

概率论与数理统计

赵子轩

2024 年 3 月 24 日

目录

第一章 随机事件及其概率	1
1.1 随机事件	1
1.1.1 随机试验	1
1.1.2 随机事件的概念	2
1.1.3 随机事件的关系	2
1.1.4 随机事件的运算	3
1.1.5 事件域	5
1.2 随机事件的概率	7
1.2.1 频率	7
1.2.2 概率	8
1.2.3 概率的连续性	12
1.2.4 古典概型	14
1.2.5 几何概型	17
1.3 条件概率	20
1.3.1 条件概率及其性质	20
1.3.2 乘法公式	22
1.3.3 全概率公式	24
1.3.4 贝叶斯公式	25
1.4 事件的独立性	26
1.5 伯努利概型	30
第二章 随机变量及其分布	31
2.1 随机变量及其分布函数	31
2.1.1 随机变量	31
2.1.2 分布函数	31
2.2 离散型随机变量及其概率分布	33
2.3 连续型随机变量及其概率密度	34
2.4 常用的概率分布	36
2.4.1 0-1 分布	36
2.4.2 二项分布	37
2.4.3 泊松分布	37

2.4.4	几何分布	39
2.4.5	超几何分布	39
2.4.6	帕斯卡分布	40
2.4.7	均匀分布	41
2.4.8	指数分布	41
2.4.9	正态分布	43
2.4.10	伽马分布	48
2.4.11	贝塔分布	52
2.4.12	柯西分布	53
2.5	随机变量的函数的分布	54
2.5.1	离散型随机变量的函数的分布	54
第三章	二维随机变量及其分布	55
3.1	二维随机变量	55
3.1.1	二维随机变量及其分布函数	55
3.1.2	边缘分布	56
3.1.3	随机变量的独立性	57
3.2	二维离散型随机变量	57
3.2.1	二维离散型随机变量及其概率分布	57
3.2.2	边缘概率分布	58
3.2.3	随机变量的独立性	58
3.3	二维连续型随机变量	59
3.3.1	二维连续型随机变量及其概率密度	59
3.3.2	边缘概率密度	60
3.3.3	随机变量的独立性	60
3.3.4	二维均匀分布	60
3.3.5	二维正态分布	60
3.4	条件分布	62
3.4.1	离散型随机变量的条件分布	62
3.4.2	连续型随机变量的条件分布	63
3.5	二维随机变量的函数的分布	64
3.5.1	二维离散型随机变量的函数的分布	64
3.5.2	二维连续型随机变量的函数的分布	65
3.6	n 维随机变量	68
第四章	随机变量的数字特征	73
4.1	数学期望	73
4.1.1	数学期望的定义	73
4.1.2	随机变量函数的数学期望	74
4.1.3	数学期望的性质	75

4.1.4	常见概率分布的数学期望	77
4.2	方差	85
4.2.1	方差的定义	85
4.2.2	方差的性质	86
4.2.3	切比雪夫不等式	89
4.2.4	常见概率分布的方差	91
4.2.5	随机变量的标准化	98
4.3	协方差与相关系数	98
4.3.1	协方差	98
4.3.2	相关系数	100
4.4	矩	103
4.4.1	矩的概念	103
4.4.2	协方差矩阵	103
4.4.3	n 维正态分布	104
附录		107
A.1	常见的概率分布	107

第一章 随机事件及其概率

1.1 随机事件

1.1.1 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** (random experiment), 简称为**试验** (experiment).

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** (sample point), 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** (sample space), 用 $\Omega = \{\omega\}$ 表示.

备注

1. 样本空间中的元素可以是数, 也可以不是数.
2. 随机现象的样本空间至少有 2 个样本点. 只有 1 个样本点的样本空间对应必然现象.
3. 根据样本点个数, 可以将样本空间分为有限与无限两类. 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**.
4. 另一种分类方式: 样本点个数为有限个或可列个时, 称为**离散样本空间**; 样本点个数为不可列无限个时, 称为**连续样本空间**.

1.1.2 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** (elementary event).

Ω 本身也是 Ω 的子集, 在每次试验中必然发生, 称为**必然事件** (certain event).

空集 \emptyset 也是 Ω 的子集, 在试验中不可能发生, 称为**不可能事件** (impossible event).

必然事件和不可能事件统称为**确定事件**.

1.1.3 随机事件的关系

1.1.3.1 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 , 或称事件 A 包含于事件 B , 记作 $A \subseteq B$.

性质 1.1.1

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

性质 1.1.2

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

1.1.3.2 事件的相等

如果事件 A 与 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 $A = B$.

1.1.3.3 事件的互斥

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 **互斥**, 或称事件 A 与事件 B **互不相容**.

任意两个基本事件一定互斥.

不可能事件 \emptyset 与任意事件互斥.

1.1.3.4 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**, 称事件 A 与事件 B 互为**逆事件**或**对立事件**, 记作 $\bar{A} = B$, 或 $\bar{B} = A$.

对立事件一定互斥. 反之, 互斥的事件不一定对立.

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

1.1.4 随机事件的运算

1.1.4.1 事件的并

如果事件 A 和事件 B 至少有一个发生, 则这样的—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 $A \cup B$ 就是集合 A 与 B 的并集.

性质 1.1.3

对于任意事件 A 与 B , 有:

1. $A \cup A = A$.
2. $A \cup \emptyset = A$.
3. $A \cup B = B \cup A$.
4. $A \cup \bar{A} = \Omega$.
5. $A \subseteq A \cup B$, $B \subseteq A \cup B$.
6. 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:

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

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

1.1.4.2 事件的交

如果事件 A 和事件 B 同时发生, 则这样的—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 $A \cap B$ 就是集合 A 与 B 的交集.

性质 1.1.4

对于任意事件 A 与 B , 有:

1. $A \cap A = A$.
2. $A \cap \emptyset = \emptyset$.
3. $A \cap B = B \cap A$.
4. $A \cap \bar{A} = \emptyset$.
5. $A \cap B \subseteq A$, $A \cap B \subseteq B$.
6. 如果 $A \subseteq B$, 则有 $A \cap B = A$.
7. 如果 A 与 B 互斥, 则有 $A \cap B = \emptyset$.

事件的交可以推广到多个事件的情形:

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

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

结论

事件 A 与 B 对立的充分必要条件是: $A \cap B = \emptyset$, 且 $A \cup B = \Omega$.

1.1.4.3 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 $A - B$.

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

事件 A 和事件 B 作为样本空间 Ω 的子集, 差事件 $A - B$ 就是集合 A 对 B 的差集.

性质 1.1.5

对于任意事件 A 与 B , 有:

1. $A - A = \emptyset$
2. $A - \emptyset = A$
3. $\emptyset - A = \emptyset$
4. $\Omega - A = \bar{A}$
5. $A - \Omega = \emptyset$
6. $A - B = A - AB = A\bar{B}$
7. $A \cup B = A \cup (B - A) = A \cup (B - AB)$
8. $A \cup B = B \cup (A - B) = B \cup (A - AB)$
9. $A - B, AB, B - A$ 两两互斥, 且
 - (1) $A \cup B = (A - B) \cup AB \cup (B - A)$
 - (2) $A = (A - B) \cup AB$
 - (3) $B = (B - A) \cup AB$

1.1.4.4 随机事件的运算性质

性质 1.1.6

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 (德摩根公式)

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

对于多个及可列个随机事件, 以上的运算性质也成立.

1.1.5 事件域

定义 1.1.1

设 Ω 为样本空间, \mathcal{F} 为 Ω 的某些子集组成的集合类. 如果 \mathcal{F} 满足:

1. $\Omega \in \mathcal{F}$;
2. 若 $A \in \mathcal{F}$, 则 $\overline{A} \in \mathcal{F}$;
3. 若 $A_i \in \mathcal{F}, i = 1, 2, \dots$, 则 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

则称 \mathcal{F} 为一个事件域, 也称为 σ 域或 σ 代数. 将 (Ω, \mathcal{F}) 称为可测空间 (measurable space).

在可测空间上才可定义概率, 这时 \mathcal{F} 中都是有概率可言的事件.

备注

事件域就是样本空间中某些子集及其运算结果组成的集合类. 对于离散样本空间, 它的所有子集就构成了事件域. 而对连续样本空间, 存在无法测量长度的子集, 这样的集合称为不可测集. 如果将不可测集也看成事件, 那么这样的事件将没有概率可言. 因此, 没有必要将连续样本空间的所有子集都看做事件, 只需将可度量的子集 (称为可测集) 看做事件即可.

事件域包含样本空间的所有可测子集, 这些子集都可以看做有概率可言的事件. 事件之间可以进行运算, 并且运算结果仍然是事件, 因此事件域要对事件的各种运算具有封闭性.

在事件的各种运算中, 并与对立是最基本的运算, 因为:

- 交运算可通过并与对立来实现 ($A \cap B = \overline{\overline{A} \cup \overline{B}}$).
- 差运算可通过对立与交来实现 ($A - B = A\overline{B}$).

所以, 只需事件域对并运算和对立运算封闭, 这就是定义 1.1.1 的来源.

若样本空间含有 n 个样本点, 则其事件域 \mathcal{F} 是由空集 \emptyset 、 n 个单元素集、 C_n^2 个双元素集、 C_n^3 个三元素集……和 Ω 组成的集合类, 此时 \mathcal{F} 中共有 $C_n^0 + C_n^1 + C_n^2 + \cdots + C_n^n = 2^n$ 个事件.

若样本空间含有可列个样本点, 则其事件域 \mathcal{F} 是由空集 \emptyset 、可列个单元素集、可列个双元素集……可列个 n 元素集……和 Ω 组成的集合类, 此时 \mathcal{F} 由可列个事件组成.

如果样本空间为全体实数, 即 $\Omega = \mathbf{R}$, 此时事件域 \mathcal{F} 中的元素无法一一列出, 而是由一个基本集合类逐步扩展形成, 具体操作如下:

1. 取基本集合类 \mathcal{P} 为“全体半直线组成的类”, 即

$$\mathcal{P} = \{(-\infty, x) \mid -\infty < x < +\infty\}$$

2. 把有限的左闭右开区间扩展进来

$$[a, b) = (-\infty, b) - (-\infty, a), \quad a, b \in \mathbf{R}$$

3. 把闭区间、单点集、左开右闭区间、开区间扩展进来

$$[a, b] = \bigcap_{i=1}^{\infty} \left[a, b + \frac{1}{i} \right)$$

$$\{b\} = [a, b] - [a, b)$$

$$(a, b] = [a, b] - \{a\}$$

$$(a, b) = [a, b) - \{a\}$$

4. 最后用(有限个或可列个)并运算和交运算把实数集中一切有限集、可列集、开集、闭集都扩展进来.

经过上述几步扩展所得集合的全体就是事件域 \mathcal{F} , 这样的事件域 \mathcal{F} 又称为**博雷尔事件域**, 域中的每个元素称为**博雷尔集**(Borel set).

性质 1.1.7

可测空间 (Ω, \mathcal{F}) 的性质:

1. $\emptyset \in \mathcal{F}$.

2. 若 $A_i \in \mathcal{F}, i = 1, 2, \dots$, 则 $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

3. 若 $A_i \in \mathcal{F}, i = 1, 2, \dots, n$, 则 $\bigcap_{i=1}^n A_i \in \mathcal{F}$.

4. 若 $A_i \in \mathcal{F}, i = 1, 2, \dots, n$, 则 $\bigcup_{i=1}^n A_i \in \mathcal{F}$.

5. 若 $A, B \in \mathcal{F}$, 则 $A - B \in \mathcal{F}$.

证明: (1) 由定义 1.1.1 可知, $\Omega \in \mathcal{F}$, 进而有

$$\overline{\Omega} = \emptyset \in \mathcal{F}$$

- (2) 若 $A_i \in \mathcal{F}$, 则 $\overline{A_i} \in \mathcal{F}$, 因此

$$\bigcup_{i=1}^{\infty} \overline{A_i} = \overline{\bigcap_{i=1}^{\infty} A_i} \in \mathcal{F}$$

进而有

$$\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$$

(3) 令 $A_i = \Omega$, $i = n+1, n+2, \dots$, 则由性质 2 可得

$$\bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^n A_i \in \mathcal{F}$$

(4) 令 $A_i = \emptyset$, $i = n+1, n+2, \dots$, 则由定义 1.1.1 可得

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^n A_i \in \mathcal{F}$$

(5) 因为 $B \in \mathcal{F}$, 所以 $\bar{B} \in \mathcal{F}$, 进而有

$$A - B = A\bar{B} \in \mathcal{F}$$

□

定义 1.1.2

对样本空间 Ω , 如果事件 A_1, A_2, \dots, A_n 互斥, 且 $\bigcup_{i=1}^n A_i = \Omega$, 则称 A_1, A_2, \dots, A_n 为样本空间 Ω 的一个分割或完全事件组.

若 $A = \{A_1, A_2, \dots, A_n\}$ 为样本空间 Ω 的一个分割, 由分割 A 产生的事件域记作 $\sigma(A)$, 则 $\sigma(A)$ 中含有 2^n 个不同的事件. 利用分割可以将事件域简化, 从而简化所研究的问题.

1.2 随机事件的概率

1.2.1 频率

定义 1.2.1

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 $\frac{n_A}{n}$ 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 1.2.1 (频率的基本性质)

1. 非负性: 对于任意事件 A , 有 $f_n(A) \geq 0$.
2. 规范性: 对于必然事件 Ω , 有 $f_n(\Omega) = 1$.
3. 有限可加性: 对于两两互斥的事件 A_1, A_2, \dots, A_m , 有

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

频率 $f_n(A)$ 依赖于试验次数 n 及每次试验的结果. 当 n 较小时, 频率的波动性一般较大. 当 n 增大时, 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.

1.2.2 概率

定义 1.2.2 (概率的公理化定义)

设 Ω 为一个样本空间, \mathcal{F} 为 Ω 的某些子集组成的事件域. 如果对于任一事件 $A \in \mathcal{F}$, 定义在 \mathcal{F} 上的一个实值函数 $P(A)$ 满足:

1. 非负性: 对于任意事件 $A \in \mathcal{F}$, 有 $P(A) \geq 0$;
2. 规范性: 对于必然事件 Ω , 有 $P(\Omega) = 1$;
3. 可列可加性: 对于两两互斥的事件 A_1, A_2, \dots , 有

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

则称集合函数 P 为可测空间 (Ω, \mathcal{F}) 上的概率 (probability), 称 $P(A)$ 为事件 A 的概率, 称三元素 (Ω, \mathcal{F}, P) 为概率空间 (probability space).

性质 1.2.2

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

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

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots$$

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

性质 1.2.3 (有限可加性)

对于两两互斥的事件 A_1, A_2, \dots, A_n , 有

$$P \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n P(A_i) \quad (1.2.1)$$

证明：根据概率的可列可加性及性质 1.2.2，有

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

□

性质 1.2.4

对于任一事件 A ，有

$$P(\bar{A}) = 1 - P(A) \quad (1.2.2)$$

证明：因为 $A \cup \bar{A} = \Omega$ ，且 $A\bar{A} = \emptyset$ ，由性质 1.2.3 及概率的规范性，得

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

即

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

□

性质 1.2.5

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

证明：因为 $A \subseteq B$ ，从而有 $B = A \cup (B - A)$ ，且 $A(B - A) = \emptyset$ ，由性质 1.2.3 可得

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

所以

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

□

推论 1.2.1 (概率的单调性)

如果 $A \subseteq B$ ，则有 $P(A) \leq P(B)$.

证明：由性质 1.2.5 可得，当 $A \subseteq B$ 时，有

$$P(B - A) = P(B) - P(A) \geq 0$$

因此 $P(A) \leq P(B)$.

□

备注

推论 1.2.1 的逆命题不成立，即当 $P(A) \leq P(B)$ 时 $A \subseteq B$ 未必成立.

性质 1.2.6

对于任一事件 A , 有 $P(A) \leq 1$.

证明: 因为 $A \subseteq \Omega$, 由推论 1.2.1 及概率的规范性可得

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

□

性质 1.2.7 (概率的减法公式)

对于任意两个事件 A 与 B , 有

$$P(A - B) = P(A) - P(AB) \quad (1.2.3)$$

证明: 由于 $A - B = A - AB$, 而 $AB \subseteq A$, 根据性质 1.2.5 可得

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

□

性质 1.2.8 (概率的加法公式)

对于任意两个事件 A 与 B , 有

$$P(A \cup B) = P(A) + P(B) - P(AB) \quad (1.2.4)$$

对任意 n 个事件 A_1, A_2, \dots, A_n , 有

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \dots + (-1)^{n-1} P(A_1 A_2 \cdots A_n) \quad (1.2.5)$$

证明: 先证式 (1.2.4). 因为 $A \cup B = A \cup (B - AB)$, 且 $A(B - AB) = \emptyset$, $AB \subseteq B$, 由性质 1.2.3 及性质 1.2.5 可得

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

式 (1.2.4) 得证.

对于式 (1.2.5), 使用数学归纳法. 当 $n = 2$ 时, 式 (1.2.5) 即为式 (1.2.4).

