbol{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}$$

设 $(X_1, X_2, \cdots, X_n)$ 为 n 维随机变量，记

$$\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},\ \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{bmatrix}$$

如果 $(X_1, X_2, \cdots, X_n)$ 具有概率密度

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})}$$

其中 $\boldsymbol{C}$ 为 $(X_1, X_2, \cdots, X_n)$ 的协方差矩阵，则称 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布.

性质 1 n 维随机变量 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布的充分必要条件是：$X_1, X_2, \cdots, X_n$ 的任意线性组合 $k_1X_1 + k_2X_2 + \cdots + k_nX_n$ 都服从一维正态分布，其中 $k_1, k_2, \cdots, k_n$ 是不全为零的常数.

性质 2 设 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布. 如果 $Y_1, Y_2, \cdots, Y_m$ 是 $X_i\,(i = 1, 2, \cdots, n)$ 的线性函数，则 $(Y_1, Y_2, \cdots, Y_m)$ 也服从多维正态分布.

性质 3 如果 $(X_1, X_2, \cdots, X_n)$ 服从 n 维正态分布，则“随机变量 $X_1, X_2, \cdots, X_n$ 相互独立”与“$X_1, X_2, \cdots, X_n$ 两两不相关”等价.

5 大数定律与中心极限定理

5.1 切比雪夫不等式

定理 5.1（切比雪夫不等式） 设随机变量 X 具有数学期望 $E(X)=\mu$ 和方差 $D(X)=\sigma^2$，则对于任意给定的正数 ε，有

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

它的等价形式是

$$P\{|X-E(X)|<\varepsilon\}\geqslant 1-\frac{D(X)}{\varepsilon^2}$$

证明：如果 X 是离散型随机变量，设 X 的概率分布为 $P\{X=x_k\}=p_k$，$k=1,2,\cdots$，根据概率的可加性可得

$$P\{|X-\mu|\geqslant\varepsilon\}=\sum_{|x_k-\mu|\geqslant\varepsilon}P\{X=x_k\}=\sum_{|x_k-\mu|\geqslant\varepsilon}p_k$$

由 $|x_k-\mu|\geqslant\varepsilon$ 得 $(x_k-\mu)^2\geqslant\varepsilon^2$，即

$$\frac{(x_k-\mu)^2}{\varepsilon^2}\geqslant 1$$

从而有

$$\begin{aligned}\sum_{|x_k-\mu|\geqslant\varepsilon}p_k&\leqslant\sum_{|x_k-\mu|\geqslant\varepsilon}\frac{(x_k-\mu)^2}{\varepsilon^2}p_k\\&\leqslant\frac{1}{\varepsilon^2}\sum_{k=1}^{\infty}(x_k-\mu)^2p_k\\&=\frac{D(X)}{\varepsilon^2}\end{aligned}$$

因此

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

如果 X 是连续型随机变量，设 X 的概率密度为 $f(x)$，则

$$\begin{aligned}P\{|X-E(X)|\geqslant\varepsilon\}&=\int\limits_{|x-\mu|\geqslant\varepsilon} f(x)\,\mathrm{d}x\\&\leqslant\int\limits_{|x-\mu|\geqslant\varepsilon}\frac{(x-\mu)^2}{\varepsilon^2}f(x)\,\mathrm{d}x\\&\leqslant\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2f(x)\,\mathrm{d}x\\&=\frac{D(X)}{\varepsilon^2}\end{aligned}$$

□

切比雪夫不等式给出了在随机变量 X 的分布未知的情况下随机事件 $\{|X-\mu|<\varepsilon\}$ 的概率的一种估计. 例如，取 $\varepsilon=3\sigma$，则 $P\{|X-\mu|<3\sigma\}\geqslant 0.8889$.

