2025 秋　初等概率论

统计推断　课程笔记

作者：招财鱼
组织：清华大学
时间：2025 Fall

If you’re dissatisfied with the world, change yourself. -- Kusanagi Motoko

目录

第一部分

初等概率论

第 1 节 古典概型和概率空间

1.1 概率模型

样本空间是一个集合：每一个概率模型都关联着一个试验，这个试验将产生一个试验结果. 该试验的所有可能结果形成样本空间，记作 Ω。

- 样本空间的试验结果必须满足：互斥，并且完整。
- 样本空间的试验结果可能有限，也可能无限。

事件: 样本空间的子集，即某些试验结果的集合。

定义 1.1 (概率模型的基本构成)

- 样本空间 Ω: 一个试验的所有可能结果的集合;
- 概率：概率就是为试验结果的集合 A (称之为事件) 确定一个非负数 $\mathbf{P}(A)$ (称为事件 A 的概率). 此非负数刻画了我们对事件 A 的认识或所产生的信念程度。

♣

公理 1.1 (概率公理)

1. (非负性) 对一切事件 A，满足 $P(A) \geq 0$.
2. (归一化) $P(\Omega) = 1$.
3. (可加性) 若 $A \cap B = \emptyset$，那么 $P(A \cup B) = P(A) + P(B)$.
   (可列可加性) 若 $A_1, A_2, \ldots$ 是互不相交的事件，那么
   $$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$

♡

定义 1.2 (古典概型 (离散均匀概率))

设样本空间 Ω 由 n 个等可能性的试验结果组成，因此每个试验结果组成的事件 (称为基本事件) 的概率是相等的。由此得到

$$P(A) = \frac{\text{含于事件 A 的试验结果数}}{n}$$

♣

用 $\mathbb{R}^r$ 表示 r 维向量空间,

$$\mathbb{R}^r = \{(x_1, x_2, \ldots, x_r) \mid x_i \in (-\infty, \infty), 1 \leq i \leq r\}.$$

对于 $\mathbb{R}^r$ 的子集 A，用 $m(A) = \int_A dx_1 dx_2 \ldots dx_r$ 表示 A 的体积。

定义 1.3 (几何概率模型 (连续均匀概率))

设样本空间 $\Omega \subset \mathbb{R}^r$ 的体积 $m(\Omega)$ 是正数，且 Ω 中的每个试验结果发生的可能性相同，则对于事件 $A \subset \Omega$，其发生的概率为

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

♣

1.2 概率的性质

性质 概率的若干直观性质：

1. 若 $A \subset B$，则 $P(A) \le P(B)$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $P(A \cup B) \le P(A) + P(B)$.
4. $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.

性质 性质 **1:** $P(\emptyset) = 0$.

证明 因为 $\Omega = \Omega \cup \emptyset \cup \emptyset \cup \dots$，由概率公理 (3) 得

$$\begin{aligned} P(\Omega) &= P(\Omega \cup \emptyset \cup \emptyset \cup \dots) \\ &= P(\Omega) + P(\emptyset) + P(\emptyset) + \dots \end{aligned}$$

由概率公理 (2)，得

$$1 = 1 + P(\emptyset) + P(\emptyset) + \dots$$

即

$$0 = P(\emptyset) + P(\emptyset) + \dots$$

再由概率公理 (1)，得

$$P(\emptyset) = 0$$

性质 性质 **2** **(概率的有限可加性)** 设事件 $A_i, i = 1, \dots, n$，且 $A_i \cap A_j = \emptyset, i \ne j$，则

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

证明 令 $A_{n+1} = A_{n+2} = \dots = \emptyset$，则 $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{\infty} A_i$，且当 $i \ne j$ 时，$A_i \cap A_j = \emptyset$，由概率公理 (3) 和概率性质 1，可得

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

性质 性质 **3:** 如果有事件 A，则 $P(A^c) = 1 - P(A)$

证明 因为 $A \cap A^c = \emptyset, A \cup A^c = \Omega$，由性质 2 和概率公理 (2) 得

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

故 $P(A^c) = 1 - P(A)$

性质 性质 **4:** 如果有事件 A, B，且 $A \subset B$，则 $P(B \setminus A) = P(B) - P(A)$

证明 因为 $B = A \cup (B \setminus A)$ 且 $A \cap (B \setminus A) = \emptyset$，由性质 2 得

$$P(B) = P(A) + P(B \setminus A)$$

即: $P(B \setminus A) = P(B) - P(A)$

推论 1.1

1. **(概率的单调性)** 如果有事件 A, B，且 $A \subset B$，则 $P(A) \le P(B)$.
2. 对任意的事件 A，有 $0 \le P(A) \le 1$.

性质 性质 **5:** 如果有事件 A, B，则 $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

证明 因为 $A\cup B=A\cup(B\setminus A)$，且 $A\cap(B\setminus A)=\emptyset$，由性质 2 和 4 得

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

推论 1.2 (有限次可加性, finite subadditivity, 或 Boole's inequality)

如果有事件 $A_i, i=1,\dots,n$，则

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

♡

性质 性质 **6 (The inclusion-exclusion formula 容斥恒等式):**

如果有事件 $A_i, i=1,\dots,n$，则

$$\begin{aligned}P\left(\bigcup_{i=1}^{n}A_i\right)=&\sum_{1\le i\le n}P(A_i)-\sum_{1\le i<j\le n}P(A_i\cap A_j)\\&+\sum_{1\le i<j<k\le n}P(A_i\cap A_j\cap A_k)-\dots\\&+(-1)^{n-1}P(A_1\cap A_2\cap\cdots\cap A_n)\end{aligned}$$

性质 性质 **7 (Bonferroni's inequality)** 如果有事件 $A_i, i=1,\dots,n$，则

$$P\left(\bigcup_{i=1}^{n}A_i\right)\ge\sum_{1\le i\le n}P(A_i)-\sum_{1\le i<j\le n}P(A_i\cap A_j)$$

(Kounias's inequality)

$$P\left(\bigcup_{i=1}^{n}A_i\right)\le\min_k\left\{\sum_{1\le i\le n}P(A_i)-\sum_{i:i\ne k}P(A_i\cap A_k)\right\}$$

性质 性质 **8 (可列次可加性, σ-subadditivity)** 如果有事件 $A_i, i=1,2,\dots$，则

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

证明 令 $B_1=A_1, B_n=A_n\setminus\bigcup_{i=1}^{n-1}A_i$，则 $\bigcup_{i=1}^{\infty}A_i=\bigcup_{i=1}^{\infty}B_i$ 且 $B_i\cap B_j=\emptyset, i\ne j, B_n\subset A_n$. 由概率公理 (3) 和概率的单调性，得

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

性质 性质 **9:** 如果有事件 $A_i, i=1,2,\dots$，则

$$P\left(\bigcap_{i=1}^{\infty}A_i\right)\ge 1-\sum_{i=1}^{\infty}P(A_i^c)$$

特别地

$$P(A_1\cap A_2)\ge 1-P(A_1^c)-P(A_2^c)$$

证明 由性质 3 和 8 得,

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

特别地, 令 $A_3 = A_4 = \cdots = \Omega$, 由性质 (1) , 可得第二个结论.

1.3 随机抽样与随机分配

有放回抽样与无放回抽样

定义 1.4

- 如果每次将抽到的球在下一次抽球前放回箱子中，则试验称做放回抽样。这时，由 n 个球形成的每一个样本可以表示为向量 $(a_1, \ldots, a_n)$，其中，$a_i(i = 1, \ldots, n)$ 是第 i 次抽到的球的编号。易见，对于放回抽样，每个 a_i 可以是 $\{1, 2, \ldots, M\}$ 中的任何一个数。
- 假设 $n \leq M$，且凡是抽到的球都不再放回，则试验称做无放回抽样。

♣

有序抽样和无序抽样

样本空间的描述，本质上与如下情形有关：诸如 $(4, 1, 2, 1)$ 和 $(1, 4, 2, 1)$ 是认为是 ‘不同’ 的基本事件’，还是 ‘同一基本事件’。因此，习惯上区分两种情形：**有序抽样和无序抽样**。

定义 1.5

- 对有序抽样，由相同元素组成的两个样本，只要其中元素的先后顺序有所不同，就视为 ‘不同’ 的样本。
- 对无序抽样，不管元素的顺序，只要由相同元素组成的样本，都视为 ‘同一个’ 样本。
- 为强调具体样本属于哪一种，对有序样本，使用记号 $(a_1, \ldots, a_n)$，而无序样本则记作 $[a_1, \ldots, a_n]$。

♣

例 1.1

<table>
<tr><th></th><th colspan="2">抽样方式</th><th>基本事件</th></tr>
<tr><td rowspan="4">自含有 M 个球箱子中的 n 次抽样 $(M = 3, n = 2)$ 相应基本空间的结构列表</td><td rowspan="2">放回</td><td>有序</td><td>(i, j), $i, j = 1, 2, 3$</td></tr>
<tr><td>无序</td><td>$[1,1][2,2][3,3][1,2][1,3][2,3]$</td></tr>
<tr><td rowspan="2">不放回</td><td>有序</td><td>$(1,2)(1,3)(2,1)(2,3)(3,1)(3,2)$</td></tr>
<tr><td>无序</td><td>$[1,2][1,3][2,3]$</td></tr>
</table>

随机抽样的样本空间

1. 有放回抽样

1. 放回有序抽样: 样本空间 Ω 具有如下构造:

$$\Omega=\{\omega:\omega=(a_1,\dots,a_n),\quad a_i\in\{1,2,\dots,M\}\},$$

且

$$\#(\Omega)=M^n.$$

2. 放回无序抽样：样本空间 Ω 具有如下构造:

$$\Omega=\{\omega:\omega=[a_1,\dots,a_n],\quad a_i\in\{1,2,\dots,M\}\},$$

且

$$\#(\Omega)=\binom{M+n-1}{n}.$$

2. 无放回抽样

1. 不放回有序抽样：样本空间 Ω 具有如下构造:

$$\Omega=\{\omega:\omega=(a_1,\dots,a_n),a_k\neq a_l,k\neq l,a_i\in\{1,2,\dots,M\}\},$$

且

$$\#(\Omega)=M(M-1)\dots(M-n+1)=P_M^n.$$

2. 不放回无序抽样：样本空间 Ω 具有如下构造:

$$\Omega=\{\omega:\omega=[a_1,\dots,a_n],a_k\neq a_l,k\neq l,a_i\in\{1,2,\dots,M\}\},$$

且

$$\#(\Omega)=\binom{M}{n}.$$

1.4 概率空间

定义 1.6 (事件域或 σ-域或 σ-代数 (σ-field or σ-algebra))

设 Ω 是样本空间, $\mathcal{F}$ 表示 Ω 的某些子集构成的集合，如果 $\mathcal{F}$ 满足以下三个条件:

1. $\Omega\in\mathcal{F}$;
2. 如果 $A\in\mathcal{F}$，则 $A^c\in\mathcal{F}$;
3. 如果 $A_n\in\mathcal{F},n=1,2,\dots$，则 $\bigcup_{n=1}^{\infty}A_n\in\mathcal{F}$

称 $\mathcal{F}$ 是 Ω 上的事件域或 σ-域或 σ-代数, 称 $\mathcal{F}$ 中的元素为事件, 称 $(\Omega,\mathcal{F})$ 为可测空间 (measurable space). ♣

注

1. $\mathcal{F}$ 中每一个事件都是可以分配概率的；圈定了全部我们关心的事件的范围；
2. Ω 的任意子集未必是事件，只有 $\mathcal{F}$ 中的元素才能称之为事件；
3. $\mathcal{F}$ 对集合的各类可列交并补运算都是封闭的，包括事件列的极限运算。
4. $\mathcal{F}$ 中任选一个事件，其概率是否可计算，取决于已知条件；

对于固定的 Ω ，可以构造出多个不同的 Ω 上的事件域。

例 1.2

σ-代数的例子

1. $\mathcal{F}=\{\Omega,\emptyset\}$, 平凡的 σ-代数;
2. $\mathcal{F}=\{\Omega$的所有子集$\}$, 最大的 σ-代数;
3. $\mathcal{F}=\{\Omega,\emptyset,A,A^c\}$ 是包含 A 的最小 σ-代数, 记作 $\mathcal{F}=\sigma(\{A\})$;

$\mathcal{F}$ 的构造

- 如果 A,B 是 Ω 的两个子集, 且 $B\neq A^c$, 那么由 $\mathcal{A}=\{A,B\}$ 所生成的 σ-域 $\mathcal{F}$ 是:
- $\mathcal{F}=\sigma(\mathcal{A})=\{\Omega,\emptyset,A,A^c,B,B^c,A\cap B,A\cap B^c,A^c\cap B,A^c\cap B^c,A\cup B,A^c\cup B,A\cup B^c,A^c\cup B^c,(A\cap B)\cup(A^c\cap B^c),(A\cap B^c)\cup(A^c\cap B)\}$

定义 1.7 (概率或概率测度 (probability measure))

设 $(\Omega,\mathcal{F})$ 是可测空间, $\mathbb{P}$ 是定义在 $\mathcal{F}$ 上的函数, 如果 $\mathbb{P}$ 满足下面三个条件:

1. **(非负性)** 对任意的 $A\in\mathcal{F},\mathbb{P}(A)\geq 0$.
2. **(完全性)** $\mathbb{P}(\Omega)=1$.
3. **(可列可加性, σ - additivity)** 对于 $\mathcal{F}$ 中互不相交 (disjoint) (或互不相容) 的事件 $A_1,A_2,\ldots$, 有

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\mathbb{P}(A_n)$$

称 $\mathbb{P}$ 为 $\mathcal{F}$ 上的概率测度 (probability measure), 简称概率 (probability), 称 $(\Omega,\mathcal{F},\mathbb{P})$ 为概率空间 (probability space).

♣

注

1. 如果 Ω 是可列的，则存在概率分配使得 Ω 的任一子集都可测，当然此时 $\mathcal{F}$ 可以包含 Ω 的所有子集；
2. 如果 Ω 是不可列的，则存在 Ω 的不可测子集，此时这些不可测子集就不能把它们当成事件了，因为我们无法确定其概率。

例 1.3 $(\Omega,\mathcal{F},\mathbb{P})$ 的构造: 简单例子

1. **掷一枚硬币.** $\Omega=\{H,T\},\mathcal{F}=\{\Omega,\emptyset,\{H\},\{T\}\},\mathbb{P}(\{H\})=0.6,\mathbb{P}(\{T\})=0.4$.
2. **掷一枚骰子.** $\Omega=\{1,2,\ldots,6\},\mathcal{F}=2^{\Omega}$ (Ω 的所有子集构成的集合, 称为幂集, power sets), $\mathbb{P}(\{i\})=1/6,i=1,\ldots,6$. 对任意的 $A\in\mathcal{F},\mathbb{P}(A)=\#(A)/6$.
3. **反复掷一枚硬币** (每次正面朝上的概率为 p), 直到正面向上.

$$\Omega=\{T^nH:n\geq 0\}\cup\{T^{\infty}\}$$
$$\mathbb{P}(T^nH)=(1-p)^np,\quad \mathbb{P}(T^{\infty})=\lim_{n\to\infty}(1-p)^n=0$$

对于 $A\in\mathscr{F}$，如果 $\mathbb{P}(A)=1$，称 A 以概率 1 发生或几乎处处发生，这里的几乎处处是指对几乎每个 $\omega\in\Omega$. 几乎处处有时又称为几乎必然，记作 a.s. (almost surely).

注 请注意区分事件域 $\mathcal{F}$、可测空间 $(\Omega,\mathcal{F})$ 与测度空间 $(\Omega,\mathcal{F},\mathbb{P})$：

1. 事件域 $\mathcal{F}$ 包含全部我们关心的事件；其中每一个事件都是可以分配概率的；
2. 对于固定的 Ω, 可以构造出多个不同的 Ω 上的事件域。
3. 对同一个可测空间 $(\Omega,\mathcal{F})$, 可构造不同的概率空间 $(\Omega,\mathcal{F},\mathbb{P})$;
4. $\mathcal{F}$ 中任选一个事件, 其概率是否可计算, 取决于已知条件;

1.5 概率的连续性

定义 1.8 (单调序列)

对给定的事件列 $\{A_i, i=1,2,\dots\}$,

1. 如果 $A_1 \subset A_2 \subset \dots$, 称事件列 $\{A_i, i=1,2,\dots\}$ 是单调递增的;
2. 如果 $A_1 \supset A_2 \supset \dots$, 称事件列 $\{A_i, i=1,2,\dots\}$ 是单调递减的;

单调增序列和单调减序列统称为单调序列.

对于单调增序列 $\{A_i\}$, 则

$$\lim_{n\to\infty} A_n = \bigcup_{i=1}^{\infty} A_i$$

对于单调减序列 $\{A_i\}$, 则

$$\lim_{n\to\infty} A_n = \bigcap_{i=1}^{\infty} A_i$$

♣

定理 1.1

设 $\{A_i\}$ 和 $\{B_j\}$ 是事件列。

1. 如果 $\{A_i\}$ 是单调增序列, 则
$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$$
2. 如果 $\{B_j\}$ 是单调减序列, 则
$$\mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

♡

定义 1.9 (事件列的上极限与下极限)

设 $\{A_i, i=1,2,\dots\}$ 是 Ω 中的事件列, 定义:

1. $\{A_i, i=1,2,\dots\}$ 的上极限, 记作 $\limsup_{n\to\infty} A_n$, 定义为:
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \{\omega\in\Omega : \omega \text{ 属于无穷多个} A_i\}$$
2. $\{A_i, i=1,2,\dots\}$ 的下极限, 记作 $\liminf_{n\to\infty} A_n$, 定义为:
$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \{\omega\in\Omega : \omega \text{ 属于所有的} A_i \text{ 除了有限个之外}\}$$
3. 如果 $\liminf\limits_{n\to\infty} A_n = \limsup\limits_{n\to\infty} A_n$, 则称事件列 $\{A_i\}$ 的极限存在, 记作 $\lim_{n\to\infty} A_n$。

♣

定理 1.2

任意一组事件列 $\{A_i, i=1,2,\dots\}$, 有

$$\mathbb{P}(\limsup_{n\to\infty} A_n) \ge \limsup_{n\to\infty} \mathbb{P}(A_n), \quad \mathbb{P}(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} \mathbb{P}(A_n)$$

♡

证明 记 $B_n = \bigcup_{k=n}^{\infty} A_k$, 则 $\{B_n\}$ 是单调减序列, 有

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

又由概率单调性, 有 $\mathbb{P}(B_n) \ge \mathbb{P}(A_k), \forall k \ge n$.

即 $\mathbb{P}(B_n) \ge \sup_{k\ge n} \mathbb{P}(A_k)$, 代入上式得到第一个不等式.

(注意, 此时 $\sup_{k\ge n} \mathbb{P}(A_k)$ 是数列, 不是事件列; 该数列关于 n 是单调的, 从而保证了其极限存在.)

第二个不等式的证明是类似的.

定理 1.3 (Borel-Cantelli 引理)

设 $\{A_n\}$ 是事件列。

1. 如果
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$
则
$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0.$$
2. 如果 $\{A_n\}$ 是相互独立的，且
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$
则
$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 1.$$

♡

1.6 条件概率

定义 1.10 (条件概率 (conditional probability))

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A, B \in \mathcal{F}$，且 $\mathbb{P}(A) > 0$，则在事件 A 发生的条件下，事件 B 发生的条件概率定义为：
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

♣

定理 1.4

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A \in \mathcal{F}$，且 $\mathbb{P}(A) > 0$，则

1. 对任意的 $B \in \mathcal{F}, \mathbb{P}(B|A) \geq 0$;
2. $\mathbb{P}(\Omega|A) = 1$;
3. 对互不相容的事件列 $\{B_i\}$，有
$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i \middle| A\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i|A).$$

♡

用 $\mathbb{P}_A(\cdot)$ 表示在事件 A 发生的条件下的条件概率，即 $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot|A)$。则 $(\Omega, \mathcal{F}, \mathbb{P}_A)$ 也是一个概率空间。

由于 $\mathbb{P}(A|A) = 1$，条件概率完全集中在 A 上，这样，我们可以将 A 以外的结果排除掉，并将 A 看成新的样本空间 $(A, A \cap \mathcal{F}, \mathbb{P}_A)$。

条件概率依然是一个概率测度，所有关于概率的性质对条件概率都成立。

1.7 乘法公式

定理 1.5 (乘法公式)

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A_i \in \mathcal{F}, i = 1, \ldots, n$，且 $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$，则
$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{n} \mathbb{P}(A_i | A_1 \cap A_2 \cap \cdots \cap A_{i-1}).$$

♡

例 1.4 配对问题 某人写了 n 封信, 将其装入 n 个信封, 并在每个信封上分别任意写上 n 个收信人的一个地址 (不重复), 求:

1. 没有一个信封上所写的地址正确的概率 q_0;
2. 恰有 r 个信封上所写的地址正确的概率 $q_r(r \leq n)$.