设式 (1.2.5) 对 $n - 1$ 成立, 则对于 n , 先对两个事件 $A_1 \cup A_2 \cup \dots \cup A_{n-1}$ 与 A_n 应用

式 (1.2.4):

$$\begin{aligned} & P(A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= P(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + P(A_n) - P((A_1 \cup A_2 \cup \cdots \cup A_{n-1})A_n) \\ &= P(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + P(A_n) - P((A_1A_n) \cup (A_2A_n) \cup \cdots \cup (A_{n-1}A_n)) \end{aligned}$$

由归纳假设, 将 $P(A_1 \cup A_2 \cup \cdots \cup A_{n-1})$ 和 $P((A_1A_n) \cup (A_2A_n) \cup \cdots \cup (A_{n-1}A_n))$ 展开, 得

$$\begin{aligned} & P(A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= \left[\sum_{i=1}^{n-1} P(A_i) - \sum_{1 \leq i < j \leq n-1} P(A_iA_j) + \sum_{1 \leq i < j < k \leq n-1} P(A_iA_jA_k) + \cdots + (-1)^{n-2} P(A_1A_2 \cdots A_{n-1}) \right] \\ &+ P(A_n) - \left[\sum_{i=1}^{n-1} P(A_iA_n) - \sum_{1 \leq i < j \leq n-1} P((A_iA_n)(A_jA_n)) + \sum_{1 \leq i < j < k \leq n-1} P((A_iA_n)(A_jA_n)(A_kA_n)) \right. \\ &\left. + \cdots + (-1)^{n-2} P((A_1A_n)(A_2A_n) \cdots (A_{n-1}A_n)) \right] \\ &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n-1} P(A_iA_j) + \sum_{1 \leq i < j < k \leq n-1} P(A_iA_jA_k) + \cdots + (-1)^{n-2} P(A_1A_2 \cdots A_{n-1}) - \\ &\sum_{i=1}^{n-1} P(A_iA_n) + \sum_{1 \leq i < j \leq n-1} P(A_iA_jA_n) - \sum_{1 \leq i < j < k \leq n-1} P(A_iA_jA_kA_n) + \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n) \\ &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_iA_j) + \sum_{1 \leq i < j < k \leq n} P(A_iA_jA_k) + \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n) \end{aligned}$$

因此式 (1.2.5) 对 n 也成立, 归纳法完成. \square

推论 1.2.2 (半可加性)

对于任意两个事件 A 与 B , 有

$$P(A \cup B) \leq P(A) + P(B) \quad (1.2.6)$$

对任意 n 个事件 A_1, A_2, \cdots, A_n , 有

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i) \quad (1.2.7)$$

证明: 根据性质 1.2.8, 对于任意两个事件 A 与 B , 有

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

整理得

$$P(AB) = P(A) + P(B) - P(A \cup B) \geq 0$$

进而可得

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

式 (1.2.6) 得证.

由式 (1.2.6) 可得

$$\begin{aligned}
 P\left(\bigcup_{i=1}^n A_i\right) &\leq P(A_1) + P\left(\bigcup_{i=2}^n A_i\right) \\
 &\leq P(A_1) + P(A_2) + P\left(\bigcup_{i=3}^n A_i\right) \\
 &\leq \cdots \\
 &\leq P(A_1) + P(A_2) + \cdots + P(A_n) \\
 &= \sum_{i=1}^n P(A_i)
 \end{aligned}$$

式 (1.2.7) 得证. □

1.2.3 概率的连续性

定义 1.2.3

对 \mathcal{F} 中任一单调不减的事件序列 $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq \cdots$, 称 $\bigcup_{i=1}^{\infty} F_i$ 为 $\{F_n\}$ 的极限事件, 记为

$$\lim_{n \rightarrow \infty} F_n = \bigcup_{i=1}^{\infty} F_i$$

对 \mathcal{F} 中任一单调不增的事件序列 $E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq \cdots$, 称 $\bigcap_{i=1}^{\infty} E_i$ 为 $\{E_n\}$ 的极限事件, 记为

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

定义 1.2.4

对 \mathcal{F} 上的一个概率 P , 若它对 \mathcal{F} 中任一单调不减的事件序列 $\{F_n\}$ 均有

$$\lim_{n \rightarrow \infty} P(F_n) = P\left(\lim_{n \rightarrow \infty} F_n\right)$$

则称概率 P 是下连续的. 若它对 \mathcal{F} 中任一单调不增的事件序列 $\{E_n\}$ 均有

$$\lim_{n \rightarrow \infty} P(E_n) = P\left(\lim_{n \rightarrow \infty} E_n\right)$$

则称概率 P 是上连续的.

性质 1.2.9 (概率的连续性)

若 P 为事件域 \mathcal{F} 上的概率, 则 P 既是下连续的, 又是上连续的.

证明：先证 P 的下连续性. 设 $\{F_n\}$ 是 \mathcal{F} 中一个单调不减的事件序列, 则

$$\lim_{n \rightarrow \infty} F_n = \bigcup_{i=1}^{\infty} F_i$$

若定义 $F_0 = \emptyset$, 则

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} (F_i - F_{i-1})$$

由于 $F_{i-1} \subseteq F_i$, 所以各个 $F_i - F_{i-1}$ 两两互斥, 再由可列可加性得

$$\begin{aligned} P(\lim_{n \rightarrow \infty} F_n) &= P\left(\bigcup_{i=1}^{\infty} F_i\right) \\ &= P\left(\bigcup_{i=1}^{\infty} (F_i - F_{i-1})\right) \\ &= \sum_{i=1}^{\infty} P(F_i - F_{i-1}) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n P(F_i - F_{i-1}) \end{aligned}$$

由有限可加性得

$$\sum_{i=1}^n P(F_i - F_{i-1}) = P\left(\bigcup_{i=1}^n (F_i - F_{i-1})\right) = P(F_n)$$

所以

$$P(\lim_{n \rightarrow \infty} F_n) = \lim_{n \rightarrow \infty} P(F_n)$$

下连续性得证.

再证 P 的上连续性. 设 $\{E_n\}$ 是 \mathcal{F} 中一个单调不增的事件序列, 则 $\{\overline{E_n}\}$ 为单调不减的事件序列, 由概率的下连续性得

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\overline{E_n}) &= P(\lim_{n \rightarrow \infty} \overline{E_n}) \\ &= P\left(\bigcup_{i=1}^{\infty} \overline{E_i}\right) \\ &= P\left(\overline{\bigcap_{i=1}^{\infty} E_i}\right) \\ &= 1 - P\left(\bigcap_{i=1}^{\infty} E_i\right) \\ &= 1 - P(\lim_{n \rightarrow \infty} E_n) \end{aligned}$$

另一方面

$$\lim_{n \rightarrow \infty} P(\overline{E_n}) = \lim_{n \rightarrow \infty} [1 - P(E_n)] = 1 - \lim_{n \rightarrow \infty} P(E_n)$$

所以

$$\lim_{n \rightarrow \infty} P(E_n) = P(\lim_{n \rightarrow \infty} E_n)$$

上连续性得证. □

定理 1.2.1

若 P 是 \mathcal{F} 上满足 $P(\Omega) = 1$ 的非负集合函数, 则 P 具有可列可加性的充分必要条件是: (1) P 是有限可加的; (2) P 是下连续的.

证明: 从性质 1.2.3 和 1.2.9 的证明过程可知, 由可列可加性可以推出有限可加性和下连续性, 因此必要性成立. 下面证明充分性.

设 $A_i \in \mathcal{F}, i = 1, 2, \dots$ 是两两互斥的事件序列, 由有限可加性可知, 对任意有限的 n 都有

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

因此正项级数 $\sum_{i=1}^{\infty} P(A_i)$ 收敛, 即

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

记 $F_n = \bigcup_{i=1}^n A_i$, 则 $\{F_n\}$ 为单调不减的事件序列, 由下连续性得

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} P(F_n) = P\left(\bigcup_{i=1}^{\infty} F_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

因此

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

可列可加性成立, 充分性得证. □

备注

从定理 1.2.1 可知, 在概率的公理化定义中, 可以将可列可加性换成有限可加性和下连续性.

1.2.4 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 $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, 基本事件 $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_n\}$ 两两互斥, 且

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

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

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

例题 1.2.1 (不放回抽样)

一批同种物品共有 N 件, 其中 M 件为甲类, 其余 $N - M$ 件为乙类. 从中随机取出 n 件 ($n \leq N$), 求事件 $A_m =$ “取出的 n 件物品中有 m 件为甲类” 的概率 ($m \leq M, n - m \leq N - M$).

解: 从 N 件物品中任取 n 件, 并且不讲次序, 所以样本空间 Ω 中样本点的总数为 C_N^n . 又因为是随机抽取的, 所以这些样本点是等可能的.

要使事件 A_m 发生, 必须从 M 件甲类物品中抽取 m 件, 再从 $N - M$ 件乙类物品中抽取 $n - m$ 件, 根据乘法原理, 事件 A_m 含有 $C_M^m C_{N-M}^{n-m}$ 个样本点, 由此得 A_m 的概率为

$$P(A_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

例题 1.2.2 (放回抽样)

一批同种物品共有 N 件, 其中 M 件为甲类, 其余 $N - M$ 件为乙类. 每次抽取一件后放回, 然后再抽取下一件, 如此重复直至抽出 n 件 ($n \leq N$) 为止. 求事件 $B_m =$ “取出的 n 件物品中有 m 件为甲类” 的概率 ($m \leq M, n - m \leq N - M$).

解: 由于每次抽取后都会放回, 因此每次抽取都是从 N 件物品中任取一件, 一共抽取 n 次, 所以样本空间 Ω 中样本点的总数为 N^n .

要使事件 B_m 发生, 必须从 M 件甲类物品中有放回地抽取 m 件, 再从 $N - M$ 件乙类物品中有放回地抽取 $n - m$ 件, 这样就有 $M^m (N - M)^{n-m}$ 种取法. 再考虑到这 m 件甲类物品可能在 n 次抽取中的任何 m 次中得到, 共有 C_n^m 种可能, 所以事件 B_m 含有 $C_n^m M^m (N - M)^{n-m}$ 个样本点, 由此得 B_m 的概率为

$$P(B_m) = \frac{C_n^m M^m (N - M)^{n-m}}{N^n} = C_n^m \left(\frac{M}{N}\right)^m \left(1 - \frac{M}{N}\right)^{n-m}$$

例题 1.2.3 (盒子模型)

设有 n 个球, 每个球都等可能地被放到 N ($n \leq N$) 个不同盒子中的任一个, 每个盒子所放球数不限. 求

- (1) 指定的 n 个盒子中各有一球的概率 p_1 ;
- (2) 恰好有 n 个盒子各有一球的概率 p_2 .

解: 每个球都可放到 N 个盒子中的任一个, n 个球的存放方式共有 N^n 种.

(1) 因为各有一球的 n 个盒子已经指定, 余下的没有球的 $N - n$ 个盒子也同时被指定, 所以只需要考虑 n 个球在这指定的 n 个盒子中各放 1 个的放法数. 第 1 个球有 n 种放法, 第 2 个球有 $n - 1$ 种放法, \dots , 第 n 个球有 1 种放法, 根据乘法原理, 总的放法数为 $n!$, 于是其概率为

$$p_1 = \frac{n!}{N^n}$$

(2) 第 1 个球可放在 N 个盒子中的任一个, 第 2 个球可放在余下的 $N - 1$ 个盒子中的任一个, \dots , 第 n 个球只可放在余下的 $N - n + 1$ 个盒子中的任一个, 根据乘法原理, 总的放法数为 $N(N - 1)(N - 2) \cdots (N - n + 1)$, 因此所求概率为

$$p_2 = \frac{N(N - 1)(N - 2) \cdots (N - n + 1)}{N^n} = \frac{N!}{N^n(N - n)!}$$

备注

盒子模型可以应用到很多实际问题中. 譬如将球解释为“粒子”, 把盒子解释为空间中的小区域, 这个问题便是统计物理学中的麦克斯韦-玻尔兹曼统计; 若 n 个“粒子”是不可辨的, 便是波色-爱因斯坦统计; 若 n 个“粒子”是不可辨的, 且每个“盒子”里最多只能放一个“粒子”, 这时就是费米-狄拉克统计.

例题 1.2.4 (生日问题)

求 n 个人的生日各不相同的概率 p_n .

解: 将 n 个人看成 n 个球, 将一年 365 天看成 $N = 365$ 个盒子, 则“ n 个人的生日各不相同”等价于“恰好有 n 个盒子各有一球”, 由例题 1.2.3 可知, 所求概率为

$$p_n = \frac{365(365 - 1)(365 - 2) \cdots (365 - n + 1)}{365^n} = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n - 1}{365}\right)$$

备注

直觉上, 30 个人的生日各不相同的概率应该还是比较大的, 至少大于 0.5. 然而, 根据例题 1.2.4 的结果, $p_{30} \approx 0.2937$, 这表明 30 个人的生日各不相同的概率比我们想象中更小. 当 $n = 23$ 时, $1 - p_{23} \approx 0.5073$, 也就是说, 23 个人中至少有两个人生日相同的概率就已经超过了 50%. 而当 $n = 60$ 时, $1 - p_{60} \approx 0.9941$, 说明 60 个人中至少有两个人生日相同的

概率超过 99%.

生日问题告诉我们, 我们的直觉并不一定准确, 有些问题的结果与我们的直觉不符. 因此, 我们不能过度依赖直觉, 应该学会用科学理性的思维去思考.

1.2.5 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, 而这个点落在 Ω 中任意两个度量相等的子区域内的概率相同, 则称这样的试验属于**几何概型** (geometric probability model).

备注

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.

对于任何有度量的区域 $A \subseteq \Omega$, 用 $m(A)$ 表示其度量, 则事件 $A =$ “随机点落在区域 A 内” 的概率定义为

$$P(A) = \frac{m(A)}{m(\Omega)}$$

例题 1.2.5 (会面问题)

甲乙两人约定在下午 6 时到 7 时之间在某处会面, 并约定先到者应等候另一个人 20 分钟, 过时即可离去. 求两人能会面的概率.

解: 设 x 和 y 分别表示甲、乙两人到达约会地点的时间, 则样本空间为

$$\Omega = \{(x, y) \mid 0 \leq x \leq 60, 0 \leq y \leq 60\}$$

设事件 $A =$ “甲乙两人能够会面”, 则

$$A = \{(x, y) \mid |x - y| \leq 20\}$$

由图 1.1 可知, Ω 的面积为 60^2 , A 的面积为 $60^2 - 40^2$, 因此

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

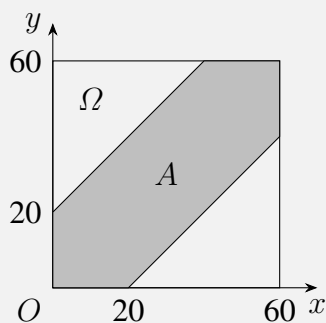


图 1.1

例题 1.2.6 (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 为针与此直线的夹角, 如图 1.2 所示, 则有

$$0 \leq x \leq a, 0 \leq \varphi \leq \pi$$

因此样本空间为

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

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

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

如图 1.3 所示. 所求概率为

$$p = P(A) = \frac{\int_0^\pi l \sin \varphi \, d\varphi}{\pi a} = \frac{2l}{\pi a} \quad (1.2.8)$$

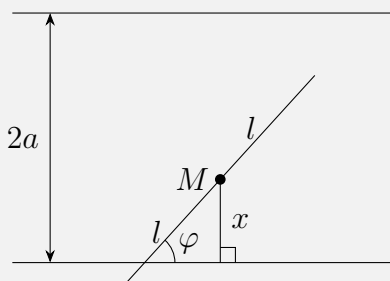


图 1.2

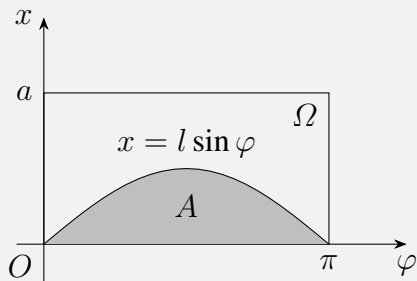


图 1.3

备注

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 $\frac{n}{N}$ 作为概率 p 的近似值, 代入式 (1.2.8) 可得

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

设计一个随机试验, 使一个事件的概率与某个未知数有关, 然后通过重复试验, 用频率估计概率, 即可求出未知数的近似值. 这种方法称为随机模拟法, 又称为蒙特卡罗法.

例题 1.2.7

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

解: 设分成的三段长度分别为 x, y 和 $a - x - y$, 则有

$$0 < x < a, 0 < y < a, 0 < a - x - y < a$$

其中 $0 < a - x - y < a$ 等价于 $0 < x + y < a$, 则样本空间为

$$\Omega = \{(x, y) \mid 0 < x < a, 0 < y < a, 0 < x + y < a\}$$

设事件 $A =$ “线段分成的三段可以构成三角形”. 由于三角形中任意两边之和大于第三边, 则事件 A 发生的条件为

$$\begin{cases} 0 < x < y + (a - x - y) \\ 0 < y < x + (a - x - y) \\ 0 < a - x - y < x + y \end{cases}$$

整理得

$$A = \{(x, y) \mid 0 < x < \frac{a}{2}, 0 < y < \frac{a}{2}, \frac{a}{2} < x + y < a\}$$

如图 1.4 所示. 则所求概率为

$$P(A) = \frac{\frac{1}{2} \left(\frac{a}{2}\right)^2}{\frac{1}{2}a^2} = \frac{1}{4}$$

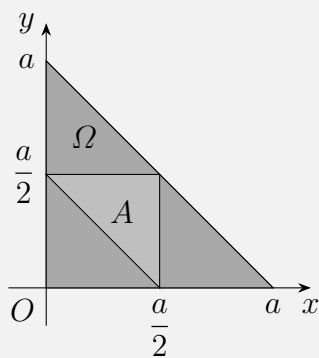


图 1.4

例题 1.2.8 (贝特朗奇论)

在一圆内任取一条弦, 求该弦的长度超过该圆内接等边三角形的边长的概率.

解:

解法一: 由于对称性, 可只考察某指定方向的弦, 作一条直径垂直于这个方向, 只有交直径于 $\frac{1}{4}$ 与 $\frac{3}{4}$ 之间的弦才能超过内接等边三角形的边长, 如图 1.5(a) 所示. 因此, 所求概率为 $\frac{1}{2}$.

解法二: 由于对称性, 可让弦的一个端点固定, 让另一端点在圆周上运动. 若在固定端点作切线, 则与此切线夹角在 60° 与 120° 之间的弦才能超过内接等边三角形的边长, 如

图 1.5(b) 所示. 因此, 所求概率为 $\frac{1}{3}$.

解法三: 圆内弦的位置由该弦的中点唯一确定. 在圆内作一同心圆, 其半径为大圆半径的一半, 这个小圆即为大圆内接等边三角形的内接圆. 只有大圆上的弦的中点落在小圆内, 此弦长才能超过内接等边三角形的边长, 如图 1.5(c) 所示. 因此, 所求概率为 $\frac{1}{4}$.