5.2 大数定律

5.2.1 依概率收敛

定义 5.1 设 $X_1,X_2,\cdots,X_n,\cdots$ 是一个随机变量序列，a 是一个常数. 如果对于任意给定的正数 ε，有

$$\lim_{n\to\infty}P\{|X_n-a|<\varepsilon\}=1$$

则称随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ **依概率收敛**于 a，记作

$$X_n\xrightarrow{P}a$$

性质：设 $X_n\xrightarrow{P}a$，$Y_n\xrightarrow{P}b$，函数 $g(x,y)$ 在点 (a,b) 连续，则

$$g(X_n,Y_n)\xrightarrow{P}g(a,b)$$

5.2.2 大数定律

定理 5.2（切比雪夫定理） 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，分别具有数学期望 $E(X_1), E(X_2), \cdots, E(X_n), \cdots$ 及方差 $D(X_1), D(X_2), \cdots, D(X_n), \cdots$，并且方差是一致有上界的（即存在正数 M，使得 $D(X_n) \leqslant M$，$n = 1, 2, \cdots$），则对任意给定的正数 ε，恒有

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

证明：

$$E(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n}\sum_{k=1}^{n} E(X_k)$$

$$D(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k)$$

根据切比雪夫不等式，得

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right\} \geqslant 1 - \frac{\sum_{k=1}^{n} D(X_k)}{n^2\varepsilon^2}$$

由于方差一致有上界，因此

$$\sum_{k=1}^{n} D(X_k) \leqslant nM$$

从而得

$$1 \geqslant P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \varepsilon\right\} \geqslant 1 - \frac{M}{n\varepsilon^2}$$

令 $n \to \infty$，则有

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

□

利用依概率收敛的概念，切比雪夫定理可叙述成：设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，分别具有数学期望和方差，并且方差一致有上界，则 $X_1, X_2, \cdots, X_n$ 的算术平均值与它们的数学期望的算术平均值之差当 $n \to \infty$ 时依概率收敛于零.

推论 5.1 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立,并且具有相同的数学期望 $E(X_k)=\mu$ 和相同的方差 $D(X_k)=\sigma^2\,(k=1,2,\cdots)$,则 $X_1, X_2, \cdots, X_n$ 的算术平均值 $\dfrac{1}{n}\sum\limits_{k=1}^{n} X_k$ 当 $n\to\infty$ 时依概率收敛于数学期望 μ,即对任意给定的正数 ε,恒有

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

推论 5.1 是实际问题中使用算术平均值的依据. 为了测量某一个量 a,在相同的条件下重复测量 n 次,得到 n 个测量结果 $x_1, x_2, \cdots, x_n$,可以认为 $x_1, x_2, \cdots, x_n$ 分别是服从同一分布、有相同数学期望 μ 和方差 σ^2 的随机变量 $X_1, X_2, \cdots, X_n$ 的试验数值. 由推论 5.1 可知,当 n 充分大时,取这 n 次测量结果的算术平均值作为 a 的近似值,所发生的误差将很小.

定理 5.3(伯努利定理) 设 n_A 是在 n 次独立重复试验中事件 A 发生的次数,p 是事件 A 在一次试验中发生的概率,则对于任意给定的正数 ε,有

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

其等价形式为

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\geqslant\varepsilon\right\}=0$$

证明: n_A 是一个随机变量,且服从二项分布 $B(n,p)$,从而有 $E(n_A)=np$,$D(n_A)=np(1-p)$.

因为

$$E(\frac{n_A}{n})=\frac{E(n_A)}{n}=p$$
$$D(\frac{n_A}{n})=\frac{D(n_A)}{n^2}=\frac{p(1-p)}{n}$$

根据切比雪夫不等式,对任意给定的正数 ε,有

$$1\geqslant P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}\geqslant 1-\frac{p(1-p)}{n\varepsilon^2}$$

令 $n\to\infty$,则有

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

□

伯努利定理表明，一个事件 A 在 n 次独立重复试验中发生的频率 $\frac{n_A}{n}$ 当 $n \to \infty$ 时依概率收敛于事件 A 发生的概率 p. 伯努利定理以严格的数学形式表达了频率的稳定性.

伯努利定理的等价形式表明，当 n 很大时，事件 A 在 n 次独立重复试验中发生的频率与 A 在一次试验中发生的概率有较大偏差的可能性很小. 根据实际推断原理，在实际应用中，当试验次数 n 很大时，可以利用事件 A 发生的频率代替事件 A 发生的概率.

定理 5.4（辛钦定理） 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 μ，则对于任意给定的正数 ε，有

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

伯努利定理是辛钦定理的特殊情况.