证明 (1) 设 $A_i =$ "第 i 个信封上所写的地址正确", $i = 1, 2, \ldots, n$, 则 $\bigcup_{i=1}^n A_i$ 表示 "n 个信封至少有一个信封上所写的地址正确", 故 $q_0 = 1 - \mathbb{P}(\bigcup_{i=1}^n A_i)$。

对任意的 $1 \leq i \leq n$, 有 $\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$。

对任意的 $1 \leq i < j \leq n$, 有

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j|A_i) = \frac{1}{n}\frac{1}{n-1} = \frac{(n-2)!}{n!}.$$

对任意的 $1 \leq i < j < k \leq n$, 有 $\mathbb{P}(A_i \cap A_j \cap A_k) = \dfrac{(n-3)!}{n!}$。

$$\cdots$$

$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \dfrac{1}{n!}$。

从而, 根据容斥原理 (inclusion-exclusion formula):

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n-1}\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \\
&= \binom{n}{1}\mathbb{P}(A_1) - \binom{n}{2}\mathbb{P}(A_1 \cap A_2) + \cdots + (-1)^{n-1}\binom{n}{n}\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \\
&= \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{(n-2)!}{n!} + \cdots + (-1)^{n-1}\frac{1}{n!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}
\end{aligned}$$

因此

$$q_0 = 1 - \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}\right) = \sum_{k=0}^n \frac{(-1)^k}{k!}$$

进一步, 由于 q_0 跟 n 有关, 故一般记为 $q_0(n)$。从而

$$\lim_{n\to\infty} q_0(n) = e^{-1} = 0.36787944117144233\ldots,$$

即, 当 n 非常大的时候, $q_0 \approx 0.37$。

(2) 首先, 指定某 r 个信封正确, 其余 $n-r$ 个不正确的概率是:

$$\underbrace{\frac{1}{n(n-1)\cdots(n-r+1)}}_{\text{r 个正确}} \times \underbrace{q_0(n-r)}_{\text{n-r 个不正确}}$$

其中, 前一项是在 n 个位置中, 指定的 r 个位置都正确的概率。后一项是剩下 $n-r$ 个信封全部装错的概率。

由于我们可以从 n 个信封中任意选择 r 个, 共有 $\binom{n}{r}$ 种选择方式。故所求概率为

$$\begin{aligned}
q_r(n) &= \binom{n}{r}\frac{1}{n(n-1)\cdots(n-r+1)}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!} \\
&= \frac{n!}{r!(n-r)!}\cdot\frac{(n-r)!}{n!}\cdot\sum_{k=0}^{n-r}\frac{(-1)^k}{k!} = \frac{1}{r!}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}
\end{aligned}$$

1.8 全概率公式

定理 1.6 (Total Probability Theorem 全概率公式)

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, $B, A_i \in \mathcal{F}, i = 1, \dots, n$ (n 可以是 ∞), 且 $\mathbb{P}(A_i) > 0$, $\{A_i\}$ 是 Ω 的一个分割, 则

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \cdots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)$$

或者写作

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

♡

注 $\{A_i\}$ 是 Ω 的一个**分割**: $\Omega = \bigcup_{i=1}^{n} A_i$, 且 $A_i \cap A_j = \emptyset, \forall i \neq j, i, j \in \{1, \dots, n\}$.

例 1.5 设一袋中有 n 个白球与 m 个黑球, 现从中**无放回**连续抽取 k 个球, 求第 k 次取得黑球的概率 ($1 \leq k \leq m + n$).

证明 (归纳法) 设 A_i 为事件 "第 i 次抽到黑球"。

显然

$$\begin{aligned}
\mathbb{P}(A_1) &= \frac{m}{n+m}; \\
\mathbb{P}(A_2) &= \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) + \mathbb{P}(A_1^c)\mathbb{P}(A_2|A_1^c) \\
&= \frac{m}{n+m} \cdot \frac{m-1}{n+m-1} + \frac{n}{n+m} \cdot \frac{m}{n+m-1} = \frac{m}{n+m}.
\end{aligned}$$

假设 "无论开始黑球、白球个数是多少, 无放回抽到第 i 次时, 抽到黑球的概率都是初始状态中黑球的比例"。

往证 "无放回抽到第 $i+1$ 次时, 抽到黑球的概率仍是初始状态中黑球的比例" , 即 $\mathbb{P}(A_{i+1}) = \frac{m}{n+m}, i + 1 \leq n + m$。

因为

$$\begin{aligned}
\mathbb{P}(A_{i+1}) &= \mathbb{P}(A_1)\mathbb{P}(A_{i+1}|A_1) + \mathbb{P}(A_1^c)\mathbb{P}(A_{i+1}|A_1^c) \\
&= \frac{m}{n+m}\mathbb{P}(A_{i+1}|A_1) + \frac{n}{n+m}\mathbb{P}(A_{i+1}|A_1^c).
\end{aligned}$$

根据归纳假设, $\mathbb{P}(A_{i+1}|A_1)$ 表示在 $m-1$ 个黑球与 n 个白球的袋中第 i 次摸得黑球的概率:

$$\mathbb{P}(A_{i+1}|A_1) = \frac{m-1}{n+m-1}.$$

同理,

$$\mathbb{P}(A_{i+1}|A_1^c) = \frac{m}{n-1+m}.$$

所以

$$\mathbb{P}(A_{i+1}) = \frac{m}{n+m} \cdot \frac{m-1}{n+m-1} + \frac{n}{n+m} \cdot \frac{m}{n-1+m} = \frac{m}{n+m}.$$

故

$$\mathbb{P}(A_k) = \frac{m}{n+m}.$$

1.9 Bayes 准则

定理 1.7 (Bayes' Rule)

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, $B, A_i \in \mathcal{F}, i = 1, \dots, n$, 且 $\mathbb{P}(B) > 0, \mathbb{P}(A_i) > 0$, $\{A_i\}$ 是 Ω 的一个分割, 则

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j)\mathbb{P}(B|A_j)}, \quad i = 1, \dots, n.$$

♡

第 2 节 随机变量和概率分布

2.1 事件的独立性

"独立性定义"：事件 B 的发生并没有给事件 A 带来新的信息，它没有改变事件 A 发生的概率，即 $\mathbb{P}(A|B) = \mathbb{P}(A)$。

定义 2.1 (独立性 (independence))

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A, B \in \mathcal{F}$，如果

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

则称 A 与 B 相互独立，简称独立。

♣

若 A 和 B 独立，则 (i) A 和 B^c 独立 (ii) A^c 和 B 独立 (iii) A^c 和 B^c 独立。

定义 2.2 (条件独立性 (conditional independence))

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A, B, C \in \mathcal{F}$，且 $\mathbb{P}(C) > 0$，如果

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$$

则称 A 与 B 在给定 C 之下条件独立。

♣

注 条件独立性和独立性之间没有特别的联系，也即二者互不存在蕴含关系。

例 2.1 考虑抛掷两次均匀的硬币。这个试验的四种可能结果都是等可能的。令

- $H_1 = \{$ 第一枚硬币正面向上 $\}$,
- $H_2 = \{$ 第二枚硬币正面向上 $\}$,
- $D = \{$ 两枚硬币的试验结果不同 $\}$。

则事件 H_1 和事件 H_2 是独立的。但是

$$\mathbb{P}(H_i|D) = \frac{1}{2}, \quad \mathbb{P}(H_1 \cap H_2|D) = 0, \quad i = 1, 2.$$

这样,

$$\mathbb{P}(H_1 \cap H_2|D) \neq \mathbb{P}(H_1|D)\mathbb{P}(H_2|D),$$

即 H_1 和 H_2 并不条件独立。

定义 2.3 (一组事件的独立性)

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A_1, \ldots, A_n \in \mathcal{F}$，如果对任意非空子集 $S \subset \{1, 2, \ldots, n\}$，都有

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i),$$

则称 $A_1, A_2, \ldots, A_n$ 是相互独立的。

♣

定义 2.4 (两两独立性 (Pairwise independence))

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间，$A_1, \ldots, A_n \in \mathcal{F}$，如果

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j), \quad \forall i \neq j;\ i, j \in \{1, 2, \ldots, n\},$$

则称 $A_1, A_2, \ldots, A_n$ 是两两独立的。

♣

例 2.2 将一个均匀的正四面体的第一面染上红、黄、蓝三色, 将其它三面分别染上红色、黄色、蓝色。设 A, B, C 分别表示掷一次四面体红色、黄色、蓝色与桌面接触的事件, 则

$$\begin{aligned}\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) &= \frac{1}{2},\\ \mathbb{P}(A \cap B) &= \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B),\\ \mathbb{P}(B \cap C) &= \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C),\\ \mathbb{P}(A \cap C) &= \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C).\end{aligned}$$

但是

$$\mathbb{P}(A \cap B \cap C) = 1/4 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

即: **两两独立并不能保证相互独立!**

2.2 随机变量

定义 2.5 (随机变量 (random variable, r.v., rv))

设 $(\Omega, \mathcal{F})$ 为可测空间, 如果 Ω 上的函数 $X(\omega)$ 满足: 对 $\forall x \in \mathbb{R}$,

$$\{\omega | X(\omega) \leq x\} \in \mathcal{F},$$

则称 $X(\omega)$ 为 $(\Omega, \mathcal{F})$ 上的随机变量, 简称随机变量。通常将随机变量 $X(\omega)$ 简记为 X。♣

注 以后用 $\{X \leq x\}$ 来表示事件 $\{\omega \,|\, X(\omega) \leq x\}$。显然,

1.
$$\{X > x\} = \{X \leq x\}^c \in \mathcal{F}$$

2.
$$\{X < x\} = \bigcup_{n=1}^{\infty} \left\{X \leq x - \frac{1}{n}\right\} \in \mathcal{F}$$

3.
$$\{X \geq x\} = \{X < x\}^c \in \mathcal{F}$$

随机变量严格定义下的性质

用 $\mathbb{R}$ 表示全体实数, 用 $\mathcal{C}$ 表示 $\mathbb{R}$ 中左开右闭的子区间的全体, 即 $\mathcal{C} = \{(a, b]\}$。令 $\mathcal{B} = \sigma(\mathcal{C})$, 通常称 $\mathcal{B}$ 为 **Borel 域**, 称 $\mathcal{B}$ 的元素为 **Borel 集**。

当 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, 下面的定理表明, 对任何 Borel 集 A, $\{X \in A\}$ 都是事件, 于是可以计算概率 $\mathbb{P}(X \in A)$。

定理 2.1

设 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量, 则对任意的 Borel 集 A, 有

$$\{X \in A\} \overset{\text{def}}{=} \{\omega | X(\omega) \in A\} \in \mathcal{F}.$$

♡

此后, 相对于原始概率空间 $\Omega, \mathscr{F}, \mathbb{P}$, 我们将更多关注经过 X 映射后的概率空间 $\mathbb{R}, \mathscr{F}, \mathbb{P}$.

随机变量的函数

定理 2.2

如果 X 是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量, $g(x)$ 是可测函数, 则 $Y=g(X)$ 是 $(\Omega, \mathcal{F})$ 上的随机变量。♡

连续函数、阶梯函数、单调函数以及这些函数的线性组合都是可测函数。完全类似的可以证明, 如果 $X_1, X_2, \ldots, X_n$ 都是可测空间 $(\Omega, \mathcal{F})$ 上的随机变量, $g(x_1, x_2, \ldots, x_n)$ 是 n 元可测函数, 则 $Y=g(X_1, X_2, \ldots, X_n)$ 是 $(\Omega, \mathcal{F})$ 上的随机变量。(n 维 Borel 域)

2.3 随机变量的独立性

定义 2.6 (随机变量的相互独立性)

设 $X_1, \ldots, X_n$ 是 $(\Omega, \mathcal{F})$ 上的随机变量, 如果对任意的实数 $x_1, \ldots, x_n$ 都有

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)=\mathbb{P}(X_1 \leq x_1) \cdots \mathbb{P}(X_n \leq x_n),$$

称随机变量 $X_1, \ldots, X_n$ 相互独立。♣

定理 2.3

设 $X_1, \ldots, X_n$ 相互独立, 则对任何 Borel 集 $A_1, \ldots, A_n$, 事件

$$\{X_1 \in A_1\}, \{X_2 \in A_2\}, \ldots, \{X_n \in A_n\}$$

相互独立。♡

随机变量序列或函数的独立性

定义 2.7 (独立序列)

如果对任意的 n, $X_1, \ldots, X_n$ 相互独立, 则称随机变量序列 $\{X_i\}$ 相互独立, 此时称 $\{X_i\}$ 为独立序列。♣

定义 2.8 (独立同分布序列)

如果随机变量 $X_1, X_2, \ldots$ 是相互独立并且具有相同的分布函数, 则称 $X_1, X_2, \ldots$ 独立同分布 (independent and identically distributed, i.i.d.)。♣

定理 2.4

设随机变量 $X_1, \ldots, X_n$ 相互独立, $g_1(x), \ldots, g_n(x)$ 是一元实可测函数, $\phi(x_1, x_2, \ldots, x_k)$ 是 k 元实可测函数, 则

1. 随机变量 $g_1(X_1), \ldots, g_n(X_n)$ 相互独立;
2. 随机变量 $\phi(X_1, X_2, \ldots, X_k), X_{k+1}, \ldots, X_n$ 相互独立。♡

例 2.3 若 (X_1, X_2) 与 (X_3, X_4) 独立, 则 X_1/X_2^2 与 $\dfrac{X_3+X_4}{2}$ 独立。

2.4 常见离散型随机变量

定义 2.9 (离散型随机变量 (discrete random variable))

如果随机变量 X 只取有限个值 $x_1, \ldots, x_m$ 或者可列个值 $x_1, x_2, \ldots$, 则称 X 是离散型随机变量, 简称离散随机变量。♣

定义 2.10 (概率分布、概率分布列)

设 X 为离散型随机变量, 称

$$\mathbb{P}(X = x_k) = p_k, \quad k \geq 1$$

为 X 的概率分布, 称 $\{p_k\}$ 是概率分布列, 简称分布列 (Probability mass function, PMF)。

当分布列 $\{p_k\}$ 的规律性不够明显时, 概率分布常写为:

X	x_1	x_2	x_3	$\ldots$
$\mathbb{P}$	p_1	p_2	p_3	$\ldots$

分布列 $\{p_k\}$ 满足以下性质:

1. $p_k \geq 0$;
2. $\displaystyle\sum_{k=1}^{\infty} p_k = 1$.

♣

两点分布、二项分布

定义 2.11 (两点分布（Bernoulli 分布）)

如果 X 只取值 0 或 1, 概率分布是 $\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$, 则称 X 服从两点分布 (**Bernoulli** 分布), 记作 $X \sim B(1, p)$ 或 $X \sim B(p)$。其分布列为:

X	0	1
$\mathbb{P}$	$1-p$	p

♣

设试验 S 成功的概率为 p, 将试验 S 重复 n 次, 用 X 表示成功的次数，求 $\mathbb{P}(X = k)$.

定义 2.12 (二项分布（Binomial 分布）)

如果随机变量 X 的概率分布如下:

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

则称 X 服从二项分布, 其中 $p \in (0, 1)$, 记作 $X \sim B(n, p)$。♣

如果随机变量 $X_1, \ldots, X_n$ 相互独立, 都服从两点分布 $B(1, p)$, 则 $S = X_1 + \cdots + X_n \sim B(n, p)$。

如果随机变量 X, Y 相互独立, 且 $X \sim B(n, p), Y \sim B(m, p)$, 则 $X + Y \sim B(m + n, p)$。

定理

二项分布的最大可能值 k_0 存在, 即满足

$$b(k_0,n,p)=\max_{0\le k\le n} b(k,n,p)$$

的整数 k_0 存在, 且

$$k_0=\begin{cases}(n+1)p,(n+1)p-1, & \text{如果}(n+1)p\text{ 为整数},\\ \lfloor (n+1)p\rfloor, & \text{如果}(n+1)p\text{ 不是整数}.\end{cases}$$

一般称 $b(k_0,n,p)$ 为二项分布的中心项

几何分布

甲向一个目标射击，直到击中为止。用 X 表示射击停止时的射击次数。如果甲每次击中的概率是 $p\in(0,1)$, 求 $\mathbb{P}(X=k)$。

定义 2.13 (几何分布（Geometric 分布）)

如果随机变量 X 的概率分布如下:

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

则称 X 服从参数为 p 的几何分布。记作 $X\sim G(p)$。

定理

取正整数值的随机变量 $X\sim G(p)$ 的充要条件是 X 有无记忆性, 即对每个 $k\ge 1$,

$$\mathbb{P}(X=k+1|X>k)=\mathbb{P}(X=1).$$

甲向一个目标射击，直到击中 r 次为止. 用 X 表示射击停止时的射击次数。如果甲每次击中的概率是 $p\in(0,1)$, 求 $\mathbb{P}(X=k)$。

定义 2.14 (Pascal 分布)

如果随机变量 X 的概率分布如下:

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

则称 X 服从帕斯卡分布。

当 $r=1$ 时, 帕斯卡分布就是几何分布。

负二项分布

令 $Y=X-r$ 则 Y 是射击停止时，射击失败的次数.