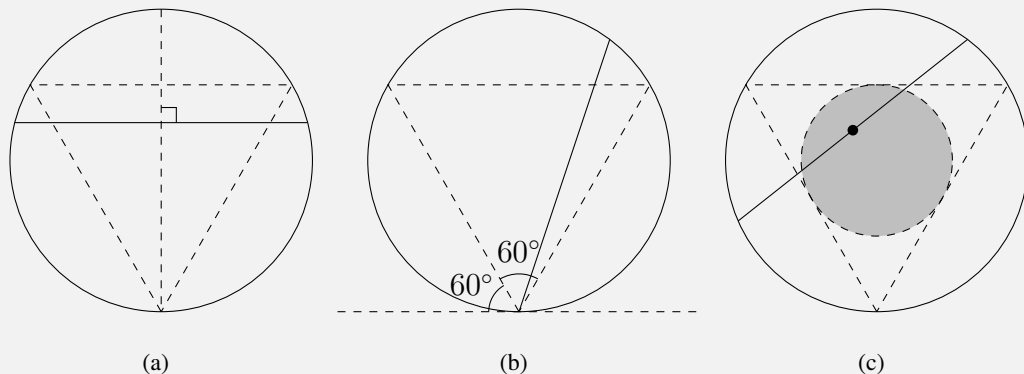


图 1.5 贝特朗奇论

备注

同一问题有三种不同答案, 这是因为在圆内取弦时规定不够具体, 不同的取法导致了不同的样本空间.

- 解法一中假定弦的中点在直径上均匀分布, 直径上的点组成样本空间 Ω_1 .
- 解法二中假定弦的另一活动端点在圆周上均匀分布, 圆周上的点组成样本空间 Ω_2 .
- 解法三中假定弦的中点在大圆内均匀分布, 大圆内的点组成样本空间 Ω_3 .

可见, 上述三个答案针对的是三个不同的样本空间, 它们都是正确的. 贝特朗奇论提醒我们, 在定义概率时需要明确指出其样本空间.

1.3 条件概率

1.3.1 条件概率及其性质

定义 1.3.1

设 A 和 B 是样本空间 Ω 中的两个事件, 若 $P(B) > 0$, 则称

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

为“在事件 B 发生的条件下事件 A 发生的概率”, 简称条件概率 (conditional probability).

备注

条件概率的本质是样本空间的改变. 计算 $P(A)$ 时, 样本空间为 Ω ; 而在事件 B 发生的条件下, 所有可能出现的样本点一定在 B 中, 因此条件概率 $P(A | B)$ 的样本空间为 B .

在定义 1.3.1 中, B 为样本空间, AB 为事件 A 落在样本空间 B 中的部分, 因此 $\frac{P(AB)}{P(B)}$ 即为条件概率 $P(A | B)$.

性质 1.3.1

若 $P(B) > 0$, 则有:

1. 非负性: $P(A | B) \geq 0, A \in \mathcal{F}$.
2. 规范性: $P(\Omega | B) = 1$.
3. 可列可加性: 若 $A_1, A_2, \dots, A_n, \dots$ 两两互斥, 则

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

证明: (1) $P(A | B) = \frac{P(AB)}{P(B)}$, 而 $P(AB) \geq 0, P(B) > 0$, 因此 $P(A | B) \geq 0$.

$$(2) P(\Omega | B) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(3) 若 $A_1, A_2, \dots, A_n, \dots$ 两两互斥, 则 $A_1 B, A_2 B, \dots, A_n B, \dots$ 也两两互斥, 因此

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i | B\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) B\right)}{P(B)} \\ &= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i B)\right)}{P(B)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i B)}{P(B)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i B)}{P(B)} \\ &= \sum_{i=1}^{\infty} P(A_i | B) \end{aligned}$$

□

备注

条件概率满足概率的公理化定义, 因此概率的任何性质对条件概率都成立.

1.3.2 乘法公式

定理 1.3.1 (乘法公式)

若 $P(B) > 0$, 则

$$P(AB) = P(B)P(A|B) \quad (1.3.1)$$

若 $P(A_1A_2\cdots A_{n-1}) > 0$ 则

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \quad (1.3.2)$$

证明: 根据定义 1.3.1, 若 $P(B) > 0$, 则有 $P(A|B) = \frac{P(AB)}{P(B)}$, 整理即可得到式 (1.3.1).

下面证明式 (1.3.2). 因为

$$P(A_1) \geq P(A_1A_2) \geq \cdots \geq P(A_1A_2\cdots A_{n-1}) > 0$$

所以式 (1.3.2) 中的条件概率均有意义, 按照条件概率的定义展开, 得

$$\begin{aligned} & P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1A_2)}{P(A_1)} \cdot \frac{P(A_1A_2A_3)}{P(A_1A_2)} \cdots \frac{P(A_1A_2\cdots A_n)}{P(A_1A_2\cdots A_{n-1})} \\ &= P(A_1A_2\cdots A_n) \end{aligned}$$

□

例题 1.3.1 (波利亚罐子模型)

设罐中有 b 个黑球、 r 个红球, 每次随机取出一个球, 取出后将原球放回, 再放入 c 个同色球和 d 个异色球. 记 B_i 为“第 i 次取出的是黑球”, R_j 为“第 j 次取出的是红球”. 若连续从罐中取出三个球, 求其中有两个红球、一个黑球的概率.

解: 设事件 A 为“连续从罐中取出三个球, 其中有两个红球、一个黑球”, 所求概率 $P(A)$ 与黑球在第几次被取出有关, 下面进行分类讨论.

(1) 黑球在第 1 次被取出: 第 1 次取球时有 b 个黑球、 r 个红球, 取出的是黑球; 第 2 次取球时有 $b+c$ 个黑球、 $r+d$ 个红球, 取出的是红球; 第 3 次取球时有 $b+c+d$ 个黑球、 $r+d+c$ 个红球, 取出的是红球. 因此

$$\begin{aligned} P(A) &= P(B_1R_2R_3) \\ &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\ &= \frac{b}{b+r} \cdot \frac{r+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d} \end{aligned}$$

(2) 黑球在第 2 次被取出: 第 1 次取球时有 b 个黑球、 r 个红球, 取出的是红球; 第 2 次取球时有 $b+d$ 个黑球、 $r+c$ 个红球, 取出的是黑球; 第 3 次取球时有 $b+d+c$ 个黑

球、 $r + c + d$ 个红球，取出的是红球. 因此

$$\begin{aligned} P(A) &= P(R_1 B_2 R_3) \\ &= P(R_1) P(B_2 | R_1) P(R_3 | R_1 B_2) \\ &= \frac{r}{b+r} \cdot \frac{b+d}{b+r+c+d} \cdot \frac{r+c+d}{b+r+2c+2d} \end{aligned}$$

(3) 黑球在第 3 次被取出: 第 1 次取球时有 b 个黑球、 r 个红球，取出的是红球; 第 2 次取球时有 $b+d$ 个黑球、 $r+c$ 个红球，取出的是红球; 第 3 次取球时有 $b+2d$ 个黑球、 $r+2c$ 个红球，取出的是黑球. 因此

$$\begin{aligned} P(A) &= P(R_1 R_2 B_3) \\ &= P(R_1) P(R_2 | R_1) P(B_3 | R_1 R_2) \\ &= \frac{r}{b+r} \cdot \frac{r+c}{b+r+c+d} \cdot \frac{b+2d}{b+r+2c+2d} \end{aligned}$$

备注

根据参数 c, d 的不同, 波利亚罐子模型可以有各种变化.

当 $c = -1, d = 0$ 时, 即为**不放回抽样**. 此时前面的抽取结果会影响后面的抽取, 但只要抽取的黑球与红球个数确定, 则抽出球的顺序不会影响其概率.

$$\begin{aligned} P(B_1 R_2 R_3) &= \frac{b}{b+r} \cdot \frac{r}{b+r-1} \cdot \frac{r-1}{b+r-2} = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)} \\ P(R_1 B_2 R_3) &= \frac{r}{b+r} \cdot \frac{b}{b+r-1} \cdot \frac{r-1}{b+r-2} = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)} \\ P(R_1 R_2 B_3) &= \frac{r}{b+r} \cdot \frac{r-1}{b+r-1} \cdot \frac{b}{b+r-2} = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)} \end{aligned}$$

当 $c = 0, d = 0$ 时, 即为**放回抽样**. 此时前面的抽取结果不会影响后面的抽取.

$$P(B_1 R_2 R_3) = P(R_1 B_2 R_3) = P(R_1 R_2 B_3) = \frac{br^2}{(b+r)^3}$$

当 $c > 0, d = 0$ 时, 称为**传染病模型**. 此时, 每次取出球后会增加下一次取到同色球的概率, 象征每次发现一个传染病患者都会增加再传染的概率.

在波利亚罐子模型中, 只要 $d = 0$, 上述三个概率都相等. 即只要抽取的黑球与红球个数确定, 则抽出球的顺序不会影响其概率.

$$P(B_1 R_2 R_3) = P(R_1 B_2 R_3) = P(R_1 R_2 B_3) = \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}$$

当 $c = 0, d > 0$ 时, 称为**安全模型**. 此模型可解释为: 每当事故发生了 (红球被取出), 安全工作就抓紧一些, 下次再发生事故的的概率就会减少; 而当事故没有发生时 (黑球被取出), 安全工作就放松一些, 下次发生事故的的概率就会增大. 此时, 上述三个概率不再相等,

它们分别为

$$\begin{aligned} P(B_1 R_2 R_3) &= \frac{b}{b+r} \cdot \frac{r+d}{b+r+d} \cdot \frac{r+d}{b+r+2d} \\ P(R_1 B_2 R_3) &= \frac{r}{b+r} \cdot \frac{b+d}{b+r+d} \cdot \frac{r+d}{b+r+2d} \\ P(R_1 R_2 B_3) &= \frac{r}{b+r} \cdot \frac{r}{b+r+d} \cdot \frac{b+2d}{b+r+2d} \end{aligned}$$

1.3.3 全概率公式

定理 1.3.2 (全概率公式)

若事件 A_1, A_2, \dots, A_n 两两互斥, 且 $P(A_i) > 0 (i = 1, 2, \dots, n)$, 则对任意事件 B , 若 $B \subseteq \bigcup_{i=1}^n A_i$, 有

$$P(B) = \sum_{i=1}^n P(A_i) P(B | A_i) \quad (1.3.3)$$

证明: 因为 $B \subseteq \bigcup_{i=1}^n A_i$, 所以

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

由于 A_1, A_2, \dots, A_n 互斥, 所以 $A_1 B, A_2 B, \dots, A_n B$ 也互斥, 由概率的有限可加性得

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

由乘法公式得

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

□

备注

1. 全概率公式的最简单形式: 若 $0 < P(A) < 1$, 则

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

2. 将条件改成“ $A_1, A_2, \dots, A_n, \dots$ 两两互斥, $P(A_i) > 0 (i = 1, 2, \dots)$, 且 $B \subseteq \bigcup_{i=1}^{\infty} A_i$ ”, 全概率公式仍然成立, 只需要将式 (1.3.3) 右侧改成可列项之和即可.

3. $P(A_i)$ 可以看做事件 A_i 对事件 B 的重要程度 (权重), $P(A_i) P(B | A_i)$ 为事件 A_i

对事件 B 的贡献. 全概率公式就是将 $A_i (i = 1, 2, \dots, n)$ 对 B 的贡献加在一起, 类似于加权平均.

例题 1.3.2 (摸彩模型)

设在 n 张彩票中有一张可中奖, 求第二个人摸到中奖彩票的概率.

解: 设 A_i 表示事件“第 i 个人摸到中奖彩票”, $i = 1, 2, \dots, n$. 若第一个人摸到中奖彩票, 则第二个人不可能再中奖, 因此

$$P(A_1) = \frac{1}{n}$$

$$P(A_2 | A_1) = 0$$

若第一个人没有摸到中奖彩票, 则第二个人要从剩下的 $n - 1$ 张彩票中摸奖, 此时有

$$P(\overline{A_1}) = \frac{n-1}{n}$$

$$P(A_2 | \overline{A_1}) = \frac{1}{n-1}$$

由全概率公式得

$$\begin{aligned} P(A_2) &= P(A_1) P(A_2 | A_1) + P(\overline{A_1}) P(A_2 | \overline{A_1}) \\ &= \frac{1}{n} \times 0 + \frac{n-1}{n} \cdot \frac{1}{n-1} \\ &= \frac{1}{n} \end{aligned}$$

备注

例题 1.3.2 表明, 摸到中奖彩票的机会与先后次序无关. 用类似的方法可得

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

如果 n 张彩票中有 k 张可中奖, $k \leq n$, 则

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

1.3.4 贝叶斯公式

如果事件 B 是由于在两两互斥的事件 A_1, A_2, \dots, 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 $P(B | A_i)$, 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, \dots, 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 $P(A_i | B)$ 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 1.3.3 (贝叶斯公式)

对于任意事件 A, B , 如果 $P(A) > 0, P(B) > 0$, 则

$$P(A | B) = \frac{P(A) P(B | A)}{P(B)} \quad (1.3.4)$$

若事件 A_1, A_2, \dots, A_n 两两互斥, 且 $P(A_i) > 0 (i = 1, 2, \dots, n)$, 则对任意事件 B , 如果 $P(B) > 0$, 且 $B \subseteq \bigcup_{i=1}^n A_i$, 有

$$P(A_i | B) = \frac{P(A_i) P(B | A_i)}{\sum_{j=1}^n P(A_j) P(B | A_j)} \quad (1.3.5)$$

证明: 如果 $P(A) > 0, P(B) > 0$, 由乘法公式可得

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

由此得

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

对于任一事件 A_i , 有

$$P(A_i | B) = \frac{P(A_i) P(B | A_i)}{P(B)}$$

利用全概率公式, 得

$$P(A_i | B) = \frac{P(A_i) P(B | A_i)}{\sum_{j=1}^n P(A_j) P(B | A_j)}$$

□

备注

$P(A_i) P(B | A_i)$ 表示事件 A_i 对事件 B 的贡献, 贝叶斯公式计算事件 A_i 的贡献在所有贡献中的占比, 以此估计事件 A_i 的发生概率.

1.4 事件的独立性

事件的独立性是指一个事件的发生不影响其他事件的发生, 也不受其他事件的影响. 对于事件 A 与 B , 当 $P(B) > 0$ 时, 条件概率 $P(A | B)$ 存在. 如果 $P(A | B) \neq P(A)$, 就意味着事件 B 的发生改变了事件 A 发生的概率, 也即事件 B 对事件 A 有某种影响. 如果事件 A 与 B 的发生不会相互影响, 则有 $P(A | B) = P(A)$, 由乘法公式可得

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

进一步有

$$P(AB) = P(A)P(B) \quad (1.4.1)$$

当 $P(A) = 0$ 或 $P(B) = 0$ 时, 式 (1.4.1) 仍然成立, 因此用式 (1.4.1) 作为事件相互独立的定义.

定义 1.4.1

对于任意事件 A 与 B , 如果 $P(AB) = P(A)P(B)$, 则称事件 A 与 B 相互独立.

备注

当事件 A 与 B 互斥, 且 $P(A) \neq 0, P(B) \neq 0$ 时, 事件 A 与 B 一定不相互独立, 因为此时 $P(AB) = 0$, 而 $P(A)P(B) \neq 0$. 反之, 当事件 A 与 B 相互独立, 且 $P(A) \neq 0, P(B) \neq 0$ 时, 事件 A 与 B 一定不互斥, 因为此时 $P(AB) = P(A)P(B) \neq 0$. 总之, 当两个事件的概率都不为 0 时, 互斥一定不相互独立, 相互独立一定不互斥.

不可能事件 \emptyset 与任意事件 A 相互独立, 因为 $P(\emptyset A) = P(\emptyset)P(A) = 0$. 同理, 概率为 0 的事件与任意事件相互独立.

必然事件 Ω 与任意事件 A 相互独立, 因为 $P(\Omega A) = P(A) = P(\Omega)P(A)$. 同理, 概率为 1 的事件与任意事件相互独立.

结论

对于同一试验 E 的任意两个事件 A 和 B , 如果 $P(A) > 0$, 则 A 与 B 相互独立的充分必要条件是 $P(B|A) = P(B)$; 如果 $P(B) > 0$, 则 A 与 B 相互独立的充分必要条件是 $P(A|B) = P(A)$.

结论

如果事件 A 与 B 相互独立, 则 A 与 \bar{B} 相互独立, \bar{A} 与 B 相互独立, \bar{A} 与 \bar{B} 相互独立.

证明:

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

因此事件 A 与 \bar{B} 相互独立.

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

因此事件 \bar{A} 与 B 相互独立.

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

因此事件 \bar{A} 与 \bar{B} 相互独立. □

定义 1.4.2

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 两两相互独立.

定义 1.4.3

如果三个事件 A, B, C 两两相互独立, 并且有 $P(ABC) = P(A)P(B)P(C)$, 则称三个事件 A, B, C 相互独立.

备注

如果三个事件两两相互独立, 不能直接推出这三个事件相互独立. 例如: 一个均匀的四面体, 一面红色, 一面黄色, 一面蓝色, 一面三种颜色都有, 抛掷该四面体, 设事件 A = “朝下的面包含红色”, 事件 B = “朝下的面包含黄色”, 事件 C = “朝下的面包含蓝色”, 则

$$\begin{aligned}
 P(A) &= P(B) = P(C) = \frac{1}{2} \\
 P(AB) &= P(BC) = P(AC) = \frac{1}{4} \\
 P(ABC) &= \frac{1}{4} \neq P(A)P(B)P(C)
 \end{aligned}$$

可得事件 A, B, C 两两相互独立, 但这三个事件不相互独立.

当 $P(ABC) = P(A)P(B)P(C)$ 成立时, 不能推出 A, B, C 两两相互独立.

结论

若 A, B, C 相互独立, 则 $A \cup B$ 与 C 相互独立, AB 与 C 相互独立, $A - B$ 与 C 相互独立.

证明:

$$\begin{aligned}
 P((A \cup B)C) &= P(AC \cup BC) \\
 &= P(AC) + P(BC) - P(ABC) \\
 &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\
 &= [P(A) + P(B) - P(A)P(B)]P(C) \\
 &= P(A \cup B)P(C)
 \end{aligned}$$

所以 $A \cup B$ 与 C 相互独立.