由辛钦定理可知，如果随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布且具有数学期望 μ，则前 n 个随机变量的算术平均值 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 依概率收敛于它们的数学期望 μ. 如果 $E(X_k^l) = \mu_l\,(k = 1, 2, \cdots)$ 存在，则 $\frac{1}{n}\sum_{k=1}^{n} X_k^l$ 依概率收敛于 $\mu_l\,(l = 1, 2, \cdots)$. 这是在数理统计中求参数点估计的矩估计法的理论基础.

5.3 中心极限定理

定义 5.2 设随机变量 $X, X_1, X_2, \cdots, X_n, \cdots$ 的分布函数依次是

$$F(x), F_1(x), F_2(x), \cdots, F_n(x), \cdots$$

如果对于 $F(x)$ 的每一个连续点 x，都有

$$\lim_{n\to\infty} F_n(x) = F(x)$$

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ **依分布收敛于** X，记作

$$X_n \xrightarrow{L} X$$

定理 5.5（独立同分布的中心极限定理） 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从相同的分布，且具有数学期望 $E(X_k) = \mu$ 和方差 $D(X_k) = \sigma^2 \neq 0\,(k = 1, 2, \cdots)$，随机变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数为 $F_n(x)$，即

$$F_n(x) = P\{Y_n \leqslant x\} = P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \leqslant x\right\},\ -\infty < x < +\infty$$

则对任意实数 x，恒有

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

定理 5.5 也叫做**莱维-林德伯格定理**.

利用定义 5.2，独立同分布的中心极限定理可叙述成：如果随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从相同的分布，且具有数学期望 $E(X_k) = \mu$ 和方差 $D(X_k) = \sigma^2 \neq 0\,(k = 1, 2, \cdots)$，则随机变量序列

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma},\ n = 1, 2, \cdots$$

依分布收敛于服从标准正态分布的随机变量 u，即 $Y_n \xrightarrow{L} u$.

定理 5.6（棣莫弗-拉普拉斯极限定理） 设随机变量 $Y_n \sim B(n, p)\,(n = 1, 2, \cdots)$，则对任意实数 x，恒有

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

证明： 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且都服从 (0-1) 分布，其概率分布为

$$P\{X_k = 0\} = 1 - p,\ P\{X_k = 1\} = p,\ 0 < p < 1,\ k = 1, 2, \cdots$$

由于 $E(X_k) = p,\ D(X_k) = p(1-p)$，根据独立同分布的中心极限定理可知，对任意实数 x，恒有

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

随机变量 $Y_n = \sum\limits_{k=1}^{n} X_k$ 服从参数为 n, p 的二项分布，即 $Y_n \sim B(n, p)$，因此有

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

□

定理 5.6 表明，正态分布是二项分布的极限分布，当 n 充分大时，可以利用该定理近似计算二项分布的概率.

定理 5.7（李雅普诺夫定理） 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有数学期望和方差

$$E(X_k) = \mu_k,\ D(X_k) = \sigma_k^2 \neq 0,\ k = 1, 2, \cdots$$

记

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

设随机变量

$$Z_n = \frac{\sum\limits_{k=1}^{n} X_k - E(\sum\limits_{k=1}^{n} X_k)}{\sqrt{D(\sum\limits_{k=1}^{n} X_k)}} = \frac{\sum\limits_{k=1}^{n} X_k - \sum\limits_{k=1}^{n} \mu_k}{B_n}$$

Z_n 的分布函数为 $F_n(x)$，即

$$F_n(x) = P\left\{\frac{\sum\limits_{k=1}^{n} X_k - \sum\limits_{k=1}^{n} \mu_k}{B_n} \leqslant x\right\},\ -\infty < x < +\infty$$

如果存在正数 δ，使得当 $n \to \infty$ 时，有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E(|X_k - \mu_k|^{2+\delta}) \to 0$$

则对任意实数 x，恒有

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

在李雅普诺夫定理的条件下，当 n 很大时，随机变量

$$Z_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

近似服从标准正态分布 $N(0,1)$. 因此，当 n 很大时，$\sum_{k=1}^{n} X_k = B_n Z_n + \sum_{k=1}^{n} \mu_k$ 近似服从正态分布 $N(\sum_{k=1}^{n} \mu_k, B_n^2)$. 这就是说，无论各随机变量 $X_k\,(k = 1, 2, \cdots)$ 服从什么分布，只要满足李雅普诺夫定理的条件，当 n 很大时，这些随机变量的和 $\sum_{k=1}^{n} X_k$ 就近似服从正态分布.