定义 2.15 (负二项分布)

如果随机变量 Y 的概率分布如下:

$$\mathbb{P}(Y=k)=\binom{k+r-1}{r-1}(1-p)^kp^r,\quad k=0,1,\dots \tag{1}$$

则称 Y 服从负二项分布, 记作 $Y\sim NB(r,p)$。

超几何分布

在包含 N 个元素的总体中，M 个是红的，NM 个是黑的. 任意选取 n 个元素组成一组. 试求所取出的这一组中，恰有 k 个红元素的概率。

定义 2.16 (超几何分布)

如果随机变量 X 的概率分布如下:

$$\mathbb{P}(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,\ldots,\min\{n,M\},$$

则称 X 服从超几何分布 (Hypergeometric distribution), 记作 $X\sim H(n,M,N)$。

♣

Poisson 分布

定义 2.17 (Poisson 分布)

如果随机变量 X 的概率分布如下:

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

则称 X 服从参数为 λ 的 **Poisson** 分布, 记作 $X\sim\mathcal{P}(\lambda)$, 其中 λ 是正常数。

♣

2.5 常见连续型随机变量

定义 2.18 (连续型随机变量、概率密度)

设随机变量 X, 如果存在非负函数 $f(x)$ 使得对任意的 $a<b$,

$$\mathbb{P}(a<X\le b)=\int_a^b f(x)dx,$$

则称 X 是连续型随机变量 (continuous r.v.), 称 $f(x)$ 是 X 的概率密度函数, 简称概率密度或者密度 (probability density function, PDF)。

♣

定理 (概率密度函数的性质)

设 $f(x)$ 是 X 的概率密度, 则

1. $\displaystyle\int_{-\infty}^{\infty} f(x)dx=1$。
2. 对任意的 Borel 集 A, 有 $\mathbb{P}(X\in A)=\displaystyle\int_A f(x)dx$。
3. $\mathbb{P}(X=a)=0$。(推论: $\mathbb{P}(a<X\le b)=\mathbb{P}(a\le X\le b)$。)

♡

均匀分布

定义 2.19 (均匀分布 U(a,b))

对 $a<b$, 如果 X 的概率密度是

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

称 X 服从区间 (a,b) 上的均匀分布, 记作 $X\sim U(a,b)$。

♣

密度函数 $f(x)$ 还可以写成示性函数的形式：

$$f(x)=\frac{1}{b-a}I_{(a,b)}(x) \text{ 或 } f(x)=\frac{1}{b-a}I_{x\in(a,b)}$$

定义 (Borel 集上的均匀分布 U(a,b))

对任意的 Borel 集 A, 如果 A 的测度 $m(A)=\int_A 1\,\mathrm{d}x<\infty$, 可以类似定义 A 上的均匀分布：如果 X 的密度是

$$f(x)=\begin{cases}\dfrac{1}{m(A)}, & x\in A,\\ 0, & x\notin A,\end{cases}$$

称 X 服从 Borel 集 A 上的均匀分布, 记作 $X\sim U(A)$。对任意的 Borel 集 B,

$$\mathbb{P}(B)=\frac{m(A\cap B)}{m(A)}.$$

♣

指数分布

定义 2.20 (指数分布 $\mathcal{E}(\lambda)$)

对正常数 λ, 如果 X 的概率密度是

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

称 X 服从参数为 λ 的指数分布, 记作 $X\sim\mathcal{E}(\lambda)$。

♣

密度函数 $f(x)$ 还可以写成示性函数的形式:

$$f(x)=\lambda e^{-\lambda x}I_{(0,\infty)} \quad \text{或} \quad f(x)=\lambda e^{-\lambda x},\ x\geq 0.$$

当随机变量 X 使得 $\mathbb{P}(X<0)=0$，则称 X 是非负随机变量

定理 2.5 (指数分布的无记忆性)

设 X 是连续型非负随机变量, 则 X 服从指数分布的充要条件是 X 有无记忆性, 即, 对任意的 $s,t\geq 0$, 有

$$\mathbb{P}(X>s+t|X>s)=\mathbb{P}(X>t). \tag{2}$$

♡

正态分布

定义 2.21 (正态分布 $\mathcal{N}(\mu,\sigma^2)$)

如果 X 的密度是

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\quad x\in\mathbb{R}, \tag{2}$$

称 X 服从参数为 (μ,σ^2) 的正态分布 (Normal distribution), 记作 $X\sim\mathcal{N}(\mu,\sigma^2)$。$N$ 是 Normal 的首字母。特别地, 当 $X\sim\mathcal{N}(0,1)$ 时, 称 X 服从标准正态分布。标准正态分布的密度函数:

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),\quad x\in\mathbb{R}.$$

♣

一些其他分布

定义 2.22 (Beta 分布 $Beta(\alpha,\beta)$)

设 α,β 是正常数, 如果 X 的密度是

$$f(x)=\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad 0<x<1.$$

其中 $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, 则称 X 服从参数为 (α,β) 的 **Beta** 分布, 记作 $X\sim Beta(\alpha,\beta)$。 ♣

定义 2.23 (Gamma 分布 $\Gamma(\alpha,\lambda)$)

设 α,λ 是正常数, 如果 X 的密度是

$$f(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},\quad x\geq 0,$$

称 X 服从参数为 (α,λ) 的 **Gamma** 分布, 记作 $X\sim\Gamma(\alpha,\lambda)$, 其中

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

称为 **Gamma** 函数。 ♣

性质 **Gamma** 函数的性质

(a) $\Gamma(n)=(n-1)!$;

(b) $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$;

(c) $\Gamma\left(\dfrac{1}{2}\right)=\sqrt{\pi}$.

α 被称为**形状参数** (shape parameter); λ 被称为**尺度参数** (rate parameter)。当 $\alpha=1$ 时，为指数分布。

定义 2.24 (Weibull 分布)

如果 X 的密度是

$$f(x)=\lambda\alpha x^{\alpha-1}\exp(-\lambda x^{\alpha}),\quad x\geq 0,$$

称 X 服从参数为 (α,λ) 的 **Weibull** 分布。 ♣

(a) 当 $\alpha=1$ 时, Weibull 分布为指数分布;

(b) 当 $\alpha=2$ 时, Weibull 分布为 **Rayleigh 分布**, 其密度为:

$$f(x;\sigma)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right),\quad x\geq 0.$$

定义 2.25 (卡方分布 (Chi-squared distribution))

设 $\nu>0$, 如果 X 的密度是

$$f(x)=\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2},\quad x\geq 0,$$

称 X 服从参数为 ν 的 χ^2_ν 分布, 记作 $X\sim\chi^2_\nu$。 ♣

设有独立同分布 (i.i.d) 的变量 $Z_j\sim\mathcal{N}(0,1), j=1,\dots,n$。令 $X=Z_1^2+\cdots+Z_n^2$, 则 $X\sim\chi^2_n$。可以验证 $\chi^2_1=\Gamma\left(\dfrac{1}{2},\dfrac{1}{2}\right)$。于是由 Gamma 分布的性质, 有 $\chi^2_n=\Gamma\left(\dfrac{n}{2},\dfrac{1}{2}\right)$。

定义 2.26 (Student's t 分布)

设 $\nu > 0$, 如果 X 的密度是

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R},$$

称 X 服从参数为 ν 的 **Student's t** 分布, 简称 **t** 分布, 记作 $X \sim t_\nu$。 ♣

设有 $Z \sim \mathcal{N}(0,1), X \sim \chi_n^2$, 且 X 与 Z 独立。令 $T = \frac{Z}{\sqrt{X/n}}$, 则有 $T \sim t_n$。

对称性: 若 $T \sim t_\nu$, 则 $-T \sim t_\nu$。

当 v 很大时，t_v 分布看起来和标准正态分布很像。当 $v = 1$ 时，性质特殊，亦称为 Cauchy 分布。

定义 2.27 (Cauchy 分布)

如果 X 的密度是

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

称 X 服从 **Cauchy** 分布。亦可记作 $X \sim t_1$。 ♣

设有独立同分布 (i.i.d) 的变量 $Z_j \sim t_1, j = 1, \ldots, n$。令 $X = \frac{Z_1 + \cdots + Z_n}{n}$, 则依然有 $X \sim t_1$。

定义 2.28 (F 分布 (Fisher Snedecor distribution))

设 m, n 是正整数, 如果 X 的密度是

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(mx+n)^{\frac{m+n}{2}}}, \quad x \geq 0,$$

称 X 服从参数为 (m, n) 的 **F** 分布, 记作 $X \sim F(m, n)$。 ♣

设有独立同分布 (i.i.d) 的变量 $X_j \sim \mathcal{N}(0,1), j = 1, \ldots, m$ 和 $Y_j \sim \mathcal{N}(0,1), j = 1, \ldots, n$。令 $F = \frac{\sum_i X_i^2/m}{\sum_j Y_j^2/n}$, 则 $F \sim F(m, n)$。

定义 2.29 (分布函数 (distribution function))

对随机变量 (random variable, r.v.) X, 称 x 的函数

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

为 X 的概率分布函数或累积分布函数 (cumulative distribution function, CDF), 简称为分布函数。 ♣

令 $H(x)$ 是 Heaviside 函数: $H(x) = I_{[0,\infty)}(x)$。则离散型随机变量的分布函数可以表示为

$$F(x) = \sum_{k=1}^{\infty} p_k H(x - x_k). \tag{1}$$

如果分布函数 $F(x)$ 可以表示为 (1) 这种形式, 称 $F(x)$ 是**离散的**。

如果 X 是连续型随机变量, 有概率密度 $f(x)$, 则

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)dt, \quad x \in \mathbb{R}$$

是连续函数, 并且在 $f(x)$ 的连续点 x 有 $f(x) = F'(x)$。称 $F(x)$ 是 $f(x)$ 的分布函数。

定理 2.6 (分布函数 F(x) 的性质)

1. $F(x)$ 单调非降;
2. $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0, F(\infty)=\lim\limits_{x\to\infty}F(x)=1.$
3. $F(x)$ 是右连续的.

♡

定理 2.7

如果对 $\forall x\in\mathbb{R}$ 都有 $\mathbb{P}(X=x)=0$, 则 $F(x)=\mathbb{P}(X\le x)$ 是连续的。

♡

定理 2.8

设 X 的分布函数 $F(x)$ 连续, 数集 A 中任何两点之间的距离大于某个正数 δ。如果在 A^c 上, 导数 $F'(x)$ 存在且连续, 则

$$f(x)=\begin{cases}F'(x), & \text{当} x\notin A,\\ 0, & \text{当} x\in A.\end{cases}$$

是 X 的密度函数。

♡

注 在该定理中, F 连续的条件是至关重要的。尽管 F 连续不能保证 X 是连续型随机变量, 但是 F 不连续能保证 X 不是连续型随机变量。当 F 是二项分布的随机变量的分布函数时, 除去有限个点外, $F'(x)=0$, 所以 X 的密度不存在。

2.6 随机变量函数的分布

一般地，有如下定理:

定理 2.9 (随机变量函数的分布)

设 X 的概率密度为 $f(x), D\subset\mathbb{R}, Y=g(X), \mathbb{P}(Y\in D)=1$, 如果

1. 对 $y\in D, \{Y=y\}=\bigcup\limits_{i=1}^{n}\{X=h_i(y)\}$;
2. 每个 $h_i(y)$ 是 D 到其值域 D_i 的可逆映射, 在 D 内有连续的导数;
3. $D_1, D_2, \ldots, D_n$ 互不相交,

则 Y 的密度函数为

$$f_Y(y)=\begin{cases}\sum\limits_{i=1}^{n} f(h_i(y))|h_i'(y)|, & y\in D\\ 0, & y\in D^c\end{cases}.$$

♡

例 2.4 设 r 是正常数, $X \sim U(0,2\pi)$, 求 $Y = r\cos X$ 的概率密度。

解 方法一:

设 $D=(-r,r)$, 则 $\mathbb{P}(Y\in D)=1$。对于 $y\in D$, 有

$$\begin{aligned}\{Y=y\} &= \{\cos X = y/r\}\\ &= \{\cos X = y/r, X\in(0,\pi)\}\cup\{\cos X = y/r, X\in(\pi,2\pi)\}\\ &= \{X=\arccos(y/r)\}\cup\{X=2\pi-\arccos(y/r)\}\end{aligned}$$

令 $h_1(y)=\arccos(y/r): D\to(0,\pi)$ 和 $h_2(y)=2\pi-\arccos(y/r): D\to(\pi,2\pi)$ 都是可逆的、连续可微的函数。利用 $f_X(x)=1/(2\pi), x\in(0,2\pi)$, 可得 Y 的密度函数

$$\begin{aligned}f_Y(y) &= \frac{1}{2\pi}\left|\frac{d}{dy}\left[\arccos\left(\frac{y}{r}\right)\right]\right| + \frac{1}{2\pi}\left|\frac{d}{dy}\left[2\pi-\arccos\left(\frac{y}{r}\right)\right]\right|\\ &= \frac{1}{2\pi}\left|-\frac{1}{\sqrt{r^2-y^2}}\right| + \frac{1}{2\pi}\left|\frac{1}{\sqrt{r^2-y^2}}\right|\\ &= \frac{1}{\pi\sqrt{r^2-y^2}},\quad y\in(-r,r).\end{aligned}$$

方法二:

对 $y\in(-r,r)$, 先求分布函数 $F_Y(y)=\mathbb{P}(Y\le y)$。

$$\begin{aligned}\mathbb{P}(Y\le y) &= \mathbb{P}(\cos X\le y/r)\\ &= \int_{\{x:\cos x\le y/r\}}\frac{1}{2\pi}dx\\ &= \frac{1}{2\pi}\{2\pi-2\arccos(y/r)\}\quad (\text{在}(0,2\pi)\ \text{区间})\\ &= 1-\frac{1}{\pi}\arccos(y/r)\end{aligned}$$

根据 $[\arccos(x)]' = -\dfrac{1}{\sqrt{1-x^2}}$, 对分布函数求导可得密度函数 $f_Y(y)$:

$$\begin{aligned}f_Y(y) = F_Y'(y) &= \frac{d}{dy}\left(1-\frac{1}{\pi}\arccos\left(\frac{y}{r}\right)\right)\\ &= -\frac{1}{\pi}\left(-\frac{1}{\sqrt{1-(y/r)^2}}\right)\cdot\frac{1}{r}\\ &= \frac{1}{\pi\sqrt{r^2-y^2}},\quad y\in(-r,r).\end{aligned}$$

第 3 节 随机向量和概率分布

定义 3.1 (n 维随机向量)

如果 $X_1, \ldots, X_n$ 都是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量, 称 $\mathbf{X} = (X_1, \ldots, X_n)$ 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的 **n** 维随机向量, 简称随机向量。♣

定义 3.2 (联合概率分布函数)

设 $\mathbf{X} = (X_1, \ldots, X_n)$ 是随机向量, 称 $\mathbb{R}^n$ 上的 n 元函数

$$F(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

为 $\mathbf{X} = (X_1, \ldots, X_n)$ 的联合概率分布函数, 简称分布函数或者联合分布。♣

联合分布函数 $F(x_1, x_2, \ldots, x_n)$ 是 $\mathbb{R}^n$ 上的右连续函数, 关于每个自变量 x_j 单调非降。
$F(\mathbf{x}) = \mathbb{P}(\mathbf{X} \le \mathbf{x})$, 其中 $\mathbf{X} \le \mathbf{x} \Leftrightarrow X_j \le x_j, j = 1, \ldots, n$。

3.1 离散型随机向量

定义 3.3 (离散型随机向量)

如果 $X_1, X_2, \ldots, X_n$ 都是离散型随机变量, 则称 $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ 为离散型随机向量。如果 $\mathbf{X}$ 所有的不同取值是

$$\mathbf{x}(j_1, j_2, \ldots, j_n) = (x_1(j_1), x_2(j_2), \ldots, x_n(j_n)), \quad j_1, j_2, \ldots, j_n \ge 1$$

则称

$$p_{j_1, j_2, \ldots, j_n} = \mathbb{P}(\mathbf{X} = \mathbf{x}(j_1, j_2, \ldots, j_n)), \quad j_1, j_2, \ldots, j_n \ge 1$$

是 $\mathbf{X}$ 的联合分布列。♣

概率分布的性质

1. $p_{j_1, j_2, \ldots, j_n} \ge 0$;
2. $\displaystyle\sum_{j_1, j_2, \ldots, j_n} p_{j_1, j_2, \ldots, j_n} = 1$.

定义 3.4 (边缘分布列)

设离散型随机向量 $\mathbf{X}$ 的联合分布列为 $p_{j_1, j_2, \ldots, j_n}$, 则称

$$\begin{aligned} &\mathbb{P}(X_1 = x_1(j_1), X_2 = x_2(j_2), \ldots, X_k = x_k(j_k)) \\ = &\sum_{j_{k+1}, j_{k+2}, \ldots, j_n} p_{j_1, j_2, \ldots, j_n}, \quad j_1, j_2, \ldots, j_k \ge 1 \end{aligned}$$

为 $\mathbf{X}$ 的边缘分布列。♣

例 3.1 设 $F(x,y)$ 是 (X,Y) 的联合分布, 则对矩形 $D=\{a<x\le b, c<y\le d\}$, 有

$$\begin{aligned}\mathbb{P}((X,Y)\in D) &= \mathbb{P}(a<X\le b, c<Y\le d)\\ &= F(b,d)-F(b,c)-F(a,d)+F(a,c)\end{aligned}$$

因此

$$F(b,d)-F(b,c)-F(a,d)+F(a,c)\ge 0,\quad a<b, c<d.$$

进一步, X,Y 分别有概率分布

$$F_X(x)=\mathbb{P}(X\le x, Y\le\infty)=F(x,\infty)$$
$$F_Y(y)=\mathbb{P}(X\le\infty, Y\le y)=F(\infty,y)$$

更一般地, 我们定义

定义 3.5 (边缘分布、边际分布 (marginal distribution))

对 $1\le k<n, S_k=\{i_1,i_2,\ldots,i_k\}$ 为指标集, 称 $\mathbf{X}_{S_k}:=(X_{i_1},X_{i_2},\ldots,X_{i_k})$ 的联合分布

$$\begin{aligned}&F_k(x_{i_1},x_{i_2},\ldots,x_{i_k})\\ &=\mathbb{P}(X_{i_1}\le x_{i_1},X_{i_2}\le x_{i_2},\ldots,X_{i_k}\le x_{i_k})\\ &=\mathbb{P}(X_j\le x_j, j\in S_k, X_t\le\infty, t\in S_k^c)\end{aligned}$$

为 $F(x_1,x_2,\ldots,x_n)$ 的 **k-**维边缘分布。♣

$F(x_1,x_2,\ldots,x_n)$ 一共有 (2^n-2) 个边缘分布。

离散型随机变量的独立性

定义 3.6 (复习: 随机变量的相互独立性)

设 $X_1,\ldots,X_n$ 是 $(\Omega,\mathcal{F})$ 上的随机变量, 如果对任意的实数 $x_1,\ldots,x_n$ 都有

$$\mathbb{P}(X_1\le x_1,\ldots,X_n\le x_n)=\mathbb{P}(X_1\le x_1)\cdots\mathbb{P}(X_n\le x_n),$$

称随机变量 $X_1,\ldots,X_n$ 相互独立。♣

等价表达: 设 $(X_1,X_2,\ldots,X_n)$ 有联合分布 $F(x_1,x_2,\ldots,x_n)$, X_i 有边缘分布 $F_i(x_i)$。则 $X_1,\ldots,X_n$ 相互独立当且仅当

$$F(x_1,x_2,\ldots,x_n)=F_1(x_1)F_2(x_2)\ldots F_n(x_n).$$

定理 3.1 (离散型随机向量的相互独立性)

设离散型随机向量 $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ 有概率分布

$$p_{j_1,j_2,\ldots,j_n}=\mathbb{P}(\mathbf{X}=\mathbf{x}(j_1,j_2,\ldots,j_n)),\quad j_1,j_2,\ldots,j_n\ge 1$$

则 $X_1,\ldots,X_n$ 相互独立的充分必要条件是: 对任意的 $(x_1(j_1),x_2(j_2),\ldots,x_n(j_n))$, 有

$$\mathbb{P}(X_1=x_1(j_1),\ldots,X_n=x_n(j_n))=\prod_{i=1}^n\mathbb{P}(X_i=x_i(j_i)).$$

♡

推论 3.1

设离散型随机向量 (X,Y) 的所有不同取值是 $(x_i,y_j), i,j\ge 1$。则 X,Y 相互独立的充分必要条件是: 对任意的 (x_i,y_j), 有

$$\mathbb{P}(X=x_i,Y=y_j)=\mathbb{P}(X=x_i)\mathbb{P}(Y=y_j).$$

♡

3.2 连续型随机向量

定义 3.7 (连续型随机向量)

设 $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ 是随机向量。如果存在 $\mathbb{R}^n$ 上的非负可积函数 $f(x_1,\ldots,x_n)$ (or $f(\mathbf{x})$), 使得对 $\mathbb{R}^n$ 的任何子立方体 $D=\{(x_1,x_2,\ldots,x_n)|a_i\le x_i\le b_i, 1\le i\le n\}$ 有

$$\mathbb{P}(\mathbf{X}\in D)=\int_D f(x_1,x_2,\ldots,x_n)dx_1dx_2\ldots dx_n:=\int_D f(\mathbf{x})d\mathbf{x}$$

就称 $\mathbf{X}$ 是连续型随机向量, 并称 $f(x_1,x_2,\ldots,x_n)$ (or $f(\mathbf{x})$) 是 $\mathbf{X}$ 的联合概率密度函数, 简称为联合密度或概率密度。

♣

定理 3.2

对任意 n 维 Borel 集 B 有:

$$\mathbb{P}(\mathbf{X}\in B)=\int_B f(\mathbf{x})d\mathbf{x}$$

♡

- 随机向量 $\mathbf{X}$ 的联合概率密度不必唯一。
- 由于 $\mathbb{R}^n$ 是 Borel 集, 有

$$\int_{R^n} f(\mathbf{x})d\mathbf{x}=\mathbb{P}(\mathbf{X}\in R^n)=1$$

定义 3.8 (边缘密度、边际密度)

被积函数

$$f_k(x_1,\ldots,x_k)=\int_{\mathbb{R}^{n-k}} f(x_1,\ldots,x_n)dx_{k+1}\ldots dx_n$$