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

所以 AB 与 C 相互独立.

$$\begin{aligned}
 P((A - B)C) &= P(\overline{A}BC) \\
 &= P(\overline{A}C - B) \\
 &= P(\overline{A}C) - P(ABC) \\
 &= P(A)P(C) - P(AB)P(C) \\
 &= [P(A) - P(AB)]P(C) \\
 &= P(A - B)P(C)
 \end{aligned}$$

所以 $A - B$ 与 C 相互独立. □

备注

如果事件 A, B, C 只有两两相互独立, 则不能证明 $A \cup B$ 与 C 相互独立、 AB 与 C 相互独立、 $A - B$ 与 C 相互独立.

定义 1.4.4

设 A_1, A_2, \dots, A_n 是同一试验 E 的 n 个事件, 如果对于这 n 个事件中的任意 k ($2 \leq k \leq n$) 个事件 $A_{i_1}, A_{i_2}, \dots, A_{i_k}$, 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, \dots, A_n 相互独立.

结论

若 n ($n \geq 2$) 个事件相互独立, 则其中任意 k ($2 \leq k \leq n$) 个事件相互独立. 将其中任意 k 个事件换成对立事件, 所得诸事件仍然相互独立.

1.5 伯努利概型

定义 1.5.1

设有两个试验 E_1 和 E_2 , 假如试验 E_1 的任一结果 (事件) 与试验 E_2 的任一结果 (事件) 都是相互独立的事件, 则称这两个试验相互独立.

定义 1.5.2

如果试验 E_1 的任一结果、试验 E_2 的任一结果、……、试验 E_n 的任一结果都是相互独立的事件, 则称试验 E_1, E_2, \dots, E_n 相互独立. 如果这 n 个独立试验是相同的, 则称其为 n 重独立重复试验.

定义 1.5.3

如果在 n 重独立重复试验中, 每次试验的可能结果只有两个 (A 或 \bar{A}), 则称这种试验为 n 重伯努利试验, 简称为伯努利试验 (Bernoulli experiment), 也称为伯努利概型 (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 $\omega_i (1 \leq i \leq n)$ 为 A 或者为 \bar{A} , 即 ω 是从 A 及 \bar{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.

设 $P(A) = p$, $P(\bar{A}) = 1 - p$, 其中 $0 < p < 1$. 如果 ω 中有 k 个 A , 则必有 $n - k$ 个 \bar{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 \bar{A} , 因此在 n 重伯努利试验中, 事件 A 恰好发生 k 次的概率 $P_n(k)$ 为

$$P_n(k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n \quad (1.5.1)$$

由二项式定理可得

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

由此可见, $C_n^k p^k (1 - p)^{n-k}$ 是二项展开式中的一项, 因此式 (1.5.1) 又称为二项概率公式.

第二章 随机变量及其分布

2.1 随机变量及其分布函数

2.1.1 随机变量

定义 2.1.1

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 2.1.2

设 X 是一个随机变量, 对于任意实数 x , 令

$$F(x) = P(X \leq x), x \in \mathbf{R}$$

称 $F(x)$ 为随机变量 X 的分布函数 (cumulative distribution function). 称随机变量 X 服从 $F(x)$, 记为 $X \sim F(x)$.

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 $\{X \leq x\}$ 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 2.1.1 (单调性)

$F(x)$ 是一个单调不减函数, 即对任意的 $x_1 < x_2$, 有 $F(x_1) \leq F(x_2)$.

证明: 对于任意实数 $x_1, x_2 (x_1 < x_2)$, 有

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

因此 $F(x)$ 是单调不减函数. □

性质 2.1.2 (有界性)

对于任意实数 x , 有 $0 \leq F(x) \leq 1$, 且

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

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

证明: 因为 $F(x) = P(X \leq x)$, 根据概率的性质可得 $0 \leq F(x) \leq 1$.

$F(x)$ 在 $(-\infty, +\infty)$ 内单调不减且有界, 由单调有界原理可知, $\lim_{x \rightarrow -\infty} F(x)$ 和 $\lim_{x \rightarrow +\infty} F(x)$ 存在.

记 $A_n = \{X \leq n\}$, 则有 $A_n \subseteq A_{n+1}$, $\bigcup_{i=1}^{\infty} A_i = \Omega$. 由海涅定理可得

$$\lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(A_n) = P(\lim_{n \rightarrow \infty} A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1$$

记 $B_n = \{X \leq -n\}$, 则有 $B_n \supseteq B_{n+1}$, $\bigcap_{i=1}^{\infty} B_i = \emptyset$. 由海涅定理可得

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow \infty} F(-n) = \lim_{n \rightarrow \infty} P(B_n) = P(\lim_{n \rightarrow \infty} B_n) = P\left(\bigcap_{i=1}^{\infty} B_i\right) = 0$$

□

性质 2.1.3 (右连续性)

$F(x)$ 处处右连续, 即对任意实数 x_0 , 有 $F(x_0^+) = F(x_0)$.

证明: 因为 $F(x)$ 是单调不减有界函数, 所以对任意实数 x_0 , 右极限 $F(x_0^+)$ 一定存在. 记 $A_n = \{X \leq x_0 + \frac{1}{n}\}$, 则有 $A_n \supseteq A_{n+1}$, $\bigcap_{i=1}^{\infty} A_i = \{X \leq x_0\}$, 由海涅定理得

$$\begin{aligned} \lim_{x \rightarrow x_0^+} F(x) &= \lim_{n \rightarrow \infty} F(x_0 + \frac{1}{n}) \\ &= \lim_{n \rightarrow \infty} P(A_n) \\ &= P(\lim_{n \rightarrow \infty} A_n) \\ &= P\left(\bigcap_{i=1}^{\infty} A_i\right) \\ &= P(X \leq x_0) \\ &= F(x_0) \end{aligned}$$

□

单调性、有界性、右连续性是分布函数的基本性质. 分布函数一定具有这些基本性质, 反过来, 满足这些基本性质的函数一定是某个随机变量的分布函数. 因此, 这三条基本性质是判定某个函数能否成为分布函数的充分必要条件.

性质 2.1.4

对于任意实数 $x_1, x_2 (x_1 < x_2)$, 有 $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$.

证明: 对于任意实数 $x_1, x_2 (x_1 < x_2)$, 由于

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

所以有

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

□

2.2 离散型随机变量及其概率分布

定义 2.2.1

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 2.2.2

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

$$P(X = x_k) = p_k, \quad k = 1, 2, \dots \quad (2.2.1)$$

并且 p_k 满足以下两个条件:

1. 非负性: $p_k \geq 0$;

2. 归一性: $\sum_{k=1}^{\infty} p_k = 1$,

则称式 (2.2.1) 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 $\{X \leq x\}$ 可以表示成 $\bigcup_{x_k \leq x} \{X = x_k\}$. 由于 $x_k (k = 1, 2, \dots)$ 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

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

它的图像是有限级 (或可列无限级) 的阶梯形曲线.

如果随机变量 X 只能取一个常数 c , 即 $P(X = c) = 1$, 这样的分布称为单点分布或退化分布, 它的分布函数是

$$F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}$$

2.3 连续型随机变量及其概率密度

定义 2.3.1

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 $F(x) = \int_{-\infty}^x f(t) dt$, 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 2.3.1 可知, 连续型随机变量的分布函数处处连续.

性质 2.3.1 (非负性)

$$f(x) \geq 0$$

性质 2.3.2 (规范性)

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

非负性和规范性是概率密度函数的基本性质. 概率密度函数一定具有这两个性质, 满足这两个性质的函数也一定是某随机变量的概率密度. 因此, 非负性和规范性是判断某个函数能否作为概率密度函数的充分必要条件.

性质 2.3.3

对于任意实数 $a, b (a < b)$, 有

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$$

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 $P(a < X \leq b)$ 等于区间 $(a, b]$ 上曲线 $y = f(x)$ 之下的曲边梯形的面积.

性质 2.3.4

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

由性质 2.3.4 可知, 在 $f(x)$ 的连续点有

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

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

$$P(x < X \leq x + \Delta x) \approx f(x)\Delta x$$

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 $\{X = x\} \subseteq \{x - \Delta x < X \leq x\}$, 因此有

$$0 \leq P(X = x) \leq P(x - \Delta x < X \leq x) = F(x) - F(x - \Delta x)$$

令 $\Delta x \rightarrow 0$, 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

$$P(x_1 < X < x_2) = P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = \int_{x_1}^{x_2} f(x) dx$$

由上述结论可知, 不可能事件的概率为 0, 但概率为 0 的事件未必是不可能事件. 类似地, 必然事件的概率为 1, 但概率为 1 的事件未必是必然事件.

备注

由于在若干点上改变概率密度 $f(x)$ 的值并不影响其积分的值, 从而不影响其分布函数 $F(x)$ 的值, 这意味着一个连续型概率分布的概率密度不唯一. 例如:

$$f_1(x) = \begin{cases} \frac{1}{a}, & 0 \leq x \leq a \\ 0, & \text{其他} \end{cases} \quad f_2(x) = \begin{cases} \frac{1}{a}, & 0 < x < a \\ 0, & \text{其他} \end{cases}$$

这两个概率密度只在 $x = 0$ 和 $x = a$ 处的取值不同, 因此

$$P(f_1(x) \neq f_2(x)) = P(X = 0) + P(X = a) = 0$$

可见这两个函数在概率意义上没有差别, 称 $f_1(x)$ 与 $f_2(x)$ “几乎处处相等”, 其意义是: 在概率论中可除去概率为 0 的事件后讨论两个函数相等和其他随机问题.

备注

除了离散型概率分布和连续型概率分布之外, 还有既非离散又非连续的分布. 例如:

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

$F(x)$ 是一个分布函数, 它既不是阶梯函数, 也不是连续函数, 所以它既不是离散型概率分布, 又不是连续型概率分布. 这是一种新的分布, 这类分布函数通常可以分解为两个分布函数的凸组合, 如上述 $F(x)$ 可分解为

$$F(x) = \frac{1}{2}F_1(x) + \frac{1}{2}F_2(x)$$

其中

$$F_1(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad F_2(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

而 $F_1(x)$ 是离散型概率分布 (单点分布) 的分布函数, $F_2(x)$ 是连续型概率分布 (均匀分布) 的分布函数.

2.4 常用的概率分布

2.4.1 0-1 分布

定义 2.4.1

如果离散型随机变量 X 只取 0 与 1 两个值, 其概率分布为

$$P(X=0) = 1-p, \quad P(X=1) = p, \quad 0 < p < 1$$

或写成

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

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 $0, 1, 2, \dots, n$. 设 $P(A) = p (0 < p < 1)$, 则由二项概率公式 (式 (1.5.1)) 可得

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

定义 2.4.2

如果随机变量 X 的概率分布为

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

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 2.4.2 可得

$$\begin{aligned} P(X = k) &\geq 0 \\ \sum_{k=0}^n P(X = k) &= \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = [p + (1-p)]^n = 1 \end{aligned}$$

特别地, 当 $n = 1$ 时, 二项分布 $B(1, p)$ 的概率分布为

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

这就是 0-1 分布. 因此, 0-1 分布是二项分布的特例.

在 n 重伯努利试验中, 设第 i 次试验中 A 出现的次数为 X_i , 则 X_i 服从 0-1 分布. 由于 n 次试验相互独立, 所以产生的 n 个随机变量 X_1, X_2, \dots, X_n 也相互独立, 将它们加起来, 得到新的随机变量

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

X 就是 n 重伯努利试验中 A 出现的总次数, 它服从二项分布 $B(n, p)$. 因此, 服从二项分布的随机变量总可以分解成 n 个相互独立且同为 0-1 分布的随机变量之和.

2.4.3 泊松分布

定义 2.4.3

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

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

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 $X \sim \pi(\lambda)$.

由定义可知

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

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

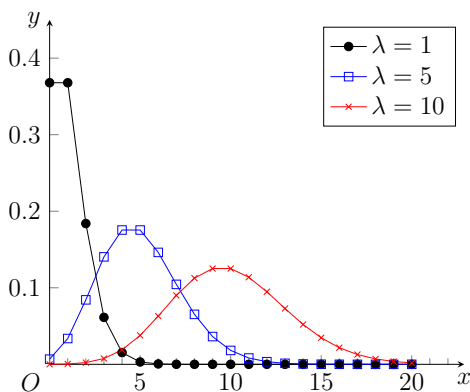


图 2.1 泊松分布

定理 2.4.1 (泊松定理)

设 $\{p_n\}$ 为一个实数序列, 且对于任意正整数 n , 有 $p_n \in [0, 1]$. 若 $n \rightarrow \infty$ 时有 $np_n \rightarrow \lambda$ ($\lambda > 0$), 则对任一固定的非负整数 k , 有

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

证明: 记 $np_n = \lambda_n$, 即 $p_n = \frac{\lambda_n}{n}$, 从而对于任意固定的非负整数 k , 有

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

对于固定的 k , 有

$$\begin{aligned} \lim_{n \rightarrow \infty} \lambda_n &= \lambda \\ \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) &= 1 \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^n &= e^{-\lambda} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda_n}{n}\right)^{-k} &= 1 \end{aligned}$$

所以

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

□

对于二项分布 $B(n, p)$, np 为定值, 当 n 很大而 p 很小时, $np = \lambda$ 大小适中, 可以将二项分布近似为泊松分布.

2.4.4 几何分布

设试验 E 只有两个对立的结果 A 和 \bar{A} , 并且 $P(A) = p$, $P(\bar{A}) = 1 - p$, 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 是一个离散型随机变量, 它的所有可能取值为 $1, 2, 3, \dots$, 其概率分布为

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

称随机变量 X 服从参数为 p 的几何分布 (geometric distribution), 记作 $X \sim Ge(p)$.

易知

$$(1 - p)^{k-1}p > 0$$

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

定理 2.4.2 (几何分布的无记忆性)

设 $X \sim Ge(p)$, 则对任意正整数 m 与 n , 有

$$P(X > m + n \mid X > m) = P(X > n) \quad (2.4.1)$$

证明: 因为

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

所以对任意正整数 m 与 n , 有

$$P(X > m + n \mid X > m) = \frac{P(X > m + n)}{P(X > m)} = \frac{(1 - p)^{m+n}}{(1 - p)^m} = (1 - p)^n = P(X > n) \quad \square$$

备注

若 $X \sim Ge(p)$, 则事件 $\{X > m\}$ 表示前 m 次试验中 A 没有出现. 假如在接下去的 n 次试验中 A 仍未出现, 这个事件记为 $\{X > m + n\}$. 定理 2.4.2 表明, 在前 m 次试验中 A 没有出现的条件下, 接下去的 n 次试验中 A 仍未出现的概率只与 n 有关, 而与之前的 m 次试验无关, 这就是无记忆性.

2.4.5 超几何分布

设有 N 件物品, 其中有 M ($M \leq N$) 件为甲类, 从中不放回地随机抽取 n ($n \leq N$) 件, 记抽出甲类物品的数量为 X , 则随机变量 X 服从超几何分布 (hypergeometric distribution), 记为 $X \sim H(n, N, M)$.

当 $n \leq N - M$ 时, X 的取值范围为 $0, 1, 2, \dots, \min\{M, n\}$; 当 $n > N - M$ 时, 一定能够抽出甲类物品, X 的取值范围为 $n + M - N, n + M - N + 1, n + M - N + 2, \dots, \min\{M, n\}$. 综合两种情况, 记 $m = \max\{0, n + M - N\}$, $r = \min\{M, n\}$, 则超几何分布的概率分布为

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k = m, m+1, m+2, \dots, r$$

备注

当 $N = M$ 时, 有

$$P(X = n) = \frac{C_M^n C_{N-M}^{n-n}}{C_N^n} = C_0^0 = 1$$

C_0^0 本身没有意义, 但为了保证组合数定义的一致性, 规定 $C_0^0 = 1$.

当 $n \ll N$ 时, 每次抽取后, 甲类物品所占比例 $p = \frac{M}{N}$ 改变很小, 此时可以将不放回抽样近似看做放回抽样, 超几何分布近似为二项分布:

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \approx C_n^k p^k (1-p)^{n-k}$$

2.4.6 帕斯卡分布

设试验 E 只有两个对立的结果 A 和 \bar{A} , 并且 $P(A) = p$, $P(\bar{A}) = 1 - p$, 其中 $0 < p < 1$. 将试验 E 独立地重复进行下去, 记事件 A 第 r 次出现时的试验次数为 X , 则 X 的可能取值为 $r, r+1, r+2, \dots$, 称随机变量 X 服从帕斯卡分布 (Pascal distribution).

帕斯卡分布是负二项分布 (negative binomial distribution) 的正整数形式, 因此可以将帕斯卡分布记作 $X \sim Nb(r, p)$.

如果 $X = k$, 意味着第 k 次试验时事件 A 第 r 次出现, 而前 $k-1$ 次试验中 A 出现 $r-1$ 次, 由二项分布知

$$P(X = k) = C_{k-1}^{r-1} p^{r-1} (1-p)^{k-r} \cdot p$$

因此帕斯卡分布的概率分布为

$$P(X = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \dots$$

当 $r = 1$ 时, 帕斯卡分布即为几何分布.

如果将第一个 A 出现的试验次数记为 X_1 , 第二个 A 出现的试验次数记为 X_2 (从第一个 A 出现之后算起), \dots , 第 r 个 A 出现的试验次数记为 X_r (从第 $r-1$ 个 A 出现之后算起), 则诸 X_i 独立同分布, 且 $X_i \sim Ge(p)$. 此时有 $X = X_1 + X_2 + \dots + X_r \sim Nb(r, p)$, 即服从帕斯卡分布的随机变量可以分解成 r 个独立同分布的几何分布随机变量之和.

2.4.7 均匀分布

定义 2.4.4

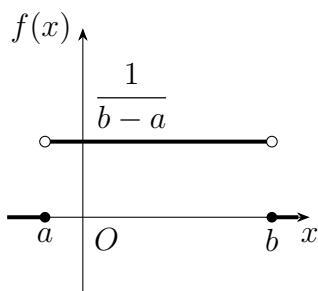
如果连续型随机变量 X 的概率密度为

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

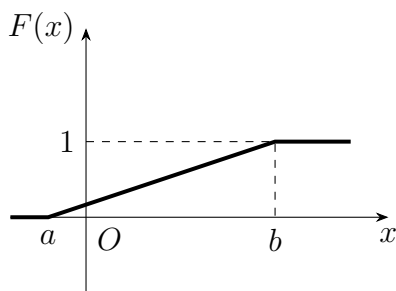
则称 X 在区间 (a, b) 上服从均匀分布 (uniform distribution), 记作 $X \sim U(a, b)$.

均匀分布 $U(a, b)$ 的分布函数为

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



(a) 均匀分布的概率密度



(b) 均匀分布的分布函数

图 2.2 均匀分布

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

$$P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} \frac{dx}{b-a} = \frac{x_2 - x_1}{b-a}$$

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.

2.4.8 指数分布

定义 2.4.5

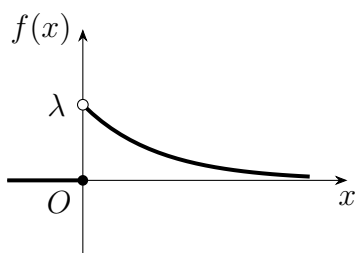
如果连续型随机变量 X 的概率密度为

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

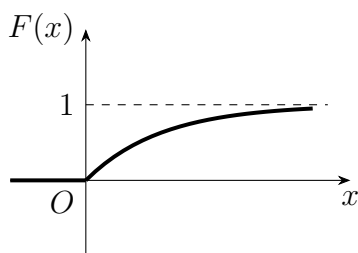
其中 $\lambda > 0$ 是常数, 则称 X 服从参数为 λ 的指数分布 (exponential distribution), 记作 $X \sim \text{Exp}(\lambda)$.

指数分布 $\text{Exp}(\lambda)$ 的分布函数为

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



(a) 指数分布的概率密度



(b) 指数分布的分布函数

图 2.3 指数分布

定理 2.4.3 (指数分布的无记忆性)

如果随机变量 $X \sim \text{Exp}(\lambda)$, 则对任意实数 $s > 0, t > 0$, 有

$$P(X > s + t \mid X > s) = P(X > t) \quad (2.4.2)$$

证明: 因为 $X \sim \text{Exp}(\lambda)$, 所以对任意实数 $s > 0$, 有

$$P(X > s) = 1 - P(X \leq s) = 1 - (1 - e^{-\lambda s}) = e^{-\lambda s}$$

又因为 $\{X > s + t\} \subseteq \{X > s\}$, 所以 $\{X > s + t\} \cap \{X > s\} = \{X > s + t\}$. 于是

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

□

备注

定理 2.4.3 的物理意义: 记 X 是某种产品的使用寿命, 若 X 服从指数分布, 那么已知此产品使用了 s 小时没发生故障, 则能再使用 t 小时而不发生故障的概率与已使用的 s 小时无关, 只相当于重新开始使用 t 小时的概率.

结论 (泊松分布与指数分布的关系)

如果在 $[0, t]$ 时间段内随机事件 A 发生的次数 $N(t)$ 服从参数为 λt 的泊松分布, 则 A 相邻两次发生的时间间隔 T 服从参数为 λ 的指数分布.

证明: 由于 $N(t) \sim P(\lambda t)$, 即

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

事件 $\{T \geq t\}$ 说明事件 A 在 $[0, t]$ 时间段内没有发生, 即 $\{T \geq t\} = \{N(t) = 0\}$, 由此可得:

- 当 $t < 0$ 时, 有 $F_T(t) = P(T \leq t) = 0$;
- 当 $t \geq 0$ 时, 有

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P(N(t) = 0) = 1 - e^{-\lambda t}$$

所以 $T \sim \text{Exp}(\lambda)$. □

备注

泊松分布常用来描述单位时间内随机事件发生的次数, 如一天之内光临店铺的顾客数、一段时间内电话来电的次数、一段时间内设备发生故障的次数等. 指数分布常用作描述时间长度, 如随机服务中的等待时间、设备的寿命、电话的通话时间等.

2.4.9 正态分布

2.4.9.1 正态分布及其性质

定义 2.4.6

如果连续型随机变量 X 的概率密度为

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

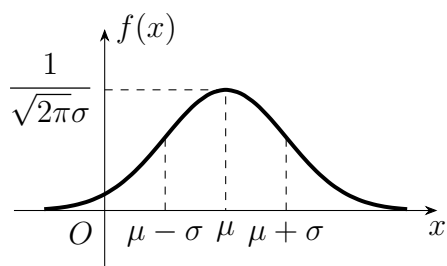
其中 $\mu, \sigma (\sigma > 0)$ 为常数, 则称 X 服从参数为 μ, σ 的正态分布 (normal distribution), 记作 $X \sim N(\mu, \sigma^2)$.

概率密度 $f(x)$ 的图像如图 2.4(a) 所示, 它是一条钟形曲线, 关于直线 $x = \mu$ 对称, 并在 $x = \mu$ 处取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$, 在 $x = \mu \pm \sigma$ 处有拐点, 以 x 轴为水平渐近线.

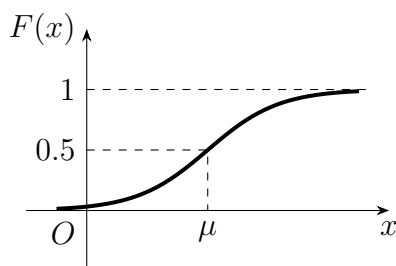
正态分布 $N(\mu, \sigma^2)$ 的分布函数为

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

它是一条光滑上升的 S 形曲线, 如图 2.4(b) 所示.



(a) 正态分布的概率密度



(b) 正态分布的分布函数

图 2.4 正态分布

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(a) 所示. 由此可见, 正态密度函数的位置由参数 μ 确定, 因此称 μ 为位置参数.

如果固定 μ , 改变 σ 的值, 则由 $y_{\max} = \frac{1}{\sqrt{2\pi}\sigma}$ 可知: σ 越小, 概率密度曲线在 $x = \mu$ 附近越陡峭, 形状高而瘦, 分布较为集中, X 落在 $x = \mu$ 附近的概率较大; σ 越大, 概率密度曲线越平坦, 形状矮而胖, 分布较为分散. 如图 2.5(b) 所示. 可见正态密度函数的尺度由参数 σ 确定, 因此称 σ 为尺度参数.

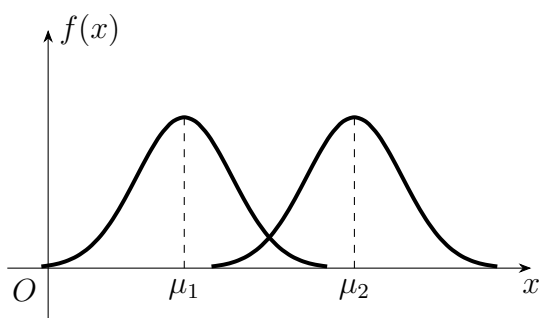
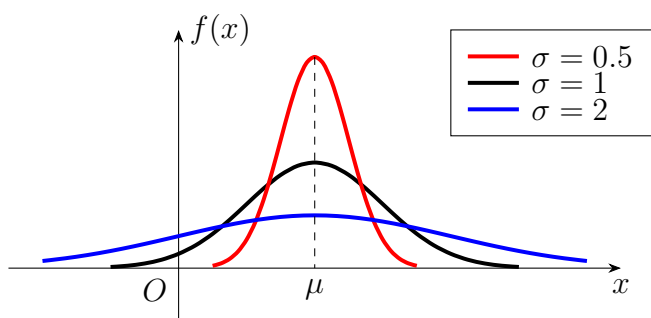
(a) 固定 σ , 改变 μ (b) 固定 μ , 改变 σ

图 2.5

2.4.9.2 标准正态分布

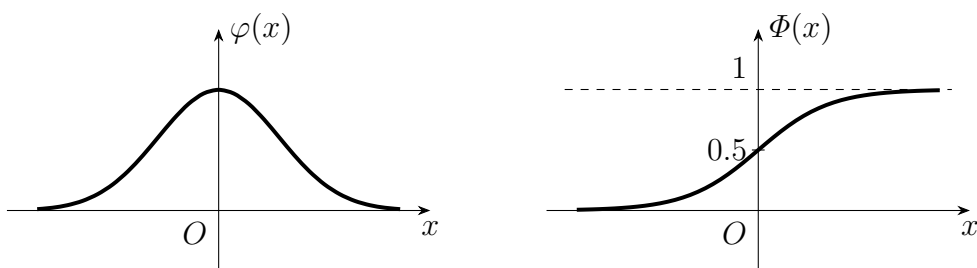
定义 2.4.7

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

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

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



(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 2.6 标准正态分布

性质 2.4.1

$$\Phi(-x) = 1 - \Phi(x) \quad (2.4.3)$$

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

$$\begin{aligned} \Phi(-x) &= \int_{-\infty}^{-x} \varphi(t) dt \\ &= \int_{-\infty}^{+\infty} \varphi(t) dt - \int_{-x}^{+\infty} \varphi(t) dt \\ &= 1 - \int_{-x}^{+\infty} \varphi(t) dt \end{aligned}$$

令 $u = -t$, 得

$$\begin{aligned} \Phi(-x) &= 1 - \int_x^{-\infty} \varphi(-u) d(-u) \\ &= 1 - \int_{-\infty}^x \varphi(u) du \\ &= 1 - \Phi(x) \end{aligned}$$

□

性质 2.4.2

若 $X \sim N(0, 1)$, 则对于任意正的实数 a , 有

$$P(|X| > a) = 2[1 - \Phi(a)] \quad (2.4.4)$$

$$P(|X| \leq a) = 2\Phi(a) - 1 \quad (2.4.5)$$

证明:

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

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

□

性质 2.4.3 (正态分布的标准化)

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

证明: 设 X 的分布函数为 $F_X(x)$. 令 $U = \frac{X - \mu}{\sigma}$, 则其分布函数为

$$F_U(u) = P(U \leq u) = P\left(\frac{X - \mu}{\sigma} \leq u\right) = P(X \leq \mu + \sigma u) = F_X(\mu + \sigma u)$$

设 X 与 U 的概率密度分别为 $f_X(x)$ 和 $f_U(u)$, 则有

$$f_U(u) = \frac{d}{du} F_U(u) = \frac{d}{du} F_X(\mu + \sigma u) = f_X(\mu + \sigma u) \cdot \sigma = \sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu + \sigma u - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

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

推论 2.4.1

如果 $X \sim N(\mu, \sigma^2)$, 则

$$P(X \leq x) = F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (2.4.6)$$

推论 2.4.2

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

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \Phi\left(\frac{x_2 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right) \quad (2.4.7)$$

结论 (3σ 规则)

若随机变量 $X \sim N(\mu, \sigma^2)$, 则

$$P(|X - \mu| < \sigma) = 2\Phi(1) - 1 = 0.6826$$

$$P(|X - \mu| < 2\sigma) = 2\Phi(2) - 1 = 0.9544$$

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

证明: 由式 (2.4.7) 和式 (2.4.3) 有

$$\begin{aligned} P(|X - \mu| < k\sigma) &= P(\mu - k\sigma < X < \mu + k\sigma) \\ &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi(k) - \Phi(-k) \\ &= \Phi(k) - [1 - \Phi(k)] \\ &= 2\Phi(k) - 1 \end{aligned}$$

所以

$$P(|X - \mu| < \sigma) = 2\Phi(1) - 1 = 0.6826$$

$$P(|X - \mu| < 2\sigma) = 2\Phi(2) - 1 = 0.9544$$

$$P(|X - \mu| < 3\sigma) = 2\Phi(3) - 1 = 0.9974$$
□

3σ 规则表明, 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.

如果某随机变量近似满足 3σ 规则, 则可认为该随机变量近似服从正态分布; 反之, 如果随机变量与 3σ 规则相差较大, 则可认为该随机变量不服从正态分布.

2.4.9.3 标准正态分布的上 α 分位点

定义 2.4.8

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

$$P(X \geq u_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx = \alpha$$

则称点 u_α 为标准正态分布的上 α 分位点.

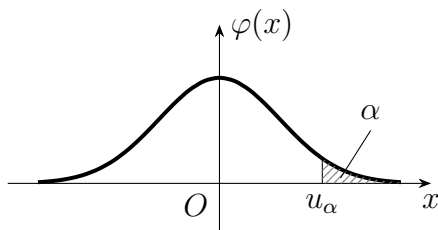


图 2.7 标准正态分布的上 α 分位点

性质 2.4.4

$$\Phi(u_\alpha) = 1 - \alpha \quad (2.4.8)$$

证明:

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

□

性质 2.4.5

标准正态分布的上 $1 - \alpha$ 分位点 $u_{1-\alpha}$ 与上 α 分位点 u_α 关于原点对称, 即

$$u_{1-\alpha} = -u_\alpha \quad (2.4.9)$$

证明:

$$\begin{aligned} P(X \geq -u_\alpha) &= 1 - P(X < -u_\alpha) \\ &= 1 - \Phi(-u_\alpha) \\ &= 1 - [1 - \Phi(u_\alpha)] \\ &= \Phi(u_\alpha) \\ &= 1 - \alpha \\ &= P(X \geq u_{1-\alpha}) \end{aligned}$$

因此 $u_{1-\alpha} = -u_\alpha$.

□

2.4.10 伽马分布

定义 2.4.9

函数

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \quad (2.4.10)$$

称为伽马函数, 也叫欧拉第二积分, 其中 $\alpha > 0$.

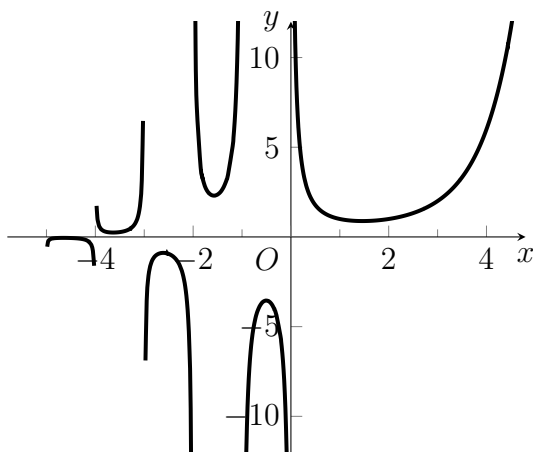


图 2.8 伽马函数

从图 2.8 可知, 伽马函数在 $x > 0$ 时是连续的, 0 及所有负整数点都是奇点.

性质 2.4.6 (伽马函数的性质)

1. $\Gamma(1) = 1$
2. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
3. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
4. 对于正整数 n , 有 $\Gamma(n + 1) = n\Gamma(n) = n!$

证明: (1) $\Gamma(1) = \int_0^{+\infty} e^{-x} dx = -e^{-x} \Big|_0^{+\infty} = 1$

(2)

$$\begin{aligned} \left[\Gamma\left(\frac{1}{2}\right) \right]^2 &= \int_0^{+\infty} x^{-\frac{1}{2}} e^{-x} dx \int_0^{+\infty} y^{-\frac{1}{2}} e^{-y} dy \\ &= \int_0^{+\infty} 2e^{-x} d\sqrt{x} \int_0^{+\infty} 2e^{-y} d\sqrt{y} \end{aligned}$$

令 $u = \sqrt{x}, v = \sqrt{y}$, 则

$$\begin{aligned} \left[\Gamma\left(\frac{1}{2}\right) \right]^2 &= \int_0^{+\infty} 2e^{-u^2} du \int_0^{+\infty} 2e^{-v^2} dv \\ &= 4 \int_0^{+\infty} \int_0^{+\infty} e^{-(u^2+v^2)} du dv \end{aligned}$$

令 $u = t \cos \theta, v = t \sin \theta$, 则

$$\begin{aligned} \left[\Gamma\left(\frac{1}{2}\right)\right]^2 &= 4 \int_0^{\frac{\pi}{2}} \int_0^{+\infty} r e^{-r^2} dr d\theta \\ &= 4 \int_0^{\frac{\pi}{2}} d\theta \int_0^{+\infty} r e^{-r^2} dr \\ &= 4 \times \frac{\pi}{2} \int_0^{+\infty} \frac{1}{2} e^{-r^2} dr^2 \\ &= \pi \end{aligned}$$

因此 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

(3)

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{+\infty} x^\alpha e^{-x} dx \\ &= - \int_0^{+\infty} x^\alpha de^{-x} \\ &= - \left(x^\alpha e^{-x} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-x} dx^\alpha \right) \\ &= \alpha \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \\ &= \alpha \Gamma(\alpha) \end{aligned}$$

(4) 对于任意正整数 n , 由性质 3 可得

$$\Gamma(n+1) = n\Gamma(n)$$

下面用数学归纳法证明 $\Gamma(n+1) = n!$. 当 $n=1$ 时, 有

$$\Gamma(1+1) = 1 \times \Gamma(1) = 1 = 1!$$

假设当 $n=k$ ($k>1$) 时 $\Gamma(k+1) = k!$ 成立, 则

$$\Gamma((k+1)+1) = (k+1)\Gamma(k+1) = (k+1)k! = (k+1)!$$

归纳法完成. □

备注

由上述性质可知, 伽马函数是阶乘运算在实数上的扩展.

将式 (2.4.10) 变形可得

$$\int_0^{+\infty} \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} dx = 1$$

令 $x = \lambda t$, 得

$$\int_0^{+\infty} \frac{(\lambda t)^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} d(\lambda t) = \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt = 1$$

这样就得到了伽马分布的定义.

定义 2.4.10

若随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

则称 X 服从伽马分布 (Gamma distribution), 记作 $X \sim Ga(\alpha, \lambda)$, 其中 $\alpha > 0$ 为形状参数 (shape parameter), $\lambda > 0$ 为尺度参数 (scale parameter).