是 $(X_1,\ldots,X_k)$ 的联合密度, 被称为 $\mathbf{X}$ 的边缘概率密度函数, 简称为边缘密度。

♣

例 3.2 设 $f(\mathbf{x})$ 是 $\mathbf{X}$ 的概率密度。$(X_{j_1},X_{j_2},\ldots,X_{j_k})$ 的边缘密度为:

$$g_k(x_{j_1},x_{j_2},\ldots,x_{j_k})=\int_{\mathbb{R}^{n-k}} f(x_1,\ldots,x_n)dx_{j_{k+1}}\ldots dx_{j_n}$$

例 3.3 设 (X,Y) 在单位圆 $D=\{(x,y)|x^2+y^2\le 1\}$ 内均匀分布, 求 X,Y 的概率密度。

解 (X,Y) 的联合密度为 $f(x,y)=\dfrac{1}{\pi}I_D$, 其中 I_D 是单位圆 D 的示性函数。X 只在 $[-1,1]$ 中取值, 则

$$\begin{aligned}f_X(x)&=\int_{-\infty}^{\infty}f(x,y)dy=\frac{1}{\pi}\int_{-\infty}^{\infty}I_{x^2+y^2\le 1}dy\\&=\frac{1}{\pi}\int_{-\infty}^{\infty}I_{|y|\le\sqrt{1-x^2}}dy=\frac{2}{\pi}\sqrt{1-x^2},\quad |x|\le 1.\end{aligned}$$

同理可得 Y 的密度:

$$f_Y(y)=\frac{2}{\pi}\sqrt{1-y^2},\quad |y|\le 1.$$

连续型随机向量的独立性

定理 3.3 (连续型随机向量的独立性)

设对每个 $i(1\le i\le n)$, 随机变量 X_i 有概率密度 $f_i(x_i)$, 则 $X_1,X_2,\ldots,X_n$ 相互独立的充分必要条件是随机向量 $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ 有联合密度

$$f_1(x_1)f_2(x_2)\ldots f_n(x_n),\quad (x_1,x_2,\ldots,x_n)\in\mathbb{R}^n.$$

♡

经典连续型随机向量

定义 3.9 (二元正态分布)

设 μ_1, μ_2 是常数, σ_1, σ_2 是正常数, $\rho \in (-1,1)$ 中的常数。如果随机向量 (X,Y) 有概率密度

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

则称 (X,Y) 服从二元正态分布, 记作 $(X,Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$。

对比一元正态分布:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

简写形式:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

其中 $\mathbf{x} = (x,y)'$, $\mu = (\mu_1, \mu_2)'$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, $|\Sigma|$ 是 Σ 的行列式。

定理

设 $(X,Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. 则 X, Y 独立的充分必要条件是 $\rho = 0$。

联合分布与联合密度

设 $f(x,y)$ 是 (X,Y) 的联合密度, 则有

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

当 $f(x,y)$ 连续时, 两者有关系

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

定理

设 (X,Y) 有连续的分布函数 $F(x,y)$, 定义

$$f(x,y) = \begin{cases} \dfrac{\partial^2 F(x,y)}{\partial x\partial y}, & \text{当该混合偏导数存在} \\ 0, & \text{其它.} \end{cases}$$

如果

$$\int_{\mathbb{R}^2} f(x,y)dxdy = 1$$

则 $f(x,y)$ 是 (X,Y) 的联合密度。

3.3 条件分布和条件密度

3.3.1 离散型条件分布

设 (X,Y) 是离散型随机向量, 有概率分布 (joint PMF)

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j) > 0, \quad i,j = 1,2,\ldots,$$

则 X, Y 分别有边缘分布 (marginal PMF)

$$p_i = \mathbb{P}(X = x_i), \quad q_j = \mathbb{P}(Y = y_j), \quad i, j = 1, 2, \dots,$$

定义 3.10 (条件概率分布)

对每个固定的 j, 称

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{ij}}{q_j}, \quad i = 1, 2, \dots,$$

为给定条件 $Y = y_j$ 下, X 的条件分布列 (conditional PMF)。

♣

定理 3.4

X, Y 独立的充分必要条件是对任何 $i, j \geq 1$,

$$\mathbb{P}(X = x_i | Y = y_j) = \mathbb{P}(X = x_i).$$

♡

- 等价于, $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j), \quad i, j \geq 1$。

3.3.2 连续型条件分布

定义 3.11 (连续型条件分布与条件密度)

设随机向量 (X, Y) 有联合密度 $f(x, y)$, Y 有边缘密度 $f_Y(y)$。
若在 y (确定的 y) 处 $f_Y(y) > 0$, 则称

$$\mathbb{P}(X \leq x | Y = y) = \frac{\displaystyle\int_{-\infty}^{x} f(s, y) ds}{f_Y(y)}, \quad x \in \mathbb{R}$$

为给定条件 $Y = y$ 下, X 的条件分布函数, 简称条件分布, 记作 $F_{X|Y}(x|y)$。
称

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f(x, y)}{\displaystyle\int_{-\infty}^{\infty} f(x, y) dx}, \quad x \in \mathbb{R}$$

为给定条件 $Y = y$ 下, X 的条件概率密度, 简称条件密度。

♣

由条件密度定义可得:

$$f(x, y) = f_Y(y) f_{X|Y}(x|y).$$

定理 3.5 (条件密度与条件分布的关系)

对使得 $f_Y(y) > 0$ 的 y,

1. $F_{X|Y}(x|y) = \mathbb{P}(X \leq x | Y = y) = \displaystyle\int_{-\infty}^{x} f_{X|Y}(s|y) ds, \quad x \in \mathbb{R}$;
2. 如果 $F_{X|Y}(x|y)$ 关于 x 连续, 且除去至多可列个点外有连续的导数, 则

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{\partial F_{X|Y}(x|y)}{\partial x}, & \text{当偏导数存在} \\ 0, & \text{其它} \end{cases}$$

   是给定条件 $Y = y$ 下, X 的条件密度。

♡

定理 3.6

X, Y 独立的充分必要条件是对 $y \in \{y | f_Y(y) > 0\}$,

$$f_{X|Y}(x|y) = f_X(x), \quad x \in \mathbb{R}.$$

- 等价于, $f(x, y) = f_X(x) f_Y(y), \quad x, y \in \mathbb{R}$。

♡

例 3.4 设计算机使用的环境指标 $Y \sim \Gamma(\alpha, \beta)$, 概率密度是

$$f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad y > 0.$$

已知 $Y = y$ 时, 软件的使用寿命 $X|Y = y \sim \mathcal{E}(y)$。求 X 的分布。

解 X 的条件密度 $f_{X|Y}(x|y) = ye^{-xy}, x > 0$。于是 (X, Y) 的联合密度为

$$f(x, y) = f_{X|Y}(x|y) f_Y(y) = ye^{-xy} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}, \quad x, y > 0.$$

最后对 $x > 0$, 有

$$\begin{aligned} f_X(x) &= \int_0^\infty f(x, y) dy = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^\alpha e^{-y(x+\beta)} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{1}{(x+\beta)^{\alpha+1}} \int_0^\infty t^\alpha e^{-t} dt \quad (\text{令} t = y(x+\beta)) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+1)}{(x+\beta)^{\alpha+1}} = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}} \quad (\text{利用} \Gamma(\alpha+1) = \alpha\Gamma(\alpha)). \end{aligned}$$

于是 X 有概率密度

$$f_X(x) = \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}}, \quad x > 0.$$

定义（条件分布和条件密度的一般定义）

设 $\mathbf{X} = (X_1, \ldots, X_n)$ 和 $\mathbf{Y} = (Y_1, \ldots, Y_m)$ 是随机向量, $f(\mathbf{x}, \mathbf{y})$ 是 $(\mathbf{X}, \mathbf{Y})$ 的联合密度, 此时 $\mathbf{Y}$ 有边缘密度

$$f_Y(\mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{x}, \mathbf{y}) d\mathbf{x}.$$

若在 $\mathbf{y}$ (确定的 $\mathbf{y}$) 处 $f_Y(\mathbf{y}) > 0$, 则称

$$\mathbb{P}(\mathbf{X} \le \mathbf{x} | \mathbf{Y} = \mathbf{y}) = \frac{\int_{\mathbf{s} \le \mathbf{x}} f(\mathbf{s}, \mathbf{y}) d\mathbf{s}}{f_Y(\mathbf{y})}$$

为给定条件 $\mathbf{Y} = \mathbf{y}$ 下, $\mathbf{X}$ 的条件分布函数, 简称条件分布, 记作 $F_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$。

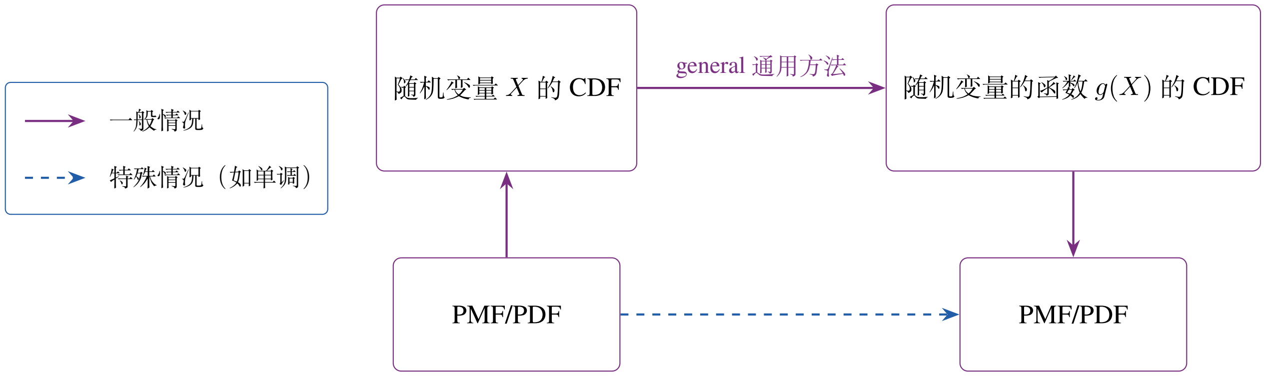
称

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

为给定条件 $\mathbf{Y} = \mathbf{y}$ 下, $\mathbf{X}$ 的条件概率密度, 简称条件密度。

♣

3.4 概率向量函数的分布



一般情形

直接求分布，再根据需要求 PDF 或 PMF

例 3.5 设 X,Y 独立且都服从 $\mathcal{N}(0,1)$. 求 $Z=\sqrt{X^2+Y^2}$ 的分布。

解 (X,Y) 有联合密度

$$f(x,y)=\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right).$$

对 $z\geq 0$, 有

$$\begin{aligned}F_Z(z)&=\mathbb{P}\left(\sqrt{X^2+Y^2}\leq z\right)\\&=\iint_{\{\sqrt{x^2+y^2}\leq z\}}\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)dxdy\\&\xrightarrow{x=r\cos\theta,y=r\sin\theta}\frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^z re^{-r^2/2}dr\\&=\int_0^z re^{-r^2/2}dr.\end{aligned}$$

求导可得 Z 的密度函数

$$f_Z(z)=ze^{-z^2/2},\quad z\geq 0.$$

此分布称为 **Rayleigh** 分布。

向坐标 (0,0) 射击时, 弹落点 (X,Y) 服从二元正态分布。$R=Z=\sqrt{X^2+Y^2}$ 为弹落点到目标的距离, 被称为脱靶量。

特殊情形

(i) Z = X + Y

例 3.6 设 (X,Y) 有联合分布列, 求 $Z=X+Y$ 的分布列。

解

$$P_Z(z)=\mathbb{P}(X+Y=z)=\sum_x\mathbb{P}(X=x,Y=z-x).$$

特别地, 当 X,Y 独立时, $Z=X+Y$ 有分布列

$$P_Z(z)=\sum_x\mathbb{P}(X=x)\mathbb{P}(Y=z-x)=\sum_x P_X(x)P_Y(z-x).$$

此时 Z 的分布列为 X 和 Y 的分布列的卷积。

例 3.7 设 (X,Y) 有联合密度 $f(x,y)$, 则 $Z=X+Y$ 有密度函数

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

特别地, 当 X,Y 独立时, $Z=X+Y$ 有密度函数

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx.$$

(ii) $V=X-Y$ (类似 (i))

例 3.8 设 (X,Y) 有联合密度 $f(x,y)$, 则 $V=X-Y$ 有密度函数

$$f_V(v)=\int_{-\infty}^{\infty} f(x,x-v)dx.$$

特别地, 当 X,Y 独立时, $V=X-Y$ 有密度函数

$$f_V(v)=\int_{-\infty}^{\infty} f_X(x)f_Y(x-v)dx.$$

3.4.1 多个函数的联合密度

定理 3.7

设 (X,Y) 有联合密度 $f(x,y)$, $U=u(X,Y),V=v(X,Y)$ 是 (X,Y) 的函数, D 是平面上的区域使得 $\mathbb{P}((U,V)\in D)=1$. 如果存在 D 上的函数 $x_i=x_i(u,v),y_i=y_i(u,v),i=1,2,\ldots,n$, 使得

1. 对 $(u,v)\in D$, 有 $\{U=u,V=v\}=\bigcup\limits_{i=1}^{n}\{X=x_i,Y=y_i\}$;
2. Δ_i: $x_i=x_i(u,v),y_i=y_i(u,v)$ 是 D 到其值域 D_i 的可逆映射, 有连续的偏导数, 并且雅克比 (Jacobian) 行列式的绝对值 $\left|\frac{\partial(x_i,y_i)}{\partial(u,v)}\right|\neq 0,(u,v)\in D,i=1,\ldots,n$.
3. $D_1,\ldots,D_n$ 互不相交.

则 (U,V) 有联合密度

$$g(u,v)=\begin{cases}\sum\limits_{i=1}^{n} f(x_i(u,v),y_i(u,v))\left|\dfrac{\partial(x_i,y_i)}{\partial(u,v)}\right|, & (u,v)\in D,\\ 0, & (u,v)\in D^c\end{cases}$$

♡

例 3.9 设 X,Y 独立, $X\sim\mathcal{N}(0,1)$, (R,Θ) 由极坐标变换

$$\begin{cases}X=R\cos\Theta\\ Y=R\sin\Theta\end{cases}$$

决定, 求 (R,Θ) 的联合密度.

解 (X,Y) 有联合密度

$$f(x,y)=\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)$$

定义

$$D=\{(r,\theta)|r>0,\theta\in[0,2\pi)\}$$

则 $\mathbb{P}((R,\Theta)\in D)=1$

$$\Delta=\begin{cases}x=r\cos\theta\\ y=r\sin\theta\end{cases}$$

建立了 D 到 $D_1 = \{(x,y)|(x,y)\neq 0\}$ 的可逆变换. 对 $(r,\theta)\in D$, 利用

$$\{R=r,\Theta=\theta\}=\{X=x,Y=y\},\quad \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|=r>0,$$

得 (R,Θ) 的联合密度

$$g(r,\theta)=f(r\cos\theta,r\sin\theta)\cdot r=\frac{1}{2\pi}\cdot r\cdot\exp\left(-\frac{r^2}{2}\right),\quad (r,\theta)\in D.$$

从中可以看出 R 和 Θ 独立, 分别有概率密度

$$g_R(r)=r\exp\left(-\frac{r^2}{2}\right),\quad r\geq 0;\quad g_\Theta(\theta)=\frac{1}{2\pi}I_{[0,2\pi)}.$$

即 R 是 Rayleigh 分布, $\Theta\sim\mathcal{U}[0,2\pi)$.

定理 3.8

设 n 维随机向量 $\mathbf{X}$ 和 $\mathbf{Y}$ 有相同的分布, $g: R^n\to R^m$ 是可测的, 则 $g(\mathbf{X})$ 与 $g(\mathbf{Y})$ 有相同的分布.

例 3.10 设 U,V 独立, $U\sim\mathcal{U}(0,1)$, 定义

$$\begin{cases} X=\sqrt{-2\log U}\cos(2\pi V)\\ Y=\sqrt{-2\log U}\sin(2\pi V)\end{cases}$$

则 X,Y 独立, 且服从 $\mathcal{N}(0,1)$. **【用来产生正态随机变量】**

证明 $R=\sqrt{-2\log U}\sim$ Rayleigh 分布, $\Theta=2\pi V\sim\mathcal{U}[0,2\pi)$. 且 R 与 Θ 独立.

由定理 1.2 易得结论成立.

3.5 次序统计量

定义 3.12

设 $X_1,X_2,\ldots,X_n$ 是概率空间 $(\Omega,\mathcal{F},\mathbb{P})$ 上的随机变量, 对 $\omega\in\Omega$, 将 $X_1(\omega),X_2(\omega),\ldots,X_n(\omega)$ 从小到大排列得到

$$X_{(1)}(\omega)\leq X_{(2)}(\omega)\leq\cdots\leq X_{(n)}(\omega)$$

称 $X_{(1)},X_{(2)},\ldots,X_{(n)}$ 为 $X_1,X_2,\ldots,X_n$ 的次序统计量 **(order statistics)**.

次序统计量的分布密度

以下设随机变量 $X_1,X_2,\ldots,X_n$ 独立同分布, 有公共的分布函数 $F(x)$ 和密度函数 $f(x)$.

性质 $(X_{(1)},X_{(2)},\ldots,X_{(n)})$ 有联合密度

$$g(x_1,\ldots,x_n)=\begin{cases} n!\prod_{i=1}^n f(x_i), & \text{if } x_1<\cdots<x_n\\ 0, & \text{others}\end{cases}$$

性质 $X_{(k)}$ 有密度

$$g_k(x_k)=n\binom{n-1}{k-1}[F(x_k)]^{k-1}[1-F(x_k)]^{n-k}f(x_k)$$

性质 对于 $-\infty\leq a<x_1<x_2<\cdots<x_k<b\leq\infty$, 有

$$\int_{a<x_1<\cdots<x_k<b}f(x_1)\cdots f(x_k)\,dx_1\cdots dx_k=\frac{\big(F(b)-F(a)\big)^k}{k!}$$

性质 对 $k_1 < k_2$, $(X_{(k_1)}, X_{(k_2)})$ 有联合密度

$$g(x_{k_1}, x_{k_2}) = \begin{cases} \dfrac{n!}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!}\left[F(x_{k_1})\right]^{k_1-1}\left[F(x_{k_2})-F(x_{k_1})\right]^{k_2-k_1-1} & \\ \qquad \times \left[1-F(x_{k_2})\right]^{n-k_2} f(x_{k_1})f(x_{k_2}), & x_{k_1} < x_{k_2} \\ 0, & \text{others} \end{cases}$$

例 3.11 某教学楼原来每个教室有 4 只灯泡用于室内照明, 新装修后每个教室有 24 只灯泡用于室内照明。装修后教室管理员总认为灯泡更容易坏了, 试解释其中的原因。

解 设所有灯泡的使用寿命相互独立, 且服从 $\mathcal{E}(\lambda)$. 用 X_i 表示第 i 只灯泡的使用寿命。

装修前等待第一只灯泡烧坏的时间长度为

$$X = \min\{X_1, X_2, X_3, X_4\} = X_{(1)} \sim \mathcal{E}(4\lambda),$$

装修后等待第一只灯泡烧坏的时间长度为

$$Y = \min\{X_1, X_2, \ldots, X_{24}\} = X_{(1)} \sim \mathcal{E}(24\lambda).$$

利用性质 2.2 的结论分别得到 X 和 Y 的密度函数

$$f_X(t) = 4\lambda e^{-4\lambda t}, \qquad f_Y(t) = 24\lambda e^{-24\lambda t}, \quad t > 0.$$

因此 $\mathbb{P}(X > t) = e^{-4\lambda t}$, $\mathbb{P}(Y > t) = e^{-24\lambda t}$。

当 $\lambda = 1/6000$ (小时) 时, 有

$$\mathbb{P}(X > 400) = e^{-4\cdot\frac{1}{6000}\cdot 400} \approx 0.7659, \qquad \mathbb{P}(X > 200) = e^{-4\cdot\frac{1}{6000}\cdot 200} \approx 0.8752,$$
$$\mathbb{P}(Y > 400) = e^{-24\cdot\frac{1}{6000}\cdot 400} \approx 0.2019, \qquad \mathbb{P}(Y > 200) = e^{-24\cdot\frac{1}{6000}\cdot 200} \approx 0.4493.$$

从中可以看到, Y 要比 X 随机地小得多, 因此 “更容易坏” 的直观感受是合理的。

常见分布补充：Beta 分布

定义 3.13 (Beta 分布 $\mathrm{Beta}(\alpha,\beta)$)

设 $\alpha, \beta > 0$。若随机变量 X 的概率密度为

$$f(x) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 < x < 1,$$

其中

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

则称 X 服从参数为 (α,β) 的 Beta 分布，记作 $X \sim \mathrm{Beta}(\alpha,\beta)$。 ♣

性质 与二项分布的共轭性

在贝叶斯统计方法中，常将二项分布的成功概率 $p \in (0,1)$ 的先验分布取为 $\mathrm{Beta}(\alpha,\beta)$。若给定 p 条件下

$$X \mid p \sim \mathrm{Binomial}(n,p),$$

则观测到 X 之后，参数的后验分布为

$$p \mid X \sim \mathrm{Beta}(\alpha + X,\ \beta + n - X).$$

Beta 分布可看作是均匀分布 $\mathscr{U}(0,1)$ 的次序统计量.