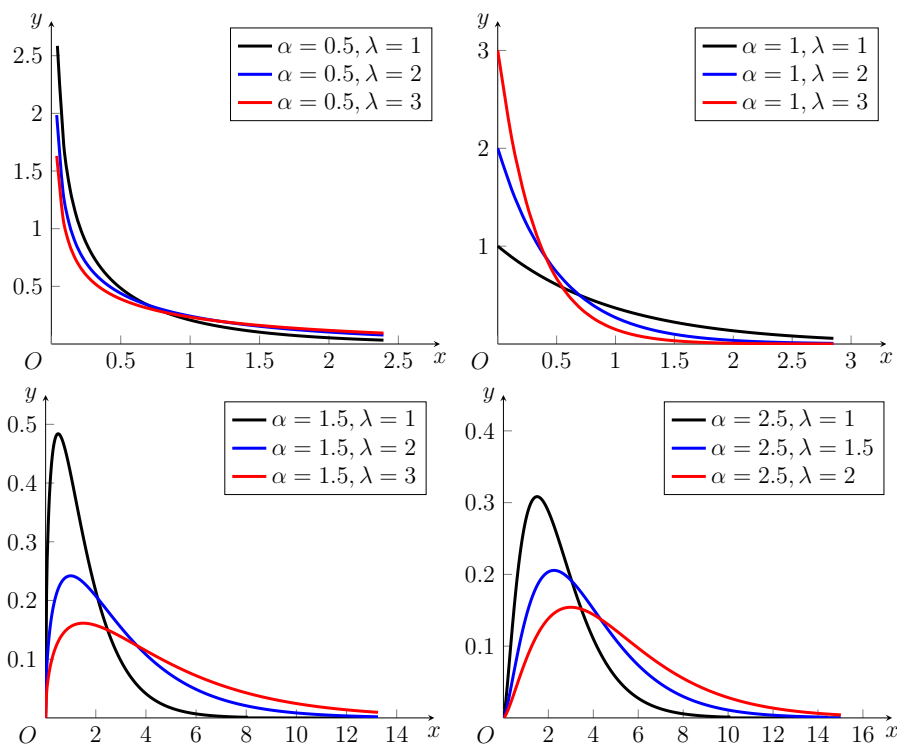


图 2.9 伽马分布的概率密度

从图 2.9 可以看出:

- 当 $0 < \alpha < 1$ 时, $f(x)$ 是严格单调下降函数, 且在 $x=0$ 处有奇点.
- 当 $\alpha=1$ 时, $f(x)$ 是严格单调下降函数, 且 $f(0)=\lambda$.
- 当 $1 < \alpha \leq 2$ 时, $f(x)$ 是单峰函数, 先上凸、后下凸, 在 $x = \frac{\alpha-1}{\lambda}$ 处取得最大值.
- 当 $\alpha > 2$ 时, $f(x)$ 是单峰函数, 先下凸、中间上凸、后下凸, 在 $x = \frac{\alpha-1}{\lambda}$ 处取得最大值. 此时 α 越大, $f(x)$ 越近似于正态分布; α 越小, 曲线的偏斜程度越严重.

伽马分布的特例:

1. 当 $\alpha=1$ 时, 伽马分布 $Ga(1, \lambda)$ 就是指数分布 $Exp(\lambda)$.
2. 当 $\alpha = \frac{n}{2}, \lambda = \frac{1}{2}$ 时, 伽马分布 $Ga(\frac{n}{2}, \frac{1}{2})$ 是自由度为 n 的卡方分布 $\chi^2(n)$.

结论 (泊松分布与伽马分布的关系)

若在 $(0, t)$ 时间段内随机事件 A 发生的次数 $N(t)$ 服从参数为 λt 的泊松分布, 则 A 第 n 次发生的时间 S_n 服从伽马分布 $Ga(n, \lambda)$.

证明: 事件 $\{S_n \leq t\}$ 表示事件 A 第 n 次发生的时间小于等于 t , 这意味着 $(0, t)$ 时间段内事件 A 发生的次数 $N(t) \geq n$, 因此 $\{S_n \leq t\} = \{N(t) \geq n\}$. 于是 S_n 的分布函数为

$$\begin{aligned} F(t) &= P(S_n \leq t) \\ &= P(N(t) \geq n) \\ &= 1 - P(N(t) \leq n-1) \\ &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \end{aligned}$$

考虑下列积分

$$\int_t^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = \frac{1}{\Gamma(n)} \int_t^{+\infty} (\lambda x)^{n-1} e^{-\lambda x} d(\lambda x)$$

令 $\lambda x = u$, 则

$$\begin{aligned} \int_t^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx &= \frac{1}{\Gamma(n)} \int_{\lambda t}^{+\infty} u^{n-1} e^{-u} du \\ &= \frac{1}{\Gamma(n)} \int_{\lambda t}^{+\infty} -u^{n-1} d(e^{-u}) \\ &= -\frac{1}{\Gamma(n)} \left(u^{n-1} e^{-u} \Big|_{\lambda t}^{+\infty} - \int_{\lambda t}^{+\infty} e^{-u} du^{n-1} \right) \\ &= \frac{1}{\Gamma(n)} \int_{\lambda t}^{+\infty} (n-1) u^{n-2} e^{-u} du + \frac{(\lambda t)^{n-1} e^{-\lambda t}}{\Gamma(n)} \\ &= \frac{n-1}{(n-1)\Gamma(n-1)} \int_{\lambda t}^{+\infty} -u^{n-2} d(e^{-u}) + \frac{(\lambda t)^{n-1} e^{-\lambda t}}{\Gamma(n)} \\ &= -\frac{1}{\Gamma(n-1)} \left(u^{n-2} e^{-u} \Big|_{\lambda t}^{+\infty} - \int_{\lambda t}^{+\infty} e^{-u} du^{n-2} \right) + \frac{(\lambda t)^{n-1} e^{-\lambda t}}{\Gamma(n)} \\ &= \frac{n-2}{\Gamma(n-1)} \int_{\lambda t}^{+\infty} u^{n-3} e^{-u} du + \frac{(\lambda t)^{n-1} e^{-\lambda t}}{\Gamma(n)} + \frac{(\lambda t)^{n-2} e^{-\lambda t}}{\Gamma(n-1)} \\ &= \frac{1}{\Gamma(n-2)} \int_{\lambda t}^{+\infty} u^{n-3} e^{-u} du + \frac{(\lambda t)^{n-1} e^{-\lambda t}}{\Gamma(n)} + \frac{(\lambda t)^{n-2} e^{-\lambda t}}{\Gamma(n-1)} \\ &= \dots \\ &= \frac{1}{\Gamma(1)} \int_{\lambda t}^{+\infty} e^{-u} du + \sum_{k=1}^{n-1} \frac{(\lambda t)^k}{\Gamma(k+1)} e^{-\lambda t} \\ &= e^{-\lambda t} + \sum_{k=1}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ &= \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \end{aligned}$$

即

$$\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = \int_t^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx$$

因此

$$F(t) = 1 - \int_t^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx$$

由伽马分布的定义可知

$$\int_0^{+\infty} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx = 1$$

所以

$$F(t) = \int_0^t \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} dx$$

S_n 的分布函数与伽马分布 $Ga(n, \lambda)$ 的分布函数相同, 因此 $S_n \sim Ga(n, \lambda)$. □

2.4.11 贝塔分布

定义 2.4.11

函数

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (2.4.11)$$

称为贝塔函数, 也叫欧拉第一积分, 其中 $a > 0, b > 0$.

性质 2.4.7

$$B(a, b) = B(b, a) \quad (2.4.12)$$

证明: 令 $y = 1 - x$, 得

$$B(a, b) = \int_1^0 (1-y)^{a-1} y^{b-1} d(1-y) = \int_0^1 (1-y)^{a-1} y^{b-1} dy = B(b, a) \quad \square$$

性质 2.4.8

$$B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)} \quad (2.4.13)$$

证明:

$$\Gamma(a) \Gamma(b) = \int_0^{+\infty} \int_0^{+\infty} x^{a-1} y^{b-1} e^{-(x+y)} dx dy$$

令 $x = uv$, $y = u(1-v)$, 其雅可比行列式 $J = -u$, 则

$$\begin{aligned} \Gamma(a) \Gamma(b) &= \int_0^1 \int_0^{+\infty} (uv)^{a-1} [u(1-v)]^{b-1} e^{-u} u du dv \\ &= \int_0^{+\infty} u^{a+b-1} e^{-u} du \int_0^1 v^{a-1} (1-v)^{b-1} dv \\ &= \Gamma(a+b) B(a, b) \end{aligned}$$

整理得

$$B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$$

□

定义 2.4.12

若随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases} \quad (2.4.14)$$

则称 X 服从贝塔分布, 记作 $X \sim Be(a, b)$, 其中 $a > 0, b > 0$, a, b 称为形状参数.

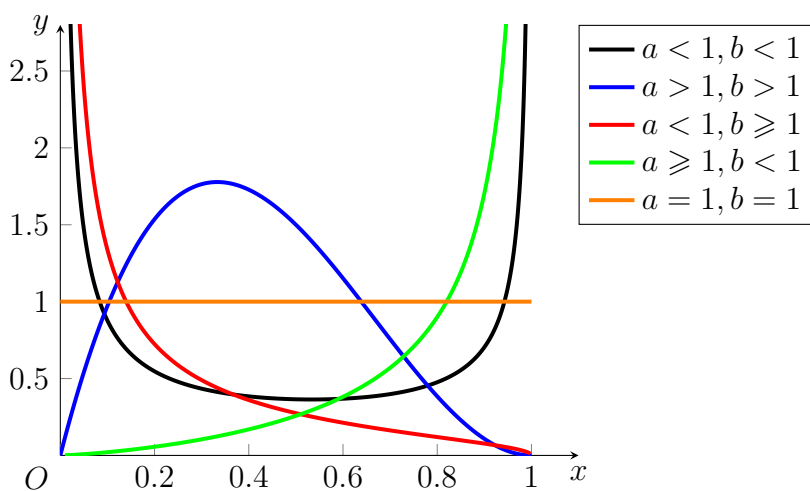


图 2.10 贝塔分布

从图 2.10 可以看出:

- 当 $a < 1, b < 1$ 时, 贝塔分布的概率密度是下凸的 U 形函数.
- 当 $a > 1, b > 1$ 时, 贝塔分布的概率密度是上凸的单峰函数.
- 当 $a < 1, b \geq 1$ 时, 贝塔分布的概率密度是下凸的单调递减函数.
- 当 $a \geq 1, b < 1$ 时, 贝塔分布的概率密度是下凸的单调递增函数.
- 当 $a = 1, b = 1$ 时, 贝塔分布退化为均匀分布 $U(0, 1)$.

2.4.12 柯西分布

定义 2.4.13

若随机变量 X 的概率密度为

$$f(x) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2} \quad (2.4.15)$$

则称随机变量 X 服从柯西分布 (Cauchy distribution), 也称为柯西-洛伦兹分布, 记作 $X \sim C(\gamma, x_0)$, 其中 γ 称为尺度参数, x_0 称为位置参数.

柯西分布 $C(\gamma, x_0)$ 的分布函数为

$$F(x) = \frac{1}{\pi} \arctan \left(\frac{x - x_0}{\gamma} \right) + \frac{1}{2} \quad (2.4.16)$$

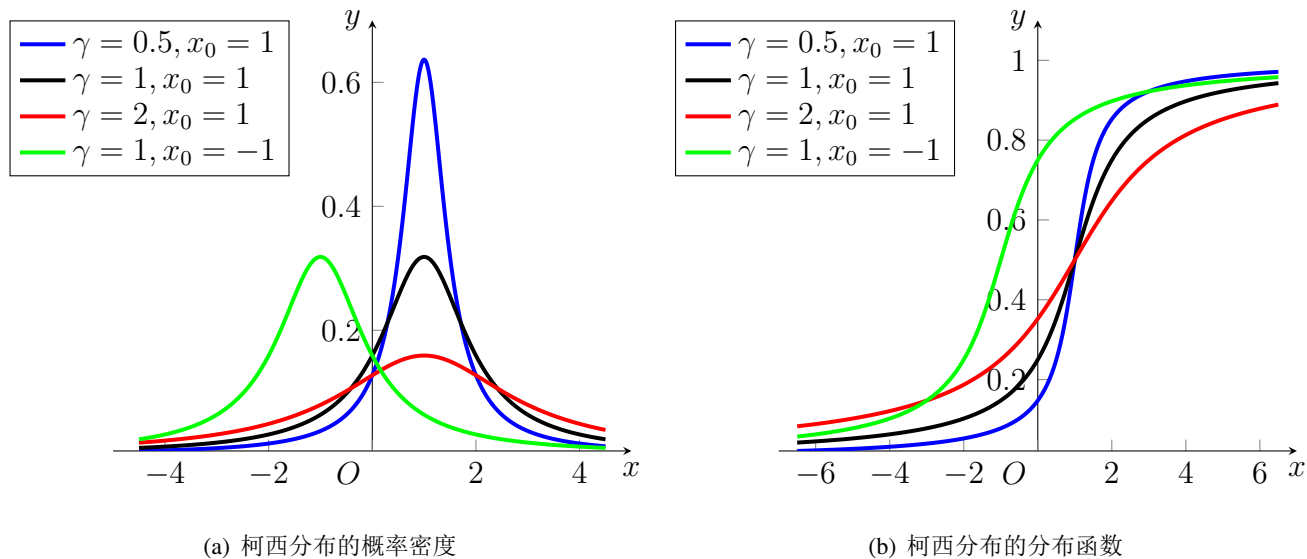


图 2.11 柯西分布

当 $\gamma = 1, x_0 = 0$ 时, 称为标准柯西分布, 此时概率密度和分布函数为

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

$$F(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$$

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

第三章 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 3.1.1

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 3.1.2

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

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

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.

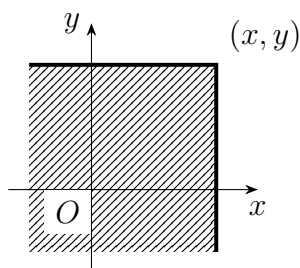


图 3.1

分布函数 $F(x, y)$ 具有如下性质:

性质 3.1.1

$0 \leq F(x, y) \leq 1$, 且

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

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

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

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

性质 3.1.2

$F(x, y)$ 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 $F(x_1, y) \leq F(x_2, y)$; 对于任意固定的 x , 当 $y_1 < y_2$ 时, 有 $F(x, y_1) \leq F(x, y_2)$.

性质 3.1.3

$F(x, y)$ 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 3.1.4

对于任意的 $x_1 < x_2, y_1 < y_2$, 有

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

3.1.2 边缘分布

定义 3.1.3

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.

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

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

3.1.3 随机变量的独立性

定义 3.1.4

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 3.2.1

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 $(x_i, y_j) (i, j = 1, 2, \dots)$, 记事件 $\{X = x_i\}$ 与 $\{Y = y_j\}$ 的交事件为 $\{X = x_i, Y = y_j\}$ 。

定义 3.2.2

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

$$P\{X = x_i, Y = y_j\} = p_{ij}, \quad i, j = 1, 2, \dots \quad (3.2.1)$$

且有

$$\begin{aligned} p_{ij} &\geq 0 \\ \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} &= 1 \end{aligned}$$

则称式 (3.2.1) 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				
	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 $P\{X = x_i, Y = y_j\} = p_{ij} (i, j = 1, 2, \dots)$, 则 (X, Y) 的分布函数为

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

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

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

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 $\{X = x_i, Y = y_j\} (j = 1, 2, \dots)$ 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 $\sum_{j=1}^{\infty} p_{ij} = p_{i\cdot}$, 则有

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

同理可得

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

定义 3.2.3

设二维离散型随机变量 (X, Y) 的概率分布为

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

随机变量 X 和 Y 的概率分布

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

$$P\{Y = y_j\} = p_{\cdot j}, \quad j = 1, 2, \dots$$

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

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

(X, Y) 关于 X 和关于 Y 的边缘概率分布依次为

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

$$P\{Y = y_j\} = p_{\cdot j}, j = 1, 2, \dots$$

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

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

即

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

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 3.3.1

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

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

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 3.3.1

$$f(x, y) \geq 0$$

性质 3.3.2

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx \, dy = 1$$

性质 3.3.3

如果 $f(x, y)$ 在点 (x, y) 处连续，则

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

性质 3.3.4

设 G 是 xOy 平面上的一个区域, 则

$$P\{(X, Y) \in G\} = \iint_G f(x, y) \, dx dy$$

$z = f(x, y)$ 表示空间 $Oxyz$ 中的一张曲面. 性质 3.3.1 和性质 3.3.1 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 3.3.1 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 3.3.2

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

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) \, dy, \quad x \in \mathbf{R} \\ f_Y(y) &= \int_{-\infty}^{+\infty} f(x, y) \, dx, \quad y \in \mathbf{R} \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

$$dt = \frac{1}{\sigma_2\sqrt{1-\rho^2}} dy$$

因此

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

因为

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

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

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

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 3.4.1

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

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \dots$$

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

$$P\{Y = y_j | X = x_i\} = \frac{P\{X = x_i, Y = y_j\}}{P\{X = x_i\}} = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \dots$$

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 3.4.2

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 $P\{y - \varepsilon < Y \leq y + \varepsilon\} > 0$. 如果对于任意实数 x , 极限

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

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

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

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

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

由此可得在条件 $Y = y$ 下 X 的条件概率密度为

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

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

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

例题 3.5.1

设 (X, Y) 的概率分布为

X \ Y	-1	1	2
	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$
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

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

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

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

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

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

3.5.2.1 $Z = X + Y$ 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

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

证明: 首先求 Z 的分布函数

$$\begin{aligned} F_Z(z) &= P\{Z \leq z\} = P\{X + Y \leq z\} \\ &= \iint_{x+y \leq z} f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f(x, y) dy dx \end{aligned}$$

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

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) \mathrm{d}x \\ f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) \mathrm{d}y \end{aligned}$$

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

$$f_Z(z) = f_X * f_Y = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) \mathrm{d}x = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) \mathrm{d}y \quad (3.5.1)$$

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 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

3.5.2.2 $M = \max(X, Y)$ 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 $\{M \leq z\} = \{X \leq z, Y \leq z\}$, 而 X 和 Y 相互独立, 所以事件 $\{X \leq z\}$ 与事件

$\{Y \leq z\}$ 相互独立, 由此可得

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

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

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

3.5.2.3 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

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

当 $z \geq 0$ 时, 有

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

因此, 随机变量 Z 的分布函数为

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

由此可得 Z 的概率密度为

$$f_Z(z) = F'_Z(z) = \begin{cases} \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} & z \geq 0 \\ 0 & z < 0 \end{cases}$$

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

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

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

3.6 n 维随机变量

定义 3.6.1

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, \dots, X_n , 称由它们构成的向量 (X_1, X_2, \dots, X_n) 为 n 维随机向量或 n 维随机变量.