3.6 随机变量的 p 分位数

设 X 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上的随机变量，$F(x)$ 是其分布函数。

定义 3.14 (随机变量的 p 分位数)

对 $p \in (0,1)$，定义

$$F^{-1}(p) := \inf\{x \in \mathbb{R} \mid F(x) \geq p\}.$$

称 $F^{-1}(p)$ 为 F 的（或 X 的）p-分位数，通常记为 ξ_p；特别地，$\xi_{1/2}$ 为 F 或 X 的中位数。♣

p-分位数的性质

性质 对所有 $p \in (0,1)$、$t \in \mathbb{R}$，有

1. $F^{-1}(p)$ 关于 p 单调非降；
2. $F^{-1}(F(x)) \leq x$，$\forall x \in \mathbb{R}$；
3. $F\big(F^{-1}(p)\big) \geq p$；
4. $F^{-1}(p) \leq t \iff p \leq F(t)$；
5. $F^{-1}(p)$ 关于 p 左连续；
6. 若 F 连续，则 $F\big(F^{-1}(p)\big) = p$。

定理 3.9 (产生服从分布函数 F(x) 的随机变量)

设 $X \sim \mathcal{U}(0,1)$，$F(x)$ 是连续分布函数。令 $Y = F^{-1}(X)$，则 $Y \sim F$（即 Y 的分布函数为 F）。♡

例 3.12 设易生成服从均匀分布的随机变量，如何生成服从下列分布函数的随机变量：

1. $\mathcal{E}(\lambda)$；2. Cauchy 分布；3. $\mathcal{N}(0,1)$；4. $F_n(x) = \dfrac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{x_i \leq x\}$，其中 $x_1, \ldots, x_n$ 为给定观测值。

解 设 $U, V \sim \mathcal{U}(0,1)$ 且相互独立。

1. 指数分布：取
$$X = -\lambda^{-1} \log(1-U).$$
则 $X \sim \mathcal{E}(\lambda)$。
2. Cauchy 分布：令
$$Z = \frac{X}{Y}, \quad \text{其中}\, X, Y \text{ 独立且}\, X, Y \sim \mathcal{N}(0,1),$$
则 Z 服从标准 Cauchy 分布。
3. 正态分布（Box–Muller 变换）：令
$$\theta = 2\pi U, \qquad R = \sqrt{-2\log V}.$$
则
$$X = R\cos\theta, \qquad Y = R\sin\theta$$
相互独立，且 $X, Y \sim \mathcal{N}(0,1)$。

4. 经验分布 F_n：注意 F_n 为阶梯函数。设

$$x_{(1)} \leq \cdots \leq x_{(n)}$$

为 $x_1 \leq \cdots \leq x_n$ 的次序统计量，则

$$F_n^{-1}(x) = \begin{cases} x_{(1)}, & 0 < x \leq \frac{1}{n}, \\ x_{(2)}, & \frac{1}{n} < x \leq \frac{2}{n}, \\ \vdots & \\ x_{(n)}, & \frac{n-1}{n} < x \leq 1. \end{cases}$$

因此 $F_n^{-1}(U)$ 的取值集合为 $\{x_{(1)}, \ldots, x_{(n)}\}$，并且

$$\mathbb{P}\big(F_n^{-1}(U) = x_{(k)}\big) = \mathbb{P}\Big(\frac{k-1}{n} < U \leq \frac{k}{n}\Big) = \frac{1}{n}, \qquad k = 1, \ldots, n.$$

即只需等概率地从观测值中抽样。

第 4 节 期望、方差、条件期望等数字特征

4.1 随机变量的数学期望

定义 4.1 (离散型随机变量的期望)

设 X 有概率分布 $\mathbb{P}(X=x_j)$，$j=1,2,\ldots$. 如果

$$\sum_{j=1}^{\infty}|x_j|\,\mathbb{P}(X=x_j)<+\infty,$$

称

$$E(X)=\sum_{j=1}^{\infty}x_j\,\mathbb{P}(X=x_j)$$

为 X 的数学期望或均值。

♣

例 4.1 两点分布的期望 设 $X\sim B(1,p)$，即 $\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$。求 $E(X)$。

解

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

例 4.2 设 A 是事件，I_A 为 A 的示性函数。求 $E(I_A)$。

解 I_A 服从两点分布，且

$$E(I_A)=1\cdot\mathbb{P}(A)+0\cdot(1-\mathbb{P}(A))=\mathbb{P}(A).$$

例 4.3 二项分布的期望 设 $X\sim B(n,p)$，即 $p_j=\mathbb{P}(X=j)=\binom{n}{j}p^j(1-p)^{n-j}$，$0\le j\le n$。求 $E(X)$。

解 可将 X 表示为 n 个独立的两点分布之和：$X=\sum_{i=1}^n X_i$，其中 $X_i\sim B(1,p)$ 相互独立。于是

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

例 4.4 设 $X\sim\text{Poisson}(\lambda)$，即

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

求 $E(X)$。

解 由级数移位可得

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

例 4.5 设 $X\sim G(p)$（几何分布），即

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

求 $E(X)$。

解 利用等比级数求和公式 $\sum_{k=1}^{\infty}kr^{k-1}=\dfrac{1}{(1-r)^2}$（$|r|<1$），取 $r=1-p\in(0,1)$，得

$$E(X)=p\sum_{k=1}^{\infty}k(1-p)^{k-1}=p\cdot\frac{1}{(1-(1-p))^2}=\frac{1}{p}.$$

定义 4.2 (连续型随机变量的期望)
设 X 的概率密度为 $f(x)$。如果

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

则定义

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

为 X 的数学期望或均值。♣

例 4.6 设 $X \sim \mathcal{U}(a,b)$，即

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x).$$

求 $E(X)$。

解

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

例 4.7 设 $X \sim \mathcal{E}(\lambda)$（指数分布），其密度

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

求 $E(X)$。

解 分部积分或由标准积分表可得

$$E(X) = \int_0^{\infty} x\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

例 4.8 设 $X \sim \Gamma(\alpha, \beta)$，其密度

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

求 $E(X)$。

解 由代换 $t = \beta x$ 且利用 $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$，有

$$E(X) = \int_0^{\infty} x f(x)\, dx = \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} dx = \frac{1}{\beta\,\Gamma(\alpha)} \int_0^{\infty} t^\alpha e^{-t} dt = \frac{\Gamma(\alpha+1)}{\beta\,\Gamma(\alpha)} = \frac{\alpha}{\beta}.$$

例 4.9 设 $X \sim \mathcal{N}(\mu, \sigma^2)$，密度

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

求 $E(X)$。

解 由对称性与换元

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-\infty}^{\infty} \mu f(x)\, dx + \int_{-\infty}^{\infty} (x-\mu) f(x)\, dx = \mu + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

令 $t = x - \mu$，则第二项为奇函数在对称区间的积分，故为 0，从而 $E(X) = \mu$。

定义 4.3 (期望的一般定义（Stieltjes 形式）)

若随机变量 X 的分布函数为 $F(x)$，则 X 的数学期望（若存在）可写为

$$E(X) = \int_\Omega X\, d\mathbb{P} = \int_{\mathbb{R}} x\, dF(x).$$

为说明 Stieltjes 积分的离散化形式，设 $G(x)$ 是任意单调非降的右连续函数，且在 $(a,b]$ 上至多有可列个跳跃点 x_j（$j=1,2,\ldots$），并满足

$$\sum_{j:x_j\in(a,b]} g(x_j)\,\big[G(x_j)-G(x_j-0)\big] < \infty.$$

则可形式地定义

$$\int_a^b g(x)\, dG(x) = \sum_{j:x_j\in(a,b]} g(x_j)\,\big[G(x_j)-G(x_j-0)\big].$$

对任意随机变量 X，若存在非负函数 $f_0(x)$ 使得

$$F(x) = F_1(x) + F_2(x), \quad F_1(x) = \int_{-\infty}^x f_0(s)\, ds,$$

且 F_2 仅在可列点 x_j 处有跳跃，满足 $\mathbb{P}(X=x_j) = F(x_j) - F(x_j-0)$。若

$$\int_{-\infty}^{\infty} |x| f_0(x)\, dx + \sum_{j=1}^{\infty} |x_j|\, \mathbb{P}(X=x_j) < \infty,$$

则 $E(X)$ 存在，并且

$$E(X) = \int_{-\infty}^{\infty} x f_0(x)\, dx + \sum_{j=1}^{\infty} x_j\, \mathbb{P}(X=x_j) = \int_{-\infty}^{\infty} x\, dF_1(x) + \int_{-\infty}^{\infty} x\, dF_2(x) = \int_{-\infty}^{\infty} x\, dF(x).$$

（若期望存在），同分布的随机变量具有相同的期望。♣

4.2 期望的性质

随机变（向）量函数的数学期望

定理 4.1

设 $\mathbf{X} = (X_1,\ldots,X_n)$ 是随机向量，$\mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$。若 $\mathbf{X}$ 具有联合分布函数 $F(\mathbf{x})$，实函数 $g(\mathbf{x})$ 满足

$$\int_{\mathbb{R}^n} |g(\mathbf{x})|\, dF(\mathbf{x}) < \infty,$$

则 $Y = g(\mathbf{X})$ 有数学期望，且

$$E(Y) =: E\big(g(\mathbf{X})\big) = \int_{\mathbb{R}^n} g(\mathbf{x})\, dF(\mathbf{x}).$$

♡

例 4.10 设 $X \sim \mathcal{U}(0,\pi/2)$，计算 $E(\cos X)$。

解 由密度 $f(x) = \dfrac{2}{\pi} I_{(0,\pi/2)}(x)$，得

$$E(\cos X) = \int_0^{\pi/2} \frac{2}{\pi}\cos(x)\, dx = \frac{2}{\pi}.$$

例 4.11 设 $X \sim \mathcal{N}(0,1)$，α 为常数，计算 $E(|X|^\alpha)$。

解 令标准正态密度 $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$。当 $\alpha > -1$ 时，偶函数对称性与换元 $x = \sqrt{2t}$ 给出

$$E(|X|^\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|^\alpha e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^\alpha e^{-x^2/2} dx = \sqrt{\frac{2^\alpha}{\pi}}\,\Gamma\left(\frac{\alpha+1}{2}\right).$$

特别地：

- $E(X^2) = 1$；
- $E|X| = \sqrt{\frac{2}{\pi}}$；
- 若 $\alpha \le -1$，则 $E|X|^\alpha = \infty$。

注：$\Gamma(1/2) = \sqrt{\pi}$。

期望的基本性质

定理 4.2

设 $E|X| < \infty$，$E|Y| < \infty$，且 $a, b, c \in \mathbb{R}$。则：

1. $EC = C$；
2. $|EX| \le E|X|$；
3. 线性性：$E(aX + bY) = a\,EX + b\,EY$；
4. 若 $X \le Y$ 几乎处处，则 $EX \le EY$；
5. 若 X 与 Y 独立，则 $E(XY) = (EX)(EY)$。

♡

最佳 (常数) 预测

定理 4.3

设 $EX = \mu$ 且 $E(X^2) < \infty$，则对所有 $c \in \mathbb{R}$ 有

$$E\big[(X-\mu)^2\big] \le E\big[(X-c)^2\big].$$

♡

由此得到等价的优化表述（最小二乘角度）：

$$EX = \arg\min_{c\in\mathbb{R}} E\big[(X-c)^2\big].$$

高阶矩

一类“特殊的随机变量函数”的期望：

定义 4.4 (Moment)

设 X 是随机变量，m 是正整数。若 $E\big(|X|^m\big) < +\infty$，称 $E\big(X^m\big)$ 为 X 的 m 阶原点矩；称 $E\big((X-EX)^m\big)$ 为 X 的 m 阶中心矩。当 $m > 2$ 时，我们将原点矩和中心矩统称为高阶矩。

♣

4.3 方差

定义 4.5 (随机变量的方差) 如果随机变量 X 的期望 $\mu = EX$ 有限，则称

$$\mathrm{var}(X) := E\big[(X-\mu)^2\big]$$

为 X 的方差（Variance），记作 $\mathrm{var}(X)$ 或 σ_X^2。当 $\mathrm{var}(X) < \infty$ 时，称 X 的方差有限。♣

例 4.12 两点分布 $B(1,p)$：设 $\mathbb{P}(X=1)=p, \mathbb{P}(X=0)=q=1-p$。注意 $X^2 = X, EX = p$，于是

$$\mathrm{var}(X) = E(X^2) - (EX)^2 = p - p^2 = p(1-p) = pq.$$

例 4.13 二项分布 $B(n,p)$：$\mathbb{P}(X=j) = \binom{n}{j} p^j (1-p)^{n-j}$，$0 \le j \le n$。将 X 写成 n 个相互独立的 $B(1,p)$ 之和可得

$$\mathrm{var}(X) = npq, \quad q = 1-p.$$

例 4.14 Poisson 分布 $\mathcal{P}(\lambda)$：$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$，$k = 0, 1, \dots$。有

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

因此 $\mathrm{var}(X) = E(X^2) - (EX)^2 = (\lambda^2+\lambda) - \lambda^2 = \lambda$。

例 4.15 几何分布 $G(p)$（取值 $k = 1, 2, \dots$）：$\mathbb{P}(X=k) = p(1-p)^{k-1}$，记 $q = 1-p$。先算

$$E\big(X(X-1)\big) = \sum_{j=1}^{\infty} j(j-1)\, pq^{j-1} = pq \sum_{j=1}^{\infty} j(j-1) q^{j-2} = pq \cdot \frac{2}{(1-q)^3} = \frac{2q}{p^2}.$$

又 $EX = \frac{1}{p}$，故

$$E(X^2) = E\big(X(X-1)\big) + EX = \frac{2q}{p^2} + \frac{1}{p}.$$

于是

$$\mathrm{var}(X) = E(X^2) - (EX)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}.$$

例 4.16 设 $X \sim \mathcal{E}(\lambda)$（指数分布），即 $f(x) = \lambda e^{-\lambda x}$，$x > 0$。则

$$E(X^2) = \int_0^{\infty} x^2\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda^2}\Gamma(3) = \frac{2}{\lambda^2}.$$

因此

$$\mathrm{var}(X) = E(X^2) - (EX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

例 4.17 设 $X \sim \mathcal{N}(\mu, \sigma^2)$，即 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$。则（由对称性与归一化）

$$\begin{aligned}
\mathrm{var}(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&\xrightarrow{t = \frac{x-\mu}{\sigma}} \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2}\, dt = \sigma^2.
\end{aligned}$$

例 4.18 设 $X \sim \Gamma(\alpha,\beta)$，密度 $f(x)=\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$，$x>0$。有

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 f(x)\,dx = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha+1}e^{-\beta x}dx \\
&\xrightarrow{t=\beta x} \frac{1}{\beta^2\,\Gamma(\alpha)}\int_0^\infty t^{\alpha+1}e^{-t}dt = \frac{\Gamma(\alpha+2)}{\beta^2\,\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\beta^2}.
\end{aligned}
$$

又 $EX=\dfrac{\alpha}{\beta}$，故

$$
\mathrm{var}(X)=E(X^2)-(EX)^2=\frac{\alpha(\alpha+1)}{\beta^2}-\left(\frac{\alpha}{\beta}\right)^2=\frac{\alpha}{\beta^2}.
$$

4.3.1 方差的性质

定理 4.4 (方差的性质)

设 $EX=\mu$ 且 $\mathrm{var}(X)<\infty$，则对任意 $a,b\in\mathbb{R}$：

1. $\mathrm{var}(aX+b)=a^2\ \mathrm{var}(X)$；
2. $\mathrm{var}(X)=E\big[(X-\mu)^2\big]\le E\big[(X-c)^2\big]$，$\forall c\in\mathbb{R}$；
3. $\mathrm{var}(X)=0 \iff X=\mu$ a.s.；
4. 若 $X_1,\dots,X_n$ 相互独立，则
$$
\mathrm{var}\left(\sum_{k=1}^n X_k\right)=\sum_{k=1}^n \mathrm{var}(X_k).
$$

4.3.2 标准差与标准化

定义 4.6 (标准差)

若 EX 有限且 $\mathrm{var}(X)=E\big[(X-\mu)^2\big]$，称

$$
\sigma_X := \sqrt{\mathrm{var}(X)}
$$

为 X 的标准差（Standard deviation）。

定义 4.7 (标准化)

设 $\mathrm{var}(X)<\infty$。令

$$
Y := \frac{X-EX}{\sqrt{\mathrm{var}(X)}}.
$$

则 $EY=0$，$\mathrm{var}(Y)=1$，称 Y 为 X 的标准化（standardization）。特别地，当 $X\sim\mathcal{N}(\mu,\sigma^2)$ 时，

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

4.4 常用的期望方差不等式

定理 4.5 (Markov 不等式)

对随机变量 X 与 $\varepsilon>0$，以及 $\alpha>0$，有

$$
\mathbb{P}\big(|X|\ge\varepsilon\big)\le\frac{E\big(|X|^\alpha\big)}{\varepsilon^\alpha}.
$$

证明 由指标函数记号与 $\{|X|\ge\varepsilon\}\subseteq\{|X|^\alpha\ge\varepsilon^\alpha\}$，得

$$\mathbb{P}(|X|\ge\varepsilon)=E\,I_{\{|X|\ge\varepsilon\}}\le E\left(\frac{|X|^\alpha}{\varepsilon^\alpha}I_{\{|X|^\alpha\ge\varepsilon^\alpha\}}\right)\le E\left(\frac{|X|^\alpha}{\varepsilon^\alpha}\right)=\frac{E\big(|X|^\alpha\big)}{\varepsilon^\alpha}.$$

推论 4.1 (Chebyshev 不等式)

若 $\mathrm{var}(X)<\infty$，则对任意 $\varepsilon>0$，

$$\mathbb{P}\big(|X-EX|\ge\varepsilon\big)\le\frac{\mathrm{var}(X)}{\varepsilon^2}.$$

♡

定理 4.6 (内积不等式（Cauchy–Schwarz）)

设 $EX^2<\infty$ 且 $EY^2<\infty$，则

$$\big|E(XY)\big|\le\sqrt{EX^2\,EY^2}.$$

等号成立当且仅当存在不全为零的常数 a,b 使 $aX+bY=0$ a.s.

♡

证明 对任意 $a,b\in\mathbb{R}$，有 $E\big((aX+bY)^2\big)\ge0$，即

$$a^2EX^2+2ab\,E(XY)+b^2EY^2\ge0.$$

将其视为关于 a/b 的二次式非负，判别式不大于零，得 $E(XY)^2\le EX^2\,EY^2$。若判别式为零，存在不全为零的 a,b 使 $E\big((aX+bY)^2\big)=0$，于是 $aX+bY=0$ a.s.

定理 4.7 (Jensen 不等式)

设 ψ 是 $\mathbb{R}$ 上的凸函数，且 X, $\psi(X)$ 都有有限期望，则

$$\psi(EX)\le E\big(\psi(X)\big).$$

若 ψ 严格凸，则等号成立当且仅当 $X=EX$ a.s.

♡

证明 凸性意味着对任意 $a,b\in\mathbb{R}$，存在某个常数 c（可取为 ψ 的某个次梯度）使得

$$\psi(b)-\psi(a)\ge c\,(b-a).$$

取 $a=EX$, $b=X$，得 $\psi(X)-\psi(EX)\ge c\,(X-EX)$。两侧取期望即得 $E\psi(X)-\psi(EX)\ge c\,E(X-EX)=0$。严格凸情形下等号当且仅当 $X=EX$ a.s.

条件概率复习

一些定义

1. $\mathbf{P(A\mid B)}$: 设事件 A,B，且 $\mathbb{P}(B)>0$。定义
$$\mathbb{P}(A\mid B):=\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$
2. $\mathbf{P(X\mid A)}$: 设随机变量 X 与事件 A，且 $\mathbb{P}(A)>0$。
   - 离散情形:
$$p_{X|A}(x)=\mathbb{P}(X=x\mid A)=\frac{\mathbb{P}(\{X=x\}\cap A)}{\mathbb{P}(A)},\qquad\sum_x p_{X|A}(x)=1.$$
   - 连续情形：存在条件密度 $f_{X|A}(x)$，使得对任意 Borel 集 B，
$$\mathbb{P}(X\in B\mid A)=\int_B f_{X|A}(x)\,dx,\qquad f_{X|A}(x)\ge0,\ \int_{\mathbb{R}}f_{X|A}(x)\,dx=1.$$
3. $\mathbf{P(X\mid Y)}$: 设随机变量 (X,Y)。

- 离散（X, Y 取可数值）：当 $\mathbb{P}(Y = y_j) > 0$ 时，

$$\mathbb{P}(X = x_i \mid Y = y_j) = \frac{\mathbb{P}(X = x_i,\, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{ij}}{q_j},$$

其中 $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$，$q_j = \sum_i p_{ij} = \mathbb{P}(Y = y_j)$。

- 连续（存在联合密度 $f(x, y)$）：当 $f_Y(y) > 0$ 时，

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx,\ x \in \mathbb{R}.$$

进而对 Borel 集 B，$\mathbb{P}(X \in B \mid Y = y) = \int_B f_{X|Y}(x \mid y)\, dx$。

4. $\mathbf{P}(\mathbf{A} \mid \mathbf{X})$：设 D 为 X 的取值域，A 为事件。若存在函数 $g : D \to [0, 1]$ 满足

$$g(x) = \mathbb{P}(A \mid X = x), \quad x \in D,$$

则称 $g(X)$ 为事件 A 关于随机变量 X 的条件概率；g 是定义在 D 上的实函数，因而 $g(X)$ 也是一个随机变量

全概率公式

设 $A_1, \ldots, A_n$ 构成 Ω 的一个分割（两两不交且并为 Ω），且 $\mathbb{P}(A_i) > 0$。

1. 对事件：

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

2. 对随机变量 X 的条件分布：

离散：

$$p_X(x) = \sum_{i=1}^{n} p_{X|A_i}(x)\, \mathbb{P}(A_i).$$

连续：

$$f_X(x) = \sum_{i=1}^{n} f_{X|A_i}(x)\, \mathbb{P}(A_i).$$

3. 对 (X, Y)：若 Y 取有限/可数值 y_i，则

$$p_X(x) = \sum_i p_{X|Y}(x \mid y_i)\, \mathbb{P}(Y = y_i);$$

若 Y 连续，

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, f_Y(y)\, dy.$$

乘法法则

1. 事件：

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \mid A)\, \mathbb{P}(A) = \mathbb{P}(A \mid B)\, \mathbb{P}(B).$$

2. $\mathbf{P}(\mathbf{X} \mid \mathbf{A})$：

离散：

$$\mathbb{P}(\{X = x_i\} \cap A) = \mathbb{P}(X = x_i \mid A)\, \mathbb{P}(A).$$

连续：写作密度的“加权-指示”形式

$$f_{X\, I_{\{X \in A\}}}(x) = f_{X|\{X \in A\}}(x)\, \mathbb{P}(X \in A), \quad 或 \quad f_X(x)\, I_A(x) = f_{X|\{X \in A\}}(x)\, \mathbb{P}(X \in A).$$

3. $\mathbf{P}(\mathbf{X} \mid \mathbf{Y})$：

离散：

$$\mathbb{P}\big(X = x_i, Y = y_j\big) = \mathbb{P}\big(X = x_i \mid Y = y_j\big)\,\mathbb{P}(Y = y_j).$$

连续：

$$f(x, y) = f_{X|Y}(x \mid y)\, f_Y(y).$$

贝叶斯准则

1. 对于事件：设 $A_1, \ldots, A_n$ 构成 Ω 的一个分割，B 是一个事件，且 $\mathbb{P}(B) > 0, \mathbb{P}(A_i) > 0$。
$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B \mid A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j)\mathbb{P}(B \mid A_j)}, \quad i = 1, \ldots, n.$$
2. 对于随机变量 (X, Y)，分情况讨论：
   (a). X, Y 均为离散型随机变量：
   $$P_{X|Y}(x \mid y) = \frac{P_{X,Y}(x, y)}{P_Y(y)} = \frac{P_X(x)P_{Y|X}(y \mid x)}{P_Y(y)},$$
   其中
   $$P_Y(y) = \sum_x P_X(x)P_{Y|X}(y \mid x).$$
   (b). (X, Y) 为连续型随机向量：
   $$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y \mid x)}{f_Y(y)},$$
   其中
   $$f_Y(y) = \int_x f_X(x)f_{Y|X}(y \mid x)\, dx.$$
   (c). X 离散，Y 连续：
   $$P_{X|Y}(x \mid y) = \frac{P_X(x)f_{Y|X}(y \mid x)}{f_Y(y)},$$
   其中
   $$f_Y(y) = \sum_x P_X(x)f_{Y|X}(y \mid x).$$
   (d). X 连续，Y 离散：
   $$f_{X|Y}(x \mid y) = \frac{f_X(x)P_{Y|X}(y \mid x)}{P_Y(y)},$$
   其中
   $$P_Y(y) = \int_x f_X(x)P_{Y|X}(y \mid x)\, dx.$$

4.5 条件在事件上的条件期望

定义 4.8 (离散型)

若有 $\sum_{i=1}^{\infty} |x_i|\mathbb{P}(X = x_i \mid A) < \infty$，则定义给定 A 发生下，X 的期望为

$$\mathbb{E}(X \mid A) = \sum_{i=1}^{\infty} x_i\mathbb{P}(X = x_i \mid A).$$

♣

定义 4.9 (连续型)

若有 $\int_{-\infty}^{\infty}|x|f_{X|A}(x \mid A)\,dx<\infty$，则定义给定 A 发生下，X 的期望为

$$\mathbb{E}(X \mid A)=\int_{-\infty}^{\infty} x f_{X|A}(x \mid A)\,dx.$$

♣

定理 4.8

设 $\mathbb{P}(A)>0$，X 是非负随机变量，则

$$\mathbb{E}(X \mid A)=\frac{\mathbb{E}(XI_A)}{\mathbb{P}(A)}.$$

♡

推论 4.2

设 $\mathbb{P}(A)>0$，$\mathbb{E}(X \mid A)$ 存在，则有

$$\mathbb{E}(X \mid A)=\frac{\mathbb{E}(XI_A)}{\mathbb{P}(A)}.$$

♡

例 4.19 设 $X\sim\mathcal{E}(\lambda)$，对 $a>0$，证明

$$\mathbb{E}(X \mid X>a)=a+\frac{1}{\lambda}.$$

解 X 有密度函数 $f(x)=\lambda e^{-\lambda x}, x>0$。利用定理 2.1，有

$$\mathbb{E}(X \mid X>a)=\frac{\mathbb{E}(XI_{\{X>a\}})}{\mathbb{P}(X>a)}=\frac{1}{e^{-\lambda a}}\int_a^{\infty} x\lambda e^{-\lambda x}\,dx=\frac{1}{\lambda}+a.$$

事实上，该结果也可由指数分布的无记忆性获得:

$$\mathbb{E}(X \mid X>a)=a+\mathbb{E}(X-a \mid X>a)=a+\mathbb{E}(X)=a+\frac{1}{\lambda}.$$

因为指数分布具有无记忆性，在条件 $X>a$ 下，$X-a$ 依然服从指数分布 $\mathcal{E}(\lambda)$。同时 $\mathbb{E}(X)=1/\lambda$。

例 4.20 设 X,Y 独立, 且 $X\sim\mathcal{E}(\lambda), Y\sim\mathcal{E}(\mu)$，计算 $\mathbb{E}(X \mid X<Y)$。

解 由定理得

$$\begin{aligned}
\mathbb{E}(X \mid X<Y)=\frac{\mathbb{E}(XI_{\{X<Y\}})}{\mathbb{P}(X<Y)}&=\frac{\int_0^{\infty}\int_0^{\infty} xI_{\{x<y\}}\lambda e^{-\lambda x}\mu e^{-\mu y}\,dy\,dx}{\int_0^{\infty}\int_0^{\infty} I_{\{x<y\}}\lambda e^{-\lambda x}\mu e^{-\mu y}\,dy\,dx}\\
&=\frac{\int_0^{\infty} x\lambda e^{-\lambda x}\left(\int_x^{\infty}\mu e^{-\mu y}\,dy\right)dx}{\int_0^{\infty}\lambda e^{-\lambda x}\left(\int_x^{\infty}\mu e^{-\mu y}\,dy\right)dx}=\frac{\int_0^{\infty} x\lambda e^{-\lambda x}e^{-\mu x}\,dx}{\int_0^{\infty}\lambda e^{-\lambda x}e^{-\mu x}\,dx}\\
&=\frac{\lambda\int_0^{\infty} xe^{-(\lambda+\mu)x}\,dx}{\lambda\int_0^{\infty} e^{-(\lambda+\mu)x}\,dx}=\frac{\lambda\cdot\frac{1}{(\lambda+\mu)^2}}{\lambda\cdot\frac{1}{\lambda+\mu}}=\frac{1}{\lambda+\mu}.
\end{aligned}$$

定理 4.9 (全期望定理)

设 $A_1, \ldots, A_n$ 构成 Ω 的一个分割，则

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{E}(X \mid A_i).$$

♡

例 4.21 利用全期望定理计算几何分布 $X \sim G(p)$ 的期望。

解 令 $A_1 = \{X = 1\}, A_2 = \{X > 1\}$。这是一个分割。

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X \mid A_1)\mathbb{P}(A_1) + \mathbb{E}(X \mid A_2)\mathbb{P}(A_2) \\ &= 1 \cdot p + \mathbb{E}(X \mid X > 1) \cdot (1 - p). \end{aligned}$$

而由于几何分布的无记忆性，$\mathbb{E}(X - 1 \mid X > 1) = \mathbb{E}(X)$，所以

$$\mathbb{E}(X \mid X > 1) = \mathbb{E}(X - 1 \mid X > 1) + 1 = \mathbb{E}(X) + 1.$$

代入上式可得

$$\mathbb{E}(X) = p + (\mathbb{E}(X) + 1)(1 - p) = p + \mathbb{E}(X) - p\mathbb{E}(X) + 1 - p.$$

整理得 $p\mathbb{E}(X) = 1$，故 $\mathbb{E}(X) = 1/p$。

4.6 条件在随机变量上的条件期望

定义 4.10 (离散型)

设 (X, Y) 为离散型随机向量。若 $\sum_{i=1}^{\infty} |x_i|\mathbb{P}(X = x_i \mid Y = y) < \infty$，则定义给定 $Y = y$ 下，X 的期望为

$$\mathbb{E}(X \mid Y = y) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i \mid Y = y).$$

♣

定义 4.11 (连续型)

设 (X, Y) 为连续型随机向量。若 $\int_{-\infty}^{\infty} |x| f_{X|Y}(x \mid y)\, dx < \infty$，则定义给定 $Y = y$ 下，X 的期望为

$$\mathbb{E}(X \mid Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx.$$

♣

定义 4.12 (E(X|Y) 的定义)

设 (X, Y) 是随机向量，$\mathbb{E}|X| < \infty$。如果

$$m(y) = \mathbb{E}(X \mid Y = y),$$

就称随机变量 $m(Y)$ 为给定 Y 时 X 的条件期望，记作 $\mathbb{E}(X \mid Y)$。

♣

定理 4.10 (重期望法则)

设 X, Y 是随机变量，则

$$\mathbb{E}[\mathbb{E}(X \mid Y)] = \mathbb{E}(X).$$

♡

例 4.22 密室逃脱问题 一位玩家在有三个门的大型密室逃脱游戏中迷了路，第一个门通到一密道走 3 小时可使他到达出口。第二个门通向使他走 5 小时后又回到原地的密道，第三个门通向使他走了 7 小时后又回到原地的密道。如果他在任何时刻都等可能地选定其中一个门。试问他到达出口平均要花多少时间？

解 设 X 表示他到达出口所需的时数，Y 表示他最初选定的门的编号，于是

$$\mathbb{P}(Y=1)=\mathbb{P}(Y=2)=\mathbb{P}(Y=3)=1/3.$$

由全期望公式，所求平均时数为

$$\mathbb{E}(X)=\sum_{i=1}^{3}\mathbb{E}(X\mid Y=i)\mathbb{P}(Y=i)=\frac{\mathbb{E}(X\mid Y=1)+\mathbb{E}(X\mid Y=2)+\mathbb{E}(X\mid Y=3)}{3}.$$

又因为

$$\begin{aligned}\mathbb{E}(X\mid Y=1)&=3,\\ \mathbb{E}(X\mid Y=2)&=5+\mathbb{E}(X),\\ \mathbb{E}(X\mid Y=3)&=7+\mathbb{E}(X).\end{aligned}$$

所以

$$\mathbb{E}(X)=\frac{3+5+\mathbb{E}(X)+7+\mathbb{E}(X)}{3}.$$

解之得 $\mathbb{E}(X)=15$。

例 4.23 设计算机使用的环境指标 $Y\sim\Gamma(\alpha,\beta)$，其概率密度是

$$f_Y(y)=\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y},\quad y>0.$$

已知 $Y=y$ 时，计算机软件的使用寿命 $X\sim\mathcal{E}(y)$。计算条件期望 $\mathbb{E}(X|Y)$，$\mathbb{E}(Y|X)$ 和 $\mathbb{E}(X)$。

解 给定 $Y=y$ 时，$X\sim\mathcal{E}(y)$，即条件密度为 $f_{X|Y}(x|y)=ye^{-xy},x>0$。于是

$$\mathbb{E}(X|Y=y)=\int_0^\infty xye^{-xy}dx=\frac{1}{y},$$

得 $\mathbb{E}(X|Y)=1/Y$。

(X,Y) 的联合密度为

$$f(x,y)=f_{X|Y}(x|y)f_Y(y)=ye^{-xy}\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y},\quad x,y>0.$$

X 有边缘密度函数

$$f_X(x)=\int_0^\infty f(x,y)dy=\frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}},\quad x>0.$$

于是在条件 $X=x$ 下，Y 有条件密度

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{(x+\beta)^{\alpha+1}}{\Gamma(\alpha+1)}y^\alpha e^{-(x+\beta)y},$$

这是 $\Gamma(\alpha+1,x+\beta)$ 分布的密度，其期望为 $(\alpha+1)/(x+\beta)$。所以有

$$\mathbb{E}(Y|X=x)=\int_0^\infty yf_{Y|X}(y|x)dy=\frac{\alpha+1}{x+\beta}.$$

于是 $\mathbb{E}(Y|X)=\dfrac{\alpha+1}{X+\beta}$。

由于

$$\begin{aligned}\mathbb{E}(X)=\mathbb{E}[\mathbb{E}(X|Y)]=\mathbb{E}(Y^{-1})&=\int_0^\infty y^{-1}f_Y(y)dy\\ &=\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty y^{\alpha-2}e^{-\beta y}dy\\ &=\frac{\beta^\alpha\Gamma(\alpha-1)\beta^{-(\alpha-1)}}{\Gamma(\alpha)}=\frac{\beta\Gamma(\alpha-1)}{(\alpha-1)\Gamma(\alpha-1)}=\frac{\beta}{\alpha-1}.\end{aligned}$$

此积分仅当 $\alpha-1>0$ 时收敛。因此

$$\mathbb{E}(X)=\begin{cases}\dfrac{\beta}{\alpha-1}, & \alpha>1,\\ \infty, & \alpha\in(0,1].\end{cases}$$

定理 4.11 (条件期望的性质)

设 X, Y 是随机变量, $g(x), h(y)$ 是实函数, $\mathbb{E}|X| < \infty, \mathbb{E}|g(X)| < \infty$。又设 r.v. $X_1, X_2, \ldots, X_n$ 的期望有限, 则

1. $|\mathbb{E}(X|Y)| \leq \mathbb{E}(|X||Y)$;
2. $[\mathbb{E}(X|Y)]^2 \leq \mathbb{E}(X^2|Y)$;
3. $\mathbb{E}[h(Y)g(X)|Y] = h(Y)\mathbb{E}[g(X)|Y]$;
4. $\mathbb{E}[\mathbb{E}(g(X)|Y)] = \mathbb{E}g(X)$.
5. 当 X, Y 独立时, $\mathbb{E}\{g(X)|Y\} = \mathbb{E}g(X)$;
6. $\mathbb{E}(c + \sum_{i=1}^n c_i X_i|Y) = c + \sum_{i=1}^n c_i \mathbb{E}(X_i|Y)$;
7. 如果 $X_1 \leq X_2$, 则 $\mathbb{E}(X_1|Y) \leq \mathbb{E}(X_2|Y)$;

♡

例 4.24 设 X, Y 是随机变量, $h(x)$ 是实函数, 若 $\mathbb{E}(X^2) < \infty, \mathbb{E}[h^2(Y)] < \infty$, 则

$$\mathbb{E}[(X - \mathbb{E}(X|Y))h(Y)] = 0$$

证明 由 Cauchy-Schwarz 不等式得

$$\mathbb{E}(|Xh(Y)|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}[h^2(Y)]} < \infty.$$

利用条件期望的性质, 可得

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}(X|Y))h(Y)] &= \mathbb{E}[Xh(Y)] - \mathbb{E}[\mathbb{E}(X|Y)h(Y)] \\
&= \mathbb{E}[Xh(Y)] - \mathbb{E}[\mathbb{E}(Xh(Y)|Y)] \\
&= \mathbb{E}[Xh(Y)] - \mathbb{E}[Xh(Y)] \\
&= 0.
\end{aligned}$$

例 4.25 最佳预测问题 设 $\mathbb{E}(X^2) < \infty, m(Y) = \mathbb{E}(X|Y)$, 则对任何实函数 $g(y)$, 有

$$\mathbb{E}[X - m(Y)]^2 \leq \mathbb{E}[X - g(Y)]^2,$$

其中等号成立当且仅当 $g(Y) = m(Y)$ a.s.

证明 当 $\mathbb{E}g^2(Y) = \infty$, 利用不等式 $b^2 \leq 2(a-b)^2 + 2a^2$ 可得

$$\infty = \mathbb{E}g^2(Y) \leq 2\mathbb{E}[X - g(Y)]^2 + 2\mathbb{E}(X^2),$$

从而 $\mathbb{E}[X - g(Y)]^2 = \infty$, 不等式成立.

下面设 $\mathbb{E}g^2(Y) < \infty$. 显然, $[\mathbb{E}(X|Y)]^2 \leq \mathbb{E}(X^2|Y)$, 从而 $\mathbb{E}[m(Y)]^2 \leq \mathbb{E}(X^2) < \infty$. 令 $h(Y) = m(Y) - g(Y)$, 则 $\mathbb{E}[h(Y)]^2 < \infty$. 于是

$$\begin{aligned}
\mathbb{E}[X - g(Y)]^2 &= \mathbb{E}[X - m(Y) + m(Y) - g(Y)]^2 \\
&= \mathbb{E}[X - m(Y)]^2 + \mathbb{E}[m(Y) - g(Y)]^2 \\
&\quad + 2\mathbb{E}((X - m(Y))[m(Y) - g(Y)]) \\
&= \mathbb{E}[X - m(Y)]^2 + \mathbb{E}[m(Y) - g(Y)]^2 \\
&\geq \mathbb{E}[X - m(Y)]^2.
\end{aligned}$$

式中等号成立当且仅当 $\mathbb{E}[m(Y) - g(Y)]^2 = 0$, 当且仅当 $g(Y) = m(Y)$ a.s.