定义 3.6.2

对于任意 n 个实数 x_1, x_2, \dots, x_n , 称 n 元函数

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

为 n 维随机变量 (X_1, X_2, \dots, X_n) 的分布函数, 或称为随机变量 X_1, X_2, \dots, X_n 的联合分布函数.

定义 3.6.3

如果 n 维随机变量 (X_1, X_2, \dots, 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, \dots, X_n 的联合概率分布) 为

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

定义 3.6.4

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

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) dt_n \cdots dt_2 dt_1$$

则称 (X_1, X_2, \dots, X_n) 为 n 维连续型随机变量, 称 $f(x_1, x_2, \dots, x_n)$ 为 (X_1, X_2, \dots, X_n) 的概率密度或 X_1, X_2, \dots, 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, \dots, X_n) 的分布函数 $F(x_1, x_2, \dots, x_n)$, 则可确定 (X_1, X_2, \dots, X_n) 的 k ($1 \leq k < n$) 维边缘分布函数: 在 $F(x_1, x_2, \dots, x_n)$ 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, \dots, X_n) 关于 X_1 的边缘分布函数为

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

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

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

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

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

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

$$f_{X_1 X_2 X_3}(x_1, x_2, x_3) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) \mathrm{d}x_4 \mathrm{d}x_5 \cdots \mathrm{d}x_n, \quad x_1, x_2, x_3 \in \mathbf{R}$$

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

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

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

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

定义 3.6.5

如果对于任意 n 个实数 x_1, x_2, \dots, x_n , 有

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

则称随机变量 X_1, X_2, \dots, X_n 是相互独立的.

结论

如果 (X_1, X_2, \dots, X_n) 是 n 维离散型随机变量, 则 X_1, X_2, \dots, X_n 相互独立的充分必要条

件是: 对于 (X_1, X_2, \dots, X_n) 的任意一组可能取的值 $x_{i_1}, x_{i_2}, \dots, x_{i_n}$, 有

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

结论

如果 (X_1, X_2, \dots, X_n) 是 n 维连续型随机变量, 则 X_1, X_2, \dots, 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, \dots, x_n , 有

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

结论

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

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

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, \dots, X_n 相互独立, 并且 $X_i \sim N(\mu_i, \sigma_i^2) (i = 1, 2, \dots, n)$, 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 $\sum_{i=1}^n c_i X_i$ 仍然服从正态分布, 且有 $\sum_{i=1}^n c_i X_i \sim N(\sum_{i=1}^n c_i \mu_i, \sum_{i=1}^n c_i^2 \sigma_i^2)$, 其中 c_1, c_2, \dots, c_n 是不全为零的常数.

定义 3.6.6

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

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

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

结论

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

结论

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

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

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

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

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

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

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

第四章 随机变量的数字特征

4.1 数学期望

4.1.1 数学期望的定义

定义 4.1.1

设离散型随机变量 X 的概率分布为 $P(X = x_k) = p_k, k = 1, 2, \dots$, 如果无穷级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 即

$$\sum_{k=1}^{\infty} |x_k| p_k < \infty$$

则称无穷级数 $\sum_{k=1}^{\infty} x_k p_k$ 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 $\int_{-\infty}^{+\infty} x f(x) dx$ 绝对收敛, 即

$$\int_{-\infty}^{+\infty} |x| f(x) dx < \infty$$

则称反常积分 $\int_{-\infty}^{+\infty} x f(x) dx$ 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

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

备注

以上定义中, 要求无穷级数绝对收敛的目的在于使数学期望唯一. 因为随机变量的取值可正可负, 取值次序可先可后, 如果此无穷级数绝对收敛, 则可保证其和不受次序变动的影响.

由于有限项的和不受次序变动的影响, 因此取有限个可能值的随机变量的数学期望总

是存在的.

数学期望的物理意义是重心. 如果把概率 $P(X = x_k) = p_k$ 看做点 x_k 上的质量, 概率分布看做质量在 x 轴上的分布, 则 X 的数学期望 $E(X)$ 就是该质量分布的重心所在位置.

4.1.2 随机变量函数的数学期望

定理 4.1.1

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 $P(X = x_k) = p_k, k = 1, 2, \dots$, 如果无穷级数 $\sum_{k=1}^{\infty} g(x_k)p_k$ 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 $\int_{-\infty}^{+\infty} g(x)f(x) dx$ 绝对收敛, 则随机变量 Y 的数学期望为

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

定理 4.1.2

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 $P(X = x_i, Y = y_j) = p_{ij}, i, j = 1, 2, \dots$, 如果无穷级数 $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} g(x_i, y_j) p_{ij}$ 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

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y) dx dy$$

绝对收敛, 则随机变量 Z 的数学期望为

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

4.1.3 数学期望的性质

性质 4.1.1

若 C 是常数, 则 $E(C) = C$.

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

性质 4.1.2

对于任意常数 C , 有 $E(CX) = CE(X)$.

证明: 若 X 为离散型随机变量, 其概率分布为 $P(X = x_i) = p_i, i = 1, 2, \dots$, 则

$$E(CX) = \sum_{i=1}^{\infty} Cx_i p_i = C \sum_{i=1}^{\infty} x_i p_i = CE(X)$$

若 X 为连续型随机变量, 其概率密度为 $f(x)$, 则

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

\square

性质 4.1.3

$E(X + Y) = E(X) + E(Y)$

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

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

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

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

□

一般地, 对任意的两个函数 $g_1(x)$ 和 $g_2(x)$, 有

$$E[g_1(x) \pm g_2(x)] = E[g_1(x)] \pm E[g_2(x)]$$

综合性质 4.1.1、4.1.2、4.1.3, 有

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

性质 4.1.4

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 $P(X = x_i, Y = y_j) = p_{ij}$, $i, j = 1, 2, \dots$, (X, Y) 关于 X 的边缘概率分布为 $P(X = x_i) = p_{i\cdot}$, $i = 1, 2, \dots$, (X, Y) 关于 Y 的边缘概率分布为 $P(Y = y_j) = p_{\cdot j}$, $j = 1, 2, \dots$. 因为 X 和 Y 相互独立, 所以 $p_{ij} = p_{i\cdot} p_{\cdot j}$, 从而有

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

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 $f(x, y) = f_X(x) f_Y(y)$, 因此

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) \, dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) \, dx dy \\ &= \int_{-\infty}^{+\infty} x f_X(x) \, dx \int_{-\infty}^{+\infty} y f_Y(y) \, dy \\ &= E(X) E(Y) \end{aligned}$$

□

性质 4.1.5 (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

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

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 $g(t) \geq 0$, 所以有

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

从而

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

□

4.1.4 常见概率分布的数学期望

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

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

□

结论

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

证明: X 的概率分布为 $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$, $k = 0, 1, 2, \dots$, 因此

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

□

结论

若随机变量 $X \sim Ge(p)$, 则 $E(X) = \frac{1}{p}$.

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

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

□

结论

若随机变量 $X \sim H(n, N, M)$, 则 $E(X) = \frac{nM}{N}$.

证明: 首先给出一些重要的引理.

引理 4.1.1

对于正整数 n, m , 如果 $n \geq m$, 则对任意正整数 $a \leq m$, 有

$$m(m-1) \cdots (m-a+1)C_n^m = n(n-1) \cdots (n-a+1)C_{n-a}^{m-a} \quad (4.1.1)$$

$$\frac{C_{n-a}^{m-a}}{C_n^m} = \frac{m(m-1) \cdots (m-a+1)}{n(n-1) \cdots (n-a+1)} \quad (4.1.2)$$

证明:

$$\begin{aligned}
 m(m-1) \cdots (m-a+1)C_n^m &= m(m-1) \cdots (m-a+1) \frac{n!}{m!(n-m)!} \\
 &= n(n-1) \cdots (n-a+1) \frac{(n-a)!}{(m-a)![(n-a)-(m-a)]!} \\
 &= n(n-1) \cdots (n-a+1)C_{n-a}^{m-a}
 \end{aligned}$$

式 (4.1.1) 得证. 由式 (4.1.1) 整理即得式 (4.1.2).

□

引理 4.1.2

对于正整数 M, N, n , 如果 $N \geq n$ 且 $N \geq M$, 则对任意正整数 a , 当 $a \leq M, a \leq n$ 时, 有

$$\sum_{k=\max\{a, n+M-N\}}^{\min\{M, n\}} C_{M-a}^{k-a} C_{N-M}^{n-k} = C_{N-a}^{n-a} \quad (4.1.3)$$

证明: 构造恒等式

$$(1+x)^{M-a}(1+x)^{N-M} = (1+x)^{N-a}$$

利用二项式定理将恒等式两侧展开, 得

$$\left(\sum_{i=0}^{M-a} C_{M-a}^i x^i \right) \left(\sum_{j=0}^{N-M} C_{N-M}^j x^j \right) = \sum_{k=0}^{N-a} C_{N-a}^k x^k$$

等号左侧展开式中 x^{n-a} 项的系数具有如下形式:

$$\sum_{k=m}^r C_{M-a}^{k-a} C_{N-M}^{n-k}$$

其中

$$\begin{cases} 0 \leq k-a \leq M-a \\ k-a \leq n-a \\ 0 \leq n-k \leq N-M \\ n-k \leq n-a \end{cases}$$

即

$$\begin{cases} k \leq M \\ k \leq n \\ k \geq n+M-N \\ k \geq a \end{cases}$$

因此 $\max\{a, n+M-N\} \leq k \leq \min\{M, n\}$, 即等号左侧展开式中 x^{n-a} 项的系数为

$$\sum_{k=\max\{a, n+M-N\}}^{\min\{M, n\}} C_{M-a}^{k-a} C_{N-M}^{n-k}$$

等号右侧展开式中 x^{n-a} 项的系数为 C_{N-a}^{n-a} , 而左右两侧的展开式中 x^{n-a} 项的系数相等, 因此

$$\sum_{k=\max\{a, n+M-N\}}^{\min\{M, n\}} C_{M-a}^{k-a} C_{N-M}^{n-k} = C_{N-a}^{n-a} \quad \square$$

下面证明超几何分布的数学期望.

(1) 当 $N = M$ 时, $P(X = n) = 1$, 因此 $E(X) = n = \frac{nM}{N}$.

(2) 当 $N > M = 1$ 时, X 的概率分布为

$$P(X=0) = \frac{C_{N-1}^n}{C_N^n}$$

$$P(X=1) = \frac{C_{N-1}^{n-1} C_1^1}{C_N^n} = \frac{n}{N}$$

此时

$$E(X) = 0 \times \frac{C_{N-1}^n}{C_N^n} + 1 \times \frac{n}{N} = \frac{n}{N} = \frac{nM}{N}$$

(3) 当 $N > M \geq 2$ 时, X 的概率分布为 $P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, $k = m, m+1, m+2, \dots, r$, 则

$$E(X) = \sum_{k=m}^r k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

由于 $k=0$ 的项恒为 0, 可以忽略, 因此 k 的取值范围可以为 $\max\{1, n+M-N\} \leq k \leq r$, 即

$$E(X) = \sum_{k=\max\{1, n+M-N\}}^r k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$

由式 (4.1.1) 可得 $kC_M^k = MC_{M-1}^{k-1}$, 因此

$$E(X) = \sum_{k=\max\{1, n+M-N\}}^r \frac{MC_{M-1}^{k-1} C_{N-M}^{n-k}}{C_N^n} = \frac{M}{C_N^n} \sum_{k=\max\{1, n+M-N\}}^r C_{M-1}^{k-1} C_{N-M}^{n-k}$$

由式 (4.1.3) 可得 $\sum_{k=\max\{1, n+M-N\}}^r C_{M-1}^{k-1} C_{N-M}^{n-k} = C_{N-1}^{n-1}$, 因此

$$E(X) = \frac{M}{C_N^n} C_{N-1}^{n-1}$$

由式 (4.1.2) 可得 $\frac{C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}$, 因此

$$E(X) = \frac{nM}{N}$$

综上, 当 $X \sim H(n, N, M)$ 时, $E(X) = \frac{nM}{N}$. □

结论

若随机变量 $X \sim Nb(r, p)$, 则 $E(X) = \frac{r}{p}$.

证明: X 的概率分布为 $P(X = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r}$, $k = r, r+1, r+2, \dots$, 因此

$$\begin{aligned} E(X) &= \sum_{k=r}^{\infty} k C_{k-1}^{r-1} p^r (1-p)^{k-r} \\ &= \sum_{k=r}^{\infty} k \frac{r}{k} C_k^r p^r (1-p)^{k-r} \\ &= \frac{r}{p} \sum_{k=r}^{\infty} C_k^r p^{r+1} (1-p)^{k-r} \end{aligned}$$

设随机变量 $Y \sim Nb(r+1, p)$, 则

$$P(Y = k+1) = C_{k+1-1}^{r+1-1} p^{r+1} (1-p)^{(k+1)-(r+1)} = C_k^r p^{r+1} (1-p)^{k-r}$$

其中 $k+1 = r+1, r+2, r+3, \dots$, 即 $k = r, r+1, r+2, \dots$, 因此

$$\sum_{k=r}^{\infty} C_k^r p^{r+1} (1-p)^{k-r} = 1$$

所以 $E(X) = \frac{r}{p}$. □

结论

若随机变量 $X \sim U(a, b)$, 则 $E(X) = \frac{a+b}{2}$.

证明: X 的概率密度为

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

因此

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$
□

结论

若随机变量 $X \sim Exp(\lambda)$, 则 $E(X) = \frac{1}{\lambda}$.

证明: X 的概率密度为

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

因此

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

□

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

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

因此

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

□

结论

若随机变量 $X \sim Ga(\alpha, \lambda)$, 则 $E(X) = \frac{\alpha}{\lambda}$.

证明:

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x \\
 &= \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\lambda x} \mathrm{d}x \\
 &= \frac{1}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^{+\infty} (\lambda x)^\alpha e^{-\lambda x} \mathrm{d}(\lambda x) \\
 &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} \\
 &= \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha)} \frac{1}{\lambda} \\
 &= \frac{\alpha}{\lambda}
 \end{aligned}$$

□

结论

若随机变量 $X \sim Be(a, b)$, 则 $E(X) = \frac{a}{a+b}$.

证明:

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x \\
 &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^a (1-x)^{b-1} \mathrm{d}x \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} B(a+1, b) \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\
 &= \frac{\Gamma(a+b)}{(a+b)\Gamma(a+b)} \frac{a\Gamma(a)}{\Gamma(a)} \\
 &= \frac{a}{a+b}
 \end{aligned}$$

□

结论

柯西分布的数学期望不存在.

证明: 柯西分布的概率密度为

$$f(x) = \frac{1}{\pi} \frac{\gamma}{(x+x_0)^2 + \gamma^2}, \quad -\infty < x < +\infty$$

而

$$\begin{aligned}
 \int_{-\infty}^{+\infty} |x|f(x) \, dx &= \int_{-\infty}^{+\infty} |x| \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \, dx \\
 &= \int_{-\infty}^0 -x \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \, dx + \int_0^{+\infty} x \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \, dx \\
 &= 2 \int_0^{+\infty} x \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \, dx \\
 &= \frac{1}{\pi} \int_0^{+\infty} \frac{2x}{\gamma \left[\left(\frac{x-x_0}{\gamma} \right)^2 + 1 \right]} \, dx \\
 &= \frac{1}{\pi} \int_0^{+\infty} \frac{2(x-x_0) + 2x_0}{\gamma \left[\left(\frac{x-x_0}{\gamma} \right)^2 + 1 \right]} \, dx \\
 &= \frac{1}{\pi} \left[\int_0^{+\infty} \frac{2(x-x_0)}{\gamma \left[\left(\frac{x-x_0}{\gamma} \right)^2 + 1 \right]} \, dx + \int_0^{+\infty} \frac{2x_0}{\gamma \left[\left(\frac{x-x_0}{\gamma} \right)^2 + 1 \right]} \, dx \right] \\
 &= \frac{1}{\pi} \left[\int_{-\frac{x_0}{\gamma}}^{+\infty} \frac{2t}{t^2 + 1} \, d(\gamma t + x_0) + \int_0^{+\infty} \frac{2x_0}{\left(\frac{x-x_0}{\gamma} \right)^2 + 1} \, d \frac{x-x_0}{\gamma} \right] \\
 &= \frac{1}{\pi} \left[\int_{-\frac{x_0}{\gamma}}^{+\infty} \frac{2\gamma t}{t^2 + 1} \, dt + 2x_0 \arctan \left(\frac{x-x_0}{\gamma} \right) \Big|_0^{+\infty} \right] \\
 &= \frac{1}{\pi} \left[\int_{-\frac{x_0}{\gamma}}^{+\infty} \frac{\gamma}{t^2 + 1} \, d(t^2 + 1) + 2x_0 \arctan \left(\frac{x-x_0}{\gamma} \right) \Big|_0^{+\infty} \right] \\
 &= \frac{1}{\pi} \left[\gamma \ln(t^2 + 1) \Big|_{-\frac{x_0}{\gamma}}^{+\infty} + 2x_0 \arctan \left(\frac{x-x_0}{\gamma} \right) \Big|_0^{+\infty} \right] \\
 &= +\infty
 \end{aligned}$$

积分 $\int_{-\infty}^{+\infty} xf(x) \, dx$ 不绝对收敛, 所以柯西分布的数学期望不存在. □

4.2 方差

4.2.1 方差的定义

定义 4.2.1

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 $P(X = x_k) = p_k, k = 1, 2, \dots$, 则由定义 4.2.1, 有

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

如果 X 是连续型随机变量, 其概率密度为 $f(x)$, 则由定义 4.2.1, 有

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

4.2.2 方差的性质

性质 4.2.1

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

证明:

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

□

推论 4.2.1

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

1. $E(X^2) \geq [E(X)]^2$.
2. 若 $E(X^2) = 0$, 则 $E(X) = 0$, 同时 $D(X) = 0$.