称 $m(Y) = \mathbb{E}(X|Y)$ 是 X 的最佳预测 (optimal forecast)。

4.7 条件方差

定义 4.13 (条件方差)

如果 $\mathbb{E}\{[X-\mathbb{E}(X|Y)]^2|Y\}$ 存在, 则称它为给定 Y 下, X 的条件方差, 记作 $\mathrm{var}(X|Y)$, 即

$$\mathrm{var}(X|Y) = \mathbb{E}\{[X-\mathbb{E}(X|Y)]^2|Y\}.$$

♣

定理 4.12 (全方差法则)

$$\mathrm{var}(X) = \mathbb{E}[\mathrm{var}(X|Y)] + \mathrm{var}(\mathbb{E}[X|Y]).$$

♡

证明

$$\begin{aligned}
\mathrm{var}(X) &= E(X-EX)^2 \\
&= E\{(X-E(X|Y)) + (E(X|Y)-EX)\}^2 \\
&= E\{[X-E(X|Y)] + [E(X|Y)-E(E[X|Y])]\}^2 \\
&= E[(X-E(X|Y))^2] + \mathrm{var}(E[X|Y]) \\
&= E(E\{[X-E(X|Y)]^2|Y\}) + \mathrm{var}(E[X|Y]) \\
&= E[\mathrm{var}(X|Y)] + \mathrm{var}(E[X|Y]).
\end{aligned}$$

例 4.26 论文数问题 设 Y 是我校明年选修概率论课程的学生人数, X_i 是第 i 位同学在学习后 5 年内写论文的篇数, X 是该段时间内这些同学写的论文的总篇数. 假设每位同学的论文数 X_i 是相互独立同分布且与选课人数 Y 也相互独立. 如果 $\mathbb{E}(X_1), \mathbb{E}(Y), \mathrm{var}(X_1), \mathrm{var}(Y)$ 已知, 求 X 的期望与方差.

解 显然 $X = X_1 + \cdots + X_Y$, 又因为

$$\begin{aligned}
\mathbb{E}(X|Y=n) &= \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = n\mathbb{E}(X_1) \\
\mathrm{var}(X|Y=n) &= \mathrm{var}(X_1 + \cdots + X_n) = n\,\mathrm{var}(X_1).
\end{aligned}$$

所以

$$\mathbb{E}(X|Y) = Y\mathbb{E}(X_1), \quad \mathrm{var}(X|Y) = Y\,\mathrm{var}(X_1).$$

从而

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}[Y\mathbb{E}(X_1)] = \mathbb{E}(Y)\mathbb{E}(X_1), \\
\mathrm{var}(X) &= \mathbb{E}[\mathrm{var}(X|Y)] + \mathrm{var}(\mathbb{E}[X|Y]) \\
&= \mathbb{E}(Y)\,\mathrm{var}(X_1) + [\mathbb{E}(X_1)]^2\,\mathrm{var}(Y).
\end{aligned}$$

4.8 协方差

定义 4.14 (协方差 (Covariance))

设 $\mathbb{E}X$ 和 $\mathbb{E}Y$ 存在, 当 $\mathbb{E}|(X-\mathbb{E}X)(Y-\mathbb{E}Y)| < \infty$ 时, 称

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

为随机变量 X, Y 的协方差 (covariance), 记作 $\mathrm{cov}(X, Y)$.

♣

定理 4.13 (协方差性质)

1. $\operatorname{cov}(X,X)=\operatorname{var}(X)$;
2. $\operatorname{cov}(X,Y)=\operatorname{cov}(Y,X)$;
3. $\operatorname{cov}(X,Y)=\mathbb{E}(XY)-(\mathbb{E}X)(\mathbb{E}Y)$; (常用计算公式)
4. 若 c 是某常数, 则 $\operatorname{cov}(X,c)=0$;
5. 双线性: $\operatorname{cov}(cX,Y)=c\operatorname{cov}(X,Y), \operatorname{cov}(X,Y+Z)=\operatorname{cov}(X,Y)+\operatorname{cov}(X,Z)$;
6. $\operatorname{var}(X_1+X_2)=\operatorname{var}(X_1)+\operatorname{var}(X_2)+2\operatorname{cov}(X_1,X_2)$

♡

4.9 相关系数

引入 X,Y 的标准化

$$\frac{X-\mathbb{E}X}{\sqrt{\operatorname{var}(X)}},\quad \frac{Y-\mathbb{E}Y}{\sqrt{\operatorname{var}(Y)}}.$$

则

$$\operatorname{cov}\left(\frac{X-\mathbb{E}X}{\sqrt{\operatorname{var}(X)}},\frac{Y-\mathbb{E}Y}{\sqrt{\operatorname{var}(Y)}}\right)=\mathbb{E}\left[\left(\frac{X-\mathbb{E}X}{\sqrt{\operatorname{var}(X)}}\right)\left(\frac{Y-\mathbb{E}Y}{\sqrt{\operatorname{var}(Y)}}\right)\right]=\frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

定义 4.15 (相关系数 (Correlation coefficient))

当 $0<\operatorname{var}(X)\cdot\operatorname{var}(Y)<\infty$ 时, 称

$$\frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

为 X,Y 的相关系数 (Pearson's correlation). 记作 $\operatorname{corr}(X,Y)$.

♣

例 4.27 设 $(X,Y)\sim\mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则 $\operatorname{corr}(X,Y)=\rho$.

解 不失一般性, 考虑标准联合正态分布 $\mathcal{N}(0,0,1,1,\rho)$. (X,Y) 有联合密度

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

则,

$$\operatorname{cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}X\cdot\mathbb{E}Y=\int_{R^2}xyf(x,y)dxdy-0\cdot 0=\rho.$$

$$\operatorname{corr}(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}=\frac{\rho}{1\cdot 1}=\rho$$

这是为什么经常记作符号 ρ_{XY}.

定理 4.14

设 X,Y 相互独立, 则有 X,Y 不相关, 即 $\operatorname{cov}(X,Y)=0$.

♡

证明

$$\operatorname{cov}(X,Y)=\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]=\mathbb{E}(X-\mathbb{E}X)\cdot\mathbb{E}(Y-\mathbb{E}Y)=0$$

例 4.28 设 (X,Y) 在单位圆 $D=\{(x,y)|x^2+y^2\le 1\}$ 内均匀分布, 则 X,Y 不相关, 也不独立.

解 (X,Y) 有联合密度

$$f(x,y)=\frac{1}{\pi}I_D.$$

则

$$\begin{aligned}\mathbb{E}(X) &= \int_{R^2} xf(x,y)dxdy \\ &= \frac{1}{\pi}\int_{-1}^{1}\left(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xdx\right)dy = 0.\end{aligned}$$

同理 $\mathbb{E}(Y)=0$. 于是

$$\mathrm{cov}(X,Y) = \int_{R^2} xyf(x,y)dxdy = \frac{1}{\pi}\int_{-1}^{1} y\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xdxdy = 0.$$

相关性与独立性的关系是：$\boxed{\text{若独立, 则不相关}}$，但反之不成立，如上例所示。

相关系数的几何意义

考虑 $\tilde{X}=X-\mathbb{E}X, \tilde{Y}=Y-\mathbb{E}Y$. 则 $\mathbb{E}\tilde{X}=0, \mathbb{E}\tilde{Y}=0$.

$$\mathrm{corr}(X,Y) = \mathrm{corr}(\tilde{X}-\mathbb{E}X, Y-\mathbb{E}Y)$$

$$\mathrm{corr}(X,Y) = \mathrm{corr}(\tilde{X},\tilde{Y}) = \frac{\mathrm{cov}(\tilde{X},\tilde{Y})}{\sqrt{\mathrm{var}(\tilde{X})}\sqrt{\mathrm{var}(\tilde{Y})}} = \frac{\mathbb{E}(\tilde{X}\tilde{Y})}{\sqrt{\mathbb{E}(\tilde{X}^2)}\sqrt{\mathbb{E}(\tilde{Y}^2)}} = \frac{\langle \tilde{X},\tilde{Y}\rangle}{|\tilde{X}|\cdot|\tilde{Y}|} = \cos(\varphi)$$

$\boxed{|\mathrm{corr}(X,Y)|\le 1; |\mathrm{corr}(X,Y)|=1 \text{ 等价于 } X,Y \text{ 共线性}}$

统计意义总结表

	$\mathbb{E}Y$	$\mathbb{E}(Y\vert X)$	$\mathrm{corr}(X,Y)$
代数	$\mathbb{E}(Y)=\int_R ydF(y)$.	$\mathbb{E}(Y\vert X)=m(X)$, $m(x)=\mathbb{E}(Y\vert X=x)$.	$\mathrm{corr}(X,Y)$ $=\frac{\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$
几何	重心	投影	夹角余弦
优化	常数中的最佳预测 $\mathbb{E}Y=\underset{c\in\mathbb{R}}{\mathrm{argmin}}\mathbb{E}(Y-c)^2$.	$g(X)$ 中的最佳预测 $\mathbb{E}(Y\vert X)=m(X)=$ $\underset{g:\mathbb{R}\to\mathbb{R}}{\mathrm{argmin}}\mathbb{E}[Y-g(X)]^2$.	拟合直线中的最佳预测 $\rho\sqrt{\frac{\mathrm{var}(Y)}{\mathrm{var}(X)}}=$ $\underset{b\in\mathbb{R}}{\mathrm{argmin}}\mathbb{E}[(Y-\mathbb{E}Y)-b(X-\mathbb{E}X)]^2$.

4.10 随机向量的期望与协方差矩阵

设 $\mathbf{X} = (X_1, \ldots, X_n)^T$ 是随机向量, 如果对每个 i, $\mu_i = \mathbb{E}X_i$ 存在, 就称 $\mathbf{X}$ 的数学期望存在, 并且定义

$$\mu := \mathbb{E}\mathbf{X} = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)^T = (\mu_1, \ldots, \mu_n)^T.$$

类似地, 对随机矩阵

$$\mathbf{Y} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}$$

可以定义其数学期望

$$\mathbb{E}\mathbf{Y} = \begin{pmatrix} \mathbb{E}X_{11} & \mathbb{E}X_{12} & \cdots & \mathbb{E}X_{1n} \\ \mathbb{E}X_{21} & \mathbb{E}X_{22} & \cdots & \mathbb{E}X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}X_{m1} & \mathbb{E}X_{m2} & \cdots & \mathbb{E}X_{mn} \end{pmatrix}.$$

设 $\mathbf{X}, \mathbf{Y}$ 如上定义, 且数学期望都存在. 容易证明, 对任何常数向量 $\mathbf{a} = (a_1, \ldots, a_n)^T$, 常数矩阵 $\mathbf{A}_{k\times m}, \mathbf{B}_{n\times j}$, 有

性质

1. $\mathbb{E}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\mathbb{E}\mathbf{X}$;
2. $(\mathbb{E}\mathbf{Y})^T = \mathbb{E}(\mathbf{Y}^T)$;
3. $\mathbb{E}(\mathbf{AY}) = \mathbf{A}\mathbb{E}(\mathbf{Y})$;
4. $\mathbb{E}(\mathbf{YB}) = \mathbb{E}(\mathbf{Y})\mathbf{B}$;
5. $\mathbb{E}(\mathbf{AYB}) = \mathbf{A}\mathbb{E}(\mathbf{Y})\mathbf{B}$;

本质上都基于期望的线性性质.

定义 4.16 (协方差矩阵 (Covariance Matrix))

如果随机向量 $\mathbf{X}$ 的数学期望 $\mu = \mathbb{E}\mathbf{X}$ 存在, 对每个分量 X_i 的方差有限, 则称

$$\Sigma = \mathbb{E}[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] = (\sigma_{ij})$$

为 $\mathbf{X}$ 的协方差矩阵, 其中 $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$. ♣

协方差矩阵 Σ 是对称矩阵. 如果行列式 $\det(\Sigma) = 0$, 就称 Σ 是退化的.

定理 4.15

设 Σ 是 $\mathbf{X} = (X_1, \ldots, X_n)$ 的协方差矩阵, 则

1. Σ 是非负定矩阵;
2. Σ 退化的充要条件是存在不全为零的常数 $a_1, \ldots, a_n$ 使得

$$\sum_{i=1}^{n} a_i(X_i - \mathbb{E}X_i) = 0 \quad a.s.$$

♡

第 5 节 特征函数和概率极限定理

5.1 概率母函数

定义 5.1 (probability-generating function)

设 X 是取非负整值的随机变量, 称

$$g(s)=\mathbb{E}(s^X)=\sum_{j=0}^{\infty}s^j\mathbb{P}(X=j), s\in[-1,1],$$

为 X 的概率母函数 (probability-generating function). 约定 $0^0=1$. ♣

定理 5.1 (概率母函数的性质)

设 $g(s)$ 是 X 的概率母函数, $g^{(k)}(s)$ 是 $g(s)$ 的 k-阶导数. 则

1. $\mathbb{P}(X=k)=\dfrac{g^{(k)}(0)}{k!}, k=0,1,\ldots;$
2. $\mathbb{E}(X)=g^{(1)}(1);$
3. 如果 $\mathbb{E}(X)<\infty$, 则 $\mathrm{var}(X)=g^{(2)}(1)+g^{(1)}(1)-[g^{(1)}(1)]^2;$
4. 如果 $X_1,\ldots,X_n$ 相互独立, $g_i(s)=\mathbb{E}(s^{X_i})$ 是 X_i 的概率母函数, 则 $Y=X_1+\cdots+X_n$ 的概率母函数为

$$g_Y(s)=g_1(s)g_2(s)\cdots g_n(s), s\in[-1,1].$$

♡

5.1.1 常见分布的概率母函数

例 5.1 二项分布 $B(n,p)$ 二项分布 $B(n,p)$ 的概率母函数是

$$g(s)=\sum_{j=0}^{n}s^j\binom{n}{j}p^jq^{n-j}=(q+sp)^n.$$

应用: 设 $X_1,X_2,\ldots,X_m$ 是相互独立, 且 $X_i\sim B(n_i,p)$, 则

$$Y=X_1+X_2+\cdots+X_m\sim B(n_1+n_2+\cdots+n_m,p).$$

例 5.2 Poisson 分布 $\mathcal{P}(\lambda)$ Poisson 分布 $\mathcal{P}(\lambda)$ 的概率母函数是

$$g(s)=\sum_{k=0}^{\infty}s^k\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(s-1)}.$$

应用: 设 $X_1,X_2,\ldots,X_m$ 是相互独立, 且 $X_i\sim\mathcal{P}(\lambda_i)$, 则

$$Y=X_1+X_2+\cdots+X_m\sim\mathcal{P}(\lambda_1+\lambda_2+\cdots+\lambda_m).$$

例 5.3 几何分布 $G(p)$ 几何分布 $G(p)$ 的概率母函数是

$$g(s)=\sum_{k=1}^{\infty}s^kpq^{k-1}=\frac{sp}{1-sq}.$$

应用: 设 $X_1,X_2,\ldots,X_m$ 相互独立, 且 $X_i\sim G(p)$, 则 $S_m=X_1+X_2+\cdots+X_m$ 有概率母函数

$$g_m(s)=\left(\frac{sp}{1-sq}\right)^m.$$

将 $g_m(s)$ Taylor 展开得到

$$\begin{aligned} g_m(s) &= (sp)^m \sum_{j=0}^{\infty} \frac{m(m+1)\cdots(m+j-1)}{j!}(sq)^j \\ &= (sp)^m \sum_{j=0}^{\infty} \binom{m+j-1}{m-1}(sq)^j \\ &= \sum_{k=m}^{\infty} \binom{k-1}{m-1} p^m q^{k-m} s^k. \end{aligned}$$

于是得到 Pascal 分布

$$\mathbb{P}(S_m = k) = \binom{k-1}{m-1} p^m q^{k-m}, k = m, m+1, \ldots$$

定理 5.2 (随机和的概率母函数)

设

1. $\{X_j\}$ 是一列独立同分布的非负整值随机变量, 且 X_1 的概率母函数为
$$g_X(s) = \sum_{k=0}^{\infty} p_k s^k, p_k = \mathbb{P}(X_1 = k).$$
2. N 为取正整数值的随机变量, 且其概率母函数为
$$g_N(s) = \sum_{n=1}^{\infty} s^n \mathbb{P}(N = n).$$
3. N 与 $\{X_j\}$ 相互独立.

则 $W = X_1 + X_2 + \cdots + X_N$ 的概率母函数为 $g_W(s) = g_N[g_X(s)]$.

♡

例 5.4 设鸡蛋个数 N 服从 $\mathcal{P}(\lambda)$, λ 为正常数. 如果每颗蛋能孵出小鸡的概率为 p, 则能孵出小鸡的鸡蛋数 W 服从 $\mathcal{P}(\lambda p)$.

解 设 $X_i = 1$ or 0, 根据第 i 个蛋有无孵出小鸡, $i = 1, 2, \ldots$, 且 $W = X_1 + X_2 + \cdots + X_N$ 为能孵出小鸡的鸡蛋数.

由于 $\{X_i\}$ 独立同分布且 X_1 的概率母函数为 $g_X(s) = q + ps$ $(q = 1 - p)$, 而 N 的概率母函数为 $g_N(s) = e^{\lambda(s-1)}$.

利用定理 5.2 得, W 的概率母函数为

$$g_W(s) = g_N[g_X(s)] = e^{\lambda(q+ps)-1} = e^{\lambda p(s-1)}.$$

由可逆性知, $W \sim \mathcal{P}(\lambda p)$.

5.1.2 随机向量的概率母函数

定义 5.2 (2-dimensional probability-generating function)

设 (X, Y) 是二维取非负整数值的随机向量, 其分布列为 $p_{ik} = \mathbb{P}(X = i, Y = k), i, k = 0, 1, 2, \ldots$, 称 (s, t) 的函数

$$g(s, t) = \mathbb{E}(s^X t^Y) = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} p_{ik} s^i t^k, s, t \in [-1, 1],$$

为 (X, Y) 的概率母函数.

♣

定理 5.3 (随机向量的概率母函数的性质)

设 $g(s,t)$ 是 (X,Y) 的概率母函数, 则

1. $|g(s,t)| \leq g(1,1) = 1, |s| \leq 1, |t| \leq 1$;
2. $g_{aX+bY+c}(s) = s^c g(s^a, s^b)$, 其中 a,b,c 均为常数;
3. X 与 Y 独立, 等价于 $g(s,t) = g_X(s)g_Y(t), \forall s,t$;
4. $g(s,1) = g_X(s), g(1,t) = g_Y(t)$;
5. $$p_{ik} = \frac{1}{i!k!}\frac{\partial^{i+k} g(s,t)}{\partial s^i \partial t^k}\bigg|_{s=t=0}, \quad i,k = 0,1,2,\ldots;$$
6. 如果 $\mathbb{E}(X), \mathbb{E}(Y) < \infty$, 则
$$\mathbb{E}(X) = \frac{\partial g(s,t)}{\partial s}\bigg|_{s=t=1}, \quad \mathbb{E}(Y) = \frac{\partial g(s,t)}{\partial t}\bigg|_{s=t=1}.$$
7. 如果 $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$, 则
$$\mathbb{E}(X^2) = \frac{\partial^2 g(s,t)}{\partial s^2}\bigg|_{s=t=1} + \frac{\partial g(s,t)}{\partial s}\bigg|_{s=t=1},$$
$$\mathbb{E}(Y^2) = \frac{\partial^2 g(s,t)}{\partial t^2}\bigg|_{s=t=1} + \frac{\partial g(s,t)}{\partial t}\bigg|_{s=t=1},$$
$$\mathbb{E}(XY) = \frac{\partial^2 g(s,t)}{\partial s \partial t}\bigg|_{s=t=1}.$$

♡

5.2 矩母函数

定义 5.3 (moment-generating function 矩母函数)

设 X 是随机向量, 称
$$M_X(s) = \mathbb{E}(e^{sX}),$$
为 X 的矩母函数 (moment-generating function).

- 离散变量: $M_X(s) = \sum_j e^{sx_j}\mathbb{P}(X = x_j)$,
- 连续变量: $M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)dx$.

仅当 $\mathbb{E}(e^{sX}) < \infty$ 时, 我们称 $M_X(s)$ 存在.