性质 4.2.2

若 C 是常数, 则 $D(C) = 0$.

证明:

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

□

性质 4.2.3

若 C 是常数, 则 $D(CX) = C^2 D(X)$.

证明:

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

□

性质 4.2.4

若 C 是常数, 则 $D(X + C) = D(X)$.

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 4.2.5

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 $E(X - E(X)) = 0$, 再由性质 4.1.4 及方差的定义, 得

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

□

结论

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

证明: 对于任意常数 C , 有

$$E[(X - C)^2] = E(X^2 - 2CX + C^2) = E(X^2) - 2CE(X) + C^2$$

将 $E(X^2) - 2CE(X) + C^2$ 看做以 C 为自变量的二次函数, 其中 $a = 1$, $b = -2E(X)$, $c = E(X^2)$. 根据二次函数的性质可知, 当且仅当 $C = -\frac{b}{2a} = E(X)$ 时函数取得最小值, 最小值为 $E(X^2) - [E(X)]^2 = D(X)$.

因此 $E[(X - C)^2] \geq D(X)$, 当且仅当 $C = E(X)$ 时等号成立, 此时 $E[(X - C)^2]$ 取得最小值. \square

结论

若随机变量 X 仅在区间 $[a, b]$ 上取值, 则

1. $a \leq E(X) \leq b$
2. $D(X) \leq \left(\frac{b-a}{2}\right)^2$

证明: (1) 如果 X 是离散型随机变量, 设 X 的概率分布为 $P(X = x_k) = p_k$, $k = 1, 2, \dots$, 则

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

由于 $x_k \in [a, b]$, 所以 $a \leq x_k \leq b$, 因此

$$a = \sum_{k=1}^{\infty} a p_k \leq E(X) \leq \sum_{k=1}^{\infty} b p_k = b$$

如果 X 是连续型随机变量, 设 X 的概率密度为 $f(x)$, 则

$$E(X) = \int_a^b x f(x) dx$$

由于 $x \in [a, b]$, 所以 $a \leq x \leq b$, 因此

$$a = \int_a^b a f(x) dx \leq E(X) \leq \int_a^b b f(x) dx = b$$

(2) 前面已经证明, 对于任意常数 C , 有 $E[(X - C)^2] \geq D(X)$. 令 $C = \frac{a+b}{2}$, 该不等式仍然成立, 即

$$D(X) \leq E\left[\left(X - \frac{a+b}{2}\right)^2\right]$$

因为 X 的取值范围为 $[a, b]$, 所以 $X - \frac{a+b}{2}$ 的取值范围为 $\left[\frac{a-b}{2}, \frac{b-a}{2}\right]$, $\left(X - \frac{a+b}{2}\right)^2$ 的取值范围为 $\left[0, \left(\frac{b-a}{2}\right)^2\right]$. 根据 (1) 所得结论, 有

$$D(X) \leq E\left[\left(X - \frac{a+b}{2}\right)^2\right] \leq \left(\frac{b-a}{2}\right)^2$$

□

4.2.3 切比雪夫不等式

定理 4.2.1 (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 $\varepsilon > 0$, 有

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

它的等价形式是

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

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

$$P(|X - E(X)| \geq \varepsilon) = \sum_{|x_k - E(X)| \geq \varepsilon} P(X = x_k) = \sum_{|x_k - E(X)| \geq \varepsilon} p_k$$

由 $|x_k - E(X)| \geq \varepsilon$ 得 $(x_k - E(X))^2 \geq \varepsilon^2$, 即

$$\frac{(x_k - E(X))^2}{\varepsilon^2} \geq 1$$

从而有

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

因此

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

如果 X 是连续型随机变量, 设 X 的概率密度为 $f(x)$, 则

$$\begin{aligned}
 P(|X - E(X)| \geq \varepsilon) &= \int_{|x - E(X)| \geq \varepsilon} f(x) \mathrm{d}x \\
 &\leq \int_{|x - E(X)| \geq \varepsilon} \frac{(x - E(X))^2}{\varepsilon^2} f(x) \mathrm{d}x \\
 &\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) \mathrm{d}x \\
 &= \frac{D(X)}{\varepsilon^2}
 \end{aligned}$$

□

在概率论中, 事件 $\{|X - E(X)| \geq \varepsilon\}$ 称为**大偏差**, 其概率 $P(|X - E(X)| \geq \varepsilon)$ 称为**大偏差发生概率**. 切比雪夫不等式给出大偏差发生概率的上界, 这个上界与方差成正比, 方差越大上界就越大. 这说明随机变量的取值大概率集中在数学期望附近, 方差越小, 发生大偏差的概率就越小, 随机变量的取值就越集中.

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 $P(|X - \mu| < 3\sigma) \geq 0.8889$.

定理 4.2.2

若随机变量 X 的方差存在, 则 $D(X) = 0$ 的充分必要条件是 X 以概率 1 取常数 C , 即 $P(X = C) = 1$, 其中 $C = E(X)$.

证明: 先证充分性. 当 $P(X = C) = 1$ 时, 有 $D(X) = D(C) = 0$, 充分性成立.

再证必要性. 当 $D(X) = 0$ 时, $E(X)$ 必定存在. 因为

$$\{|X - E(X)| > 0\} = \bigcup_{n=1}^{\infty} \{|X - E(X)| \geq \frac{1}{n}\}$$

所以有

$$\begin{aligned}
 P(|X - E(X)| > 0) &= P\left(\bigcup_{n=1}^{\infty} \{|X - E(X)| \geq \frac{1}{n}\}\right) \\
 &\leq \sum_{n=1}^{\infty} P(|X - E(X)| \geq \frac{1}{n})
 \end{aligned}$$

根据切比雪夫不等式, 有

$$P(|X - E(X)| \geq \frac{1}{n}) \leq \frac{D(x)}{(1/n)^2}$$

因此

$$P(|X - E(X)| > 0) \leq \sum_{n=1}^{\infty} \frac{D(x)}{(1/n)^2} = 0$$

由概率的非负性可知

$$P(|X - E(X)| > 0) = 0$$

又由于 $P(|X - E(X)| < 0) = 0$, 从而有

$$P(|X - E(X)| = 0) = 1$$

即

$$P(X = E(X)) = 1$$

必要性成立. □

定理 4.2.2 表明, 方差为 0 就意味着随机变量的取值集中在一点.

例题 4.2.1

设 $g(x)$ 为随机变量 X 取值的集合上的非负不减函数, 且 $E(g(X))$ 存在, 证明: 对任意的 $\varepsilon > 0$, 有

$$P(X > \varepsilon) \leq \frac{E(g(X))}{g(\varepsilon)}$$

证明: 因为 $g(x)$ 是非负不减函数, 所以当 $x > \varepsilon$ 时有 $g(x) > g(\varepsilon)$, 进而有 $\frac{g(x)}{g(\varepsilon)} > 1$.

如果 X 是离散型随机变量, 设 X 的概率分布为 $P(X = x_k) = p_k, k = 1, 2, \dots$, 则

$$P(X > \varepsilon) = \sum_{x_k > \varepsilon} p_k \leq \sum_{x_k > \varepsilon} \frac{g(x_k)}{g(\varepsilon)} p_k \leq \frac{1}{g(\varepsilon)} \sum_{k=1}^{\infty} g(x_k) p_k = \frac{E(g(X))}{g(\varepsilon)}$$

如果 X 是连续型随机变量, 设 X 的概率密度为 $f(x)$, 则

$$P(X > \varepsilon) = \int_{\varepsilon}^{+\infty} f(x) dx \leq \int_{\varepsilon}^{+\infty} \frac{g(x)}{g(\varepsilon)} f(x) dx \leq \frac{1}{g(\varepsilon)} \int_{-\infty}^{+\infty} g(x) f(x) dx = \frac{E(g(X))}{g(\varepsilon)}$$

□

4.2.4 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

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

□

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

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

则有

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

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, \dots, X_n 相互独立, 且均服从参数为 p 的 0-1 分布. 由于 $D(X_i) = p(1 - p)$, $i = 1, 2, \dots, n$, 所以根据方差的性质 4.2.5, 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

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

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 $X \sim Ge(p)$, 则 $D(X) = \frac{1-p}{p^2}$.

证明: 若 $X \sim Ge(p)$, 令 $q = 1 - p$, 则

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

由于 $E(X) = \frac{1}{p}$, 因此

$$D(X) = E(X^2) - [E(X)]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

□

结论

若随机变量 $X \sim H(n, N, M)$, 则 $D(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$.

证明: (1) 当 $N = M$ 时, $P(X = n) = 1$, $D(X) = 0$, 待证结论成立.

(2) 当 $N > M = 1$ 时, X 的概率分布为

$$\begin{aligned}
 P(X = 0) &= \frac{C_1^0 C_{N-1}^n}{C_N^n} = \frac{C_{N-1}^n}{C_N^n} \\
 P(X = 1) &= \frac{C_1^1 C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}
 \end{aligned}$$

进而有

$$E(X) = 0 \times \frac{C_{N-1}^n}{C_N^n} + 1 \times \frac{n}{N} = \frac{n}{N}$$

$$E(X^2) = 0^2 \times \frac{C_{N-1}^n}{C_N^n} + 1^2 \times \frac{n}{N} = \frac{n}{N}$$

因此

$$D(X) = E(X^2) - [E(X)]^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n(N-n)}{N^2}$$

待证结论成立.

(3) 当 $N > M \geq 2$ 时, 有

$$\begin{aligned} E(X^2) &= \sum_{k=m}^r k^2 \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \sum_{k=\max\{1, n+M-N\}}^r k(k-1+1) \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \sum_{k=\max\{1, n+M-N\}}^r k(k-1) \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} + \sum_{k=\max\{1, n+M-N\}}^r k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \\ &= \frac{M}{C_N^n} \sum_{k=\max\{2, n+M-N\}}^r (k-1) C_{M-1}^{k-1} C_{N-M}^{n-k} + \frac{nM}{N} \\ &= \frac{M(M-1)}{C_N^n} \sum_{k=\max\{2, n+M-N\}}^r C_{M-2}^{k-2} C_{N-M}^{n-k} + \frac{nM}{N} \\ &= \frac{M(M-1)}{C_N^n} C_{N-2}^{n-2} + \frac{nM}{N} \\ &= M(M-1) \frac{n(n-1)}{N(N-1)} + \frac{nM}{N} \end{aligned}$$

因此

$$D(X) = E(X^2) - [E(X)]^2 = M(M-1) \frac{n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2 = \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

综上, 当 $X \sim H(n, N, M)$ 时, $D(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$. □

结论

若随机变量 $X \sim Nb(r, p)$, 则 $D(X) = \frac{r(1-p)}{p^2}$.

证明:

$$\begin{aligned}
 E(X^2) &= \sum_{k=r}^{\infty} k^2 C_{k-1}^{r-1} p^r (1-p)^{k-r} \\
 &= \sum_{k=r}^{\infty} k^2 \frac{r}{k} C_k^r p^r (1-p)^{k-r} \\
 &= \frac{r}{p} \sum_{k=r}^{\infty} k C_k^r p^{r+1} (1-p)^{k-r} \\
 &= \frac{r}{p} \left[\sum_{k=r}^{\infty} (k+1) C_k^r p^{r+1} (1-p)^{k-r} - \sum_{k=r}^{\infty} C_k^r p^{r+1} (1-p)^{k-r} \right] \\
 &= \frac{r}{p} \left[\sum_{k=r}^{\infty} (k+1) C_k^r p^{r+1} (1-p)^{k-r} - 1 \right]
 \end{aligned}$$

设随机变量 $Y \sim Nb(r+1, p)$, 则

$$P(Y = k+1) = C_{k+1-1}^{r+1-1} p^{r+1} (1-p)^{(k+1)-(r+1)} = C_k^r p^{r+1} (1-p)^{k-r}$$

其中 $k+1 = r+1, r+2, r+3, \dots$, 即 $k = r, r+1, r+2, \dots$, 则

$$\begin{aligned}
 \sum_{k=r}^{\infty} (k+1) C_k^r p^{r+1} (1-p)^{k-r} &= E(Y) = \frac{r+1}{p} \\
 \sum_{k=r}^{\infty} C_k^r p^{r+1} (1-p)^{k-r} &= 1
 \end{aligned}$$

因此

$$E(X^2) = \frac{r}{p} \left(\frac{r+1}{p} - 1 \right)$$

由于 $E(X) = \frac{r}{p}$, 因此

$$D(X) = E(X^2) - [E(X)]^2 = \frac{r}{p} \left(\frac{r+1}{p} - 1 \right) - \left(\frac{r}{p} \right)^2 = \frac{r(1-p)}{p^2}$$

□

结论

若随机变量 $X \sim U(a, b)$, 则 $D(X) = \frac{(b-a)^2}{12}$.

证明: 若随机变量 X 在区间 $[a, b]$ 上服从均匀分布, 则

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_a^b = \frac{b^2 + ab + a^2}{3}$$

由于 $E(X) = \frac{a+b}{2}$, 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 $X \sim \text{Exp}(\lambda)$, 则 $D(X) = \frac{1}{\lambda^2}$.

证明: 由于 $E(X) = \frac{1}{\lambda}$, 则

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

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

□

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

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

□

结论

若随机变量 $X \sim Ga(\alpha, \lambda)$, 则 $D(X) = \frac{\alpha}{\lambda^2}$.

证明:

$$\begin{aligned}
 E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) \, dx \\
 &= \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\lambda x} \, dx \\
 &= \frac{1}{\Gamma(\alpha)} \frac{1}{\lambda^2} \int_0^{+\infty} (\lambda x)^{\alpha+1} e^{-\lambda x} \, d(\lambda x) \\
 &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{1}{\lambda^2} \\
 &= \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\Gamma(\alpha)} \frac{1}{\lambda^2} \\
 &= \frac{\alpha(\alpha+1)}{\lambda^2}
 \end{aligned}$$

由于 $E(X) = \frac{\alpha}{\lambda}$, 因此

$$D(X) = E(X^2) - [E(X)]^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$$

□

结论

若随机变量 $X \sim Be(a, b)$, 则 $D(X) = \frac{ab}{(a+b)^2(a+b+1)}$.

证明:

$$\begin{aligned}
 E(X^2) &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a+1}(1-x)^{b-1} dx \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} B(a+2, b) \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} \\
 &= \frac{\Gamma(a+b)}{(a+b)(a+b+1)\Gamma(a+b)} \frac{a(a+1)\Gamma(a)}{\Gamma(a)} \\
 &= \frac{a(a+1)}{(a+b)(a+b+1)}
 \end{aligned}$$

由于 $E(X) = \frac{a}{a+b}$, 因此

$$D(X) = E(X^2) - [E(X)]^2 = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2(a+b+1)}$$

□

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

4.3 协方差与相关系数

4.3.1 协方差

在方差性质 4.2.5 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 4.3.1

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 $\text{Cov}(X, Y)$, 即

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

性质 4.3.1

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

性质 4.3.2

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

证明:

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

□

性质 4.3.3

对于随机变量 X, Y 和 Z , 有

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

证明:

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

□

性质 4.3.4

对任意随机变量 X, Y , 有

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

4.3.2 相关系数

定义 4.3.2

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

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

当 $\rho_{XY} = 0$ 时, 称 X 与 Y 不相关.

性质 4.3.5

$$|\rho_{XY}| \leq 1$$

证明: 由柯西-施瓦茨不等式可得

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

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

$$|\rho_{XY}| = \left| \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \right| \leq 1 \quad \square$$

性质 4.3.6

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 $\text{Cov}(X, Y) = 0$, 进而 $\rho_{XY} = 0$. □

性质 4.3.7

$|\rho_{XY}| = 1$ 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

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

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

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

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 4.2.2, $D(Y - bX) = 0$ 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. $\text{Cov}(X, Y) = 0$.
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

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

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

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

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

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

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

而

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

因此

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

所以

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

□

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 4.4.1

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

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

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

如果 $E([X - E(X)]^k [Y - E(Y)]^l)$, $k, l = 1, 2, \dots$ 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 $\text{Cov}(X, Y)$ 是 X 和 Y 的二阶混合中心矩.

4.4.2 协方差矩阵

定义 4.4.2

设二维随机变量 (X_1, X_2) 关于 X_1 和 X_2 的二阶中心矩和二阶混合中心矩

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

都存在, 则称矩阵

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

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, \dots, X_n) 关于 X_1, X_2, \dots, X_n 的二阶中心矩和二阶混合中心矩

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

都存在, 则称矩阵

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

为 n 维随机变量 (X_1, X_2, \dots, X_n) 的协方差矩阵.

由于 $c_{ij} = c_{ji}$ ($i \neq j$, $i, j = 1, 2, \dots, n$), 所以 \mathbf{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 $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1) = \rho\sigma_1\sigma_2$, 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

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

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

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

令

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

则

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

因此二维正态随机变量 (X_1, X_2) 的概率密度可以写成

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

设 (X_1, X_2, \dots, X_n) 为 n 维随机变量, 记

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

如果 (X_1, X_2, \dots, X_n) 具有概率密度

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

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

性质 4.4.1

n 维随机变量 (X_1, X_2, \dots, X_n) 服从 n 维正态分布的充分必要条件是:

X_1, X_2, \dots, X_n 的任意线性组合 $k_1X_1 + k_2X_2 + \dots + k_nX_n$ 都服从一维正态分布, 其中 k_1, k_2, \dots, k_n 是不全为零的常数.

性质 4.4.2

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

性质 4.4.3

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

附录

A.1 常见的概率分布

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	$B(n, p)$	$P(X = k) = C_n^k p^k(1 - p)^{n-k}, k = 0, 1, \dots, n$	np	$np(1 - p)$
泊松分布	$P(\lambda)$	$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$	λ	λ
几何分布	$Ge(p)$	$P(X = k) = (1 - p)^{k-1}p, k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
均匀分布	$U(a, b)$	$f(x) = \begin{cases} \frac{1}{b - a}, & a < x < b \\ 0, & \text{其他} \end{cases}$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
指数分布	$Exp(\lambda)$	$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布	$N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$	μ	σ^2