♣

定理 5.4 (矩母函数的性质)

设 $M(s)$ 是 X 的矩母函数, $M^{(k)}(s)$ 是 $M(s)$ 的 k-阶导数. 则

1. $Y = aX + b$ 的矩母函数为 $M_Y(s) = e^{sb}M(sa)$;
2. $\mathbb{E}X^k = M^{(k)}(0), k = 1,2,\ldots$;
3. $M(0) = 1$; 当 X 仅取非负整数值时, $P(X = 0) = \lim_{s\to-\infty} M(s)$;
4. 可逆性: 如果存在一个正数 a, 使得对任意 $s \in [-a,a]$ 有 $M(s) < \infty$, 则 $M(s)$ 唯一地决定 X 的分布函数;
5. 如果 $X_1, X_2, \ldots, X_n$ 相互独立, $M_{X_i}(s) = \mathbb{E}(e^{sX_i})$ 是 X_i 的矩母函数, 则 $Y = X_1 + \cdots + X_n$ 的矩母函数为
$$M_Y(s) = M_{X_1}(s)\cdots M_{X_n}(s)$$

♡

5.2.1 部分常见随机变量的矩母函数

例 5.5 设离散型随机变量 X 的分布列为

X	2	3	5
$\mathbb{P}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

求其矩母函数 $M(s)$, 期望 $\mathbb{E}X$, 二阶矩 $\mathbb{E}X^2$.

解

$$\begin{aligned} M(s) &= \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}. \\ \mathbb{E}X &= M^{(1)}(0) = \frac{1}{2}2e^{2s} + \frac{1}{6}3e^{3s} + \frac{1}{3}5e^{5s}\Big|_{s=0} \\ &= \frac{1}{2}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{3}\cdot 5 = \frac{19}{6}. \\ \mathbb{E}X^2 &= M^{(2)}(0) = \frac{1}{2}4e^{2s} + \frac{1}{6}9e^{3s} + \frac{1}{3}25e^{5s}\Big|_{s=0} = \frac{71}{6}. \end{aligned}$$

例 5.6 指数分布 $\mathcal{E}(\lambda)$ 连续型随机变量 X 有概率密度函数 $f(x) = \lambda e^{-\lambda x} I_{\{x\geq 0\}}$, 求其矩母函数 $M(s)$, 期望 $\mathbb{E}X$, 二阶矩 $\mathbb{E}X^2$.

解 当 $s < \lambda$ 时, 有

$$M(s) = \lambda \int_0^\infty e^{sx} e^{-\lambda x} dx = \lambda \frac{e^{(s-\lambda)x}}{s-\lambda}\Big|_0^\infty = \frac{\lambda}{\lambda - s}.$$

当 $s \geq \lambda$ 时, 上积分为无穷. 当 $s < \lambda$ 时,

$$\begin{aligned} \mathbb{E}X &= M^{(1)}(0) = \frac{\lambda}{(\lambda-s)^2}\Big|_{s=0} = \frac{1}{\lambda}, \\ \mathbb{E}X^2 &= M^{(2)}(0) = \frac{2\lambda}{(\lambda-s)^3}\Big|_{s=0} = \frac{2}{\lambda^2}. \end{aligned}$$

例 5.7 混合分布 已知一银行有三位交易员, 每位交易员为顾客服务的时间皆服从指数分布. 两位快速交易员对应的参数为 $\lambda = 6$, 一位慢速交易员对应的参数为 $\lambda = 4$. 假设你来到银行等概率随机选择了一名交易员, 请求出你接受服务的时间的矩母函数.

解 设 X 为你接受服务的时间, 则 $f_X(x) = \left[\frac{2}{3}6e^{-6x} + \frac{1}{3}4e^{-4x}\right] I_{\{x\geq 0\}}$. 当 $s < 4$ 时, 有

$$M_X(s) = \int_0^\infty e^{sx}\left[\frac{2}{3}6e^{-6x} + \frac{1}{3}4e^{-4x}\right]dx = \frac{2}{3}\frac{6}{6-s} + \frac{1}{3}\frac{4}{4-s}.$$

反演: 若已知 X 的矩母函数形如 $p_1\frac{\lambda_1}{\lambda_1 - s} + p_2\frac{\lambda_2}{\lambda_2 - s}$, 则 X 是 X_1, X_2 的混合变量, 其中 $X_1 \sim \mathcal{E}(\lambda_1), X_2 \sim \mathcal{E}(\lambda_2)$, 且这两个变量被选中的概率分别为 p_1, p_2 相应密度函数为 $f_X(x) = p_1 f_{X_1}(x) + p_2 f_{X_2}(x)$.

例 5.8 正态分布 设连续型随机变量 $Z = X + Y$, 其中 $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, 且 X 和 Y 相互独立. 求 Z 的概率密度函数.

解

1. 若 $V \sim \mathcal{N}(0,1)$, 则由定义可计算得到 $M_V(s) = e^{s^2/2}$.
2. 若 $W \sim \mathcal{N}(\mu, \sigma^2)$, 则由矩母函数性质 1 得到 $M_W(s) = e^{\mu s + (\sigma^2 s^2/2)}$.
3. 进而由性质 5 得 $M_Z(s) = M_X(s)M_Y(s) = e^{(\mu_1+\mu_2)s + \frac{(\sigma_1^2+\sigma_2^2)s^2}{2}}$.
4. 可见 $M_Z(s)$ 的形式符合正态分布对应的矩母函数, 由性质 4 推断 $Z \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$ 从而可直接得其概率密度函数.

$\boxed{\text{独立的正态分布之和, 依然是正态分布!}}$

随机数个相互独立的随机变量之和

设随机变量 $X_1, X_2, \ldots$ 独立同分布, 有共同的矩母函数 $M_X(s)$, N 为一个取正整数值的随机变量, 且独立于该随机变量列.

令 $W = X_1 + \cdots + X_N$. 求 $M_W(s)$.

$$\begin{aligned}\mathbb{E}(e^{sW}|N=n) &= \mathbb{E}(e^{sX_1}\cdots e^{sX_N}|N=n)\\ &= \mathbb{E}(e^{sX_1}\cdots e^{sX_N}) = \mathbb{E}(e^{sX_1})\cdots\mathbb{E}(e^{sX_N})\\ &= (M_X(s))^n.\end{aligned}$$

$$\begin{aligned}M_W(s) = \mathbb{E}(e^{sW}) = \mathbb{E}[\mathbb{E}(e^{sW}|N=n)] &= \mathbb{E}\left[(M_X(s))^N\right]\\ &= \sum_{n=1}^{\infty}(M_X(s))^n\mathbb{P}(N=n).\end{aligned}$$

对比 $M_N(s) = \mathbb{E}(e^{sN}) = \sum_{n=1}^{\infty}(e^s)^n\mathbb{P}(N=n)$.

例 5.9 小明为了买《概率论》这本书跑到一个满是书店的街道. 每家店有此书的概率皆为 p, 且与其它店相互独立. 小明逛每个店只找这本书, 找到就走, 或者这家店肯定没有他才走. 他一直逛下去直到买到此书. 假设在每个店内他花费的时间都是一个随机变量, 服从参数为 λ 的指数分布, 并且与其他任何事情都独立. 求小明逛书店的总时间的分布.

解 $X_i \sim \mathcal{E}(\lambda), N \sim G(p), W = X_1 + \cdots + X_N$. 当 $s < \lambda$ 时, 有

$$M_{X_i}(s) = \frac{\lambda}{\lambda - s}.$$

而 $M_N(s) = \dfrac{pe^s}{1-(1-p)e^s}$. 于是

$$M_W(s) = \frac{pM_X(s)}{1-(1-p)M_X(s)} = \frac{p\lambda}{p\lambda - s}.$$

即 $Y \sim \mathcal{E}(p\lambda)$.

5.2.2 随机向量的矩母函数

定义 5.4 (多元矩母函数)

设 $\vec{X} = \{X_1, \ldots, X_n\}'$ 是一个随机向量, 则对应的多元矩母函数可定义为

$$M_{\vec{X}}(\vec{s}) = \mathbb{E}\left(e^{\vec{s}'\vec{X}}\right) = \mathbb{E}\left(e^{s_1X_1+\cdots+s_nX_n}\right),$$

其中 $\vec{s} = (s_1, \ldots, s_n)' \in \mathbb{R}^n$.

♣

5.3 特征函数

定义 5.5 (characteristic function (C.F.) 特征函数)

对随机变量 X, 称

$$\phi(t) := \mathbb{E}\left(e^{itX}\right) = \mathbb{E}\cos(tX) + i\mathbb{E}\sin(tX), t \in \mathbb{R},$$

为 X 的特征函数 (characteristic function), 其中 $i = \sqrt{-1}$.

♣

定理 5.5 (特征函数的性质)

设 $\phi(t)=\mathbb{E}(e^{itX})$, 则

1. $|\phi(t)|\le\phi(0)=1,\phi(-t)=\overline{\phi(t)}$;
2. $\phi(t)$ 在 $(-\infty,\infty)$ 上一致连续 $(|\phi(t+h)-\phi(t)|\le\mathbb{E}|e^{ihX}-1|)$;
3. 如果 $\mathbb{E}(|X|^k)<\infty$, 则
$$\phi^{(k)}(t)=i^k\mathbb{E}(X^ke^{itX}),\phi^{(k)}(0)=i^k\mathbb{E}(X^k).$$
4. 对任意常数 a,b, 有
$$\phi_{aX+b}(t)=e^{itb}\phi_X(at).$$
5. 如果 X_k 有特征函数 $\phi_k(t)$, 且 $X_1,\dots,X_n$ 相互独立, 则 $Y=X_1+\cdots+X_n$ 的特征函数为
$$\phi_Y(t)=\prod_{k=1}^n\phi_k(t).$$

5.3.1 常见分布的特征函数

例 5.10 二项分布 $B(n,p)$ 二项分布 $B(n,p)$ 的特征函数是
$$\phi(t)=\mathbb{E}\left(e^{itX}\right)=\sum_{j=0}^n e^{itj}\binom{n}{j}p^jq^{n-j}=\left(q+pe^{it}\right)^n.$$

例 5.11 Poisson 分布 $\mathcal{P}(\lambda)$ Poisson 分布 $\mathcal{P}(\lambda)$ 的特征函数是
$$\phi(t)=\mathbb{E}\left(e^{itX}\right)=\sum_{k=0}^n e^{itk}\frac{\lambda^k}{k!}e^{-\lambda}=e^{\lambda(e^{it}-1)}.$$

例 5.12 几何分布 $G(p)$ 几何分布 $G(p)$ 的特征函数是
$$\phi(t)=\mathbb{E}\left(e^{itX}\right)=\sum_{k=1}^{+\infty}e^{itk}pq^{k-1}=\frac{pe^{it}}{1-qe^{it}}.$$

例 5.13 均匀分布 $U(a,b)$ 均匀分布 $U(a,b)$ 有特征函数
$$\phi(t)=\frac{e^{itb}-e^{ita}}{it(b-a)}.$$
特别地, $U(-c,c)$ 有特征函数 $\dfrac{\sin(ct)}{(ct)}$.

例 5.14 指数分布 $\mathcal{E}(\lambda)$ 指数分布 $\mathcal{E}(\lambda)$ 有特征函数
$$\phi(t)=\left(1-\frac{it}{\lambda}\right)^{-1}.$$

例 5.15 正态分布 $\mathcal{N}(\mu,\sigma^2)$ 正态分布 $\mathcal{N}(\mu,\sigma^2)$ 有特征函数
$$\phi(t)=\exp\left(i\mu t-\frac{1}{2}\sigma^2t^2\right).$$

解 (i) 先求 $\mathcal{N}(0,1)$ 的特征函数.

【方法 1: 形式运算】将 i 视为常数, 形式地运算得到
$$\begin{aligned}\phi(t)=\mathbb{E}\left(e^{itX}\right)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{itx}e^{-x^2/2}dx\\&=e^{-t^2/2}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-it)^2}{2}}dx=e^{-t^2/2}\end{aligned}$$

【方法 2: 严格的数学推导】

$$\begin{aligned}\phi(t) = \mathbb{E}\left(e^{itX}\right) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(tx)e^{-x^2/2}dx\end{aligned}$$

得

$$\begin{aligned}\phi'(t) &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x\sin(tx)e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sin(tx)de^{-x^2/2} \\ &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t\cos(tx)e^{-x^2/2}dx \\ &= -t\phi(t).\end{aligned}$$

即:

$$\frac{d}{dt}\left[\phi(t)e^{t^2/2}\right] = \left[\phi'(t) + t\phi(t)\right]e^{t^2/2} = 0$$

得 $\phi(t)e^{t^2/2} = C$. 因为 $C = \phi(0) = 1$ 得到

$$\phi(t) = e^{-t^2/2}.$$

(ii) 现设 $Y \sim \mathcal{N}(\mu, \sigma^2)$, 则

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

因此,

$$\mathbb{E}\left(e^{itY}\right) = \mathbb{E}\left(e^{it(\mu+\sigma X)}\right) = e^{it\mu}\mathbb{E}\left(e^{it\sigma X}\right) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$$

设 $X_1, \ldots, X_m$ 相互独立, $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, 则

$$Y = X_1 + \cdots + X_m \sim \mathcal{N}\left(\sum_{j=1}^m \mu_j, \sum_{j=1}^m \sigma_j^2\right).$$

例 5.16 Cauchy 分布 Cauchy 分布, 其密度 $f(x) = \dfrac{1}{\pi(1+x^2)}, x \in \mathbb{R}$, 有特征函数

$$\phi(t) = e^{-|t|}.$$

例 5.17 Laplace 分布 Laplace 分布, 其密度 $f(x) = \dfrac{1}{2}e^{-|x|}$, 有特征函数

$$\phi(t) = \frac{1}{1+t^2}.$$

特征函数与分布函数

$\boxed{\text{随机变量的特征函数和分布函数相互唯一决定}}$

定理 5.6 (逆转公式)

设 $\phi(t)$ 是 X 的特征函数, $F(x)$ 是 X 的分布函数. 如果 $F(x)$ 在 a, b 连续, 则

$$\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi(t)dt = F(b) - F(a).$$

♡

定理 5.7

如果 X 的特征函数满足 $\int_{\mathbb{R}} |\phi(t)|dt < \infty$, 则 X 有连续密度函数

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

♡

定理 5.8

如果 $\mathbb{E}|X|^n < \infty$, 则 X 的特征函数 $\phi(t)$ 满足

$$\phi(t) = \sum_{m=0}^{n} \frac{\mathbb{E}\left[(itX)^m\right]}{m!} + o(t^n).$$

特别地, 如果 $\mathbb{E}(X^2) < \infty$, 则

$$\phi(t) = 1 + it\mathbb{E}(X) - \frac{1}{2} t^2 \mathbb{E}(X^2) + o(t^2).$$

♡

5.3.2 混合分布的特征函数

定理 5.9 (混合分布)

设分布函数 $F_1(x), \dots, F_m(x)$ 的特征函数分别为 $\phi_1(t), \dots, \phi_m(t)$, $\lambda_i \geq 0$, 且 $\sum_{i=1}^{m} \lambda_i = 1$. 则 $F(x) = \sum_{i=1}^{m} \lambda_i F_i(x)$ 的特征函数为 $\phi(t) = \sum_{i=1}^{m} \lambda_i \phi_i(t)$.

♡

注 $Y = \sum_{i=1} \lambda_i F_i$ 表示 Y 以概率 λ_i 等于 X_i, 对 $\{X_k, k = 1, \dots, m\}$ 之间的独立性无要求。此时 $Y \neq \sum_{i=1}^{m} \lambda_i X_i$.

5.3.3 随机向量的特征函数

定义 5.6 (随机向量的特征函数)

设 $X = (X_1, \dots, X_n)$ 是随机向量, X 的特征函数定义为

$$\phi(t) = \mathbb{E}\left(e^{it'X}\right), t = (t_1, \dots, t_n) \in \mathbb{R}^n.$$

♣

定理 5.10

设 $X = (X_1, \dots, X_n)$ 是随机向量, 则 $X_1, \dots, X_n$ 相互独立的充要条件是

$$\phi(t) = \phi_1(t_1)\phi_2(t_2)\dots\phi_n(t_n),$$

其中 $\phi_k(t_k)$ 是 X_k 的特征函数, $k = 1, 2, \dots, n$.

♡

例 5.18 设 X 是 Cauchy 分布, 其有密度函数 $f(x)=\dfrac{1}{\pi(1+x^2)}, x\in\mathbb{R}$. 令 $Y=aX,(a>0)$. 证明 $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)$.

证明 首先 $\phi_X(t)=e^{-|t|}$. 故

$$\phi_Y(t)=\mathbb{E}\left(e^{itY}\right)=\mathbb{E}\left(e^{i(at)X}\right)=e^{-|at|}=e^{-a|t|}$$

和

$$\begin{aligned}\phi_{X+Y}(t)&=\mathbb{E}\left(e^{it(X+Y)}\right)=\mathbb{E}\left(e^{it(1+a)X}\right)\\&=e^{-|(1+a)t|}=e^{-(1+a)|t|}.\end{aligned}$$

因此

$$\phi_{X+Y}(t)=e^{-(1+a)|t|}=e^{-|t|}\cdot e^{-a|t|}=\phi_X(t)\phi_Y(t).$$

此例说明: 尽管 $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)$, 但是推不出他们对应的随机变量是独立的.

5.4 大数定律

定理 5.11 (Weak law of large numbers, WLLN)

设 $\{X_n\}$ 是独立同分布的随机变量序列, 且 $\mathrm{Var}(X_1)<\infty$, 则对任意给定的 $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i-EX_1\right|\geq\varepsilon\right)=0,$$

也称 $\frac{1}{n}\sum_{k=1}^n X_k$ 依概率收敛到 EX_1, 记作

$$\frac{1}{n}\sum_{k=1}^n X_k\xrightarrow{p}EX_1.$$

♡

证明 设 $\{X_n\}$ 是独立同分布的随机序列, 且 $\mathrm{Var}(X_1)<\infty$.

往证: 对 $\forall\varepsilon>0$, $\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i-EX_1\right|\geq\varepsilon\right)=0$.

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i-EX_1\right|\geq\varepsilon\right)&=\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i-E\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right|\geq\varepsilon\right)\\&\leq\frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)}{\varepsilon^2}=\frac{\mathrm{Var}(X_1)}{n\varepsilon^2}\to 0, n\to+\infty\end{aligned}$$

Recall: 切比雪夫不等式

对随机变量 X 和 $\varepsilon>0$, 有

$$\mathbb{P}(|X-EX|\geq\varepsilon)\leq\frac{\mathrm{Var}(X)}{\varepsilon^2}$$

定理 5.12 (辛钦大数定律)

设 $\{X_n\}$ 是 $i.i.d.$, 且 $\mathbb{E}|X_1|<\infty$, 则

$$\frac{1}{n}\sum_{k=1}^n X_k\xrightarrow{p}EX_1$$

♡

定理 5.13 (WLLN 扩展 b)

设 $\{X_n\}$ 是独立同分布的随机变量序列, 则存在实数列 $\{a_n\}$ 使得

$$\frac{1}{n}\sum_{k=1}^{n} X_k - a_n \xrightarrow{p} 0$$

的充分必要条件是

$$n\mathbb{P}(|X_1| \geq n) \to 0$$

此时, 可取 $a_n = \mathbb{E}(X_1 I_{\{|X_1|<n\}})$.

定理 5.14 (Strong law of large numbers, SLLN)

设 $\{X_n\}$ 是独立同分布的随机变量序列:

(1) 若 $\mathbb{E}|X_1| < \infty$, 则

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} EX_1 \text{ 几乎处处收敛;}$$

(2) 反之, 若

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} C \text{ 几乎处处收敛,}$$

则 $\mathbb{E}|X_1| < \infty$, 且 $C = EX_1$.

注 几乎处处收敛, 以概率 1 收敛: $\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = EX_1\right) = 1$

例 5.19 经验分布 设 $\{X_j\}$ 是概率空间 $(\Omega, \mathcal{F}, \mathbb{P})$ 上独立同分布的随机序列, 用 x_j 表示 X_j 的观测值, 即对某个确定的 $\omega \in \Omega$, 有

$$x_j = X_j(\omega), \quad j = 1, 2, \ldots$$

则以概率 1, 观测数列 $\{x_j\}$ 可以决定 X_j 的分布函数 $F(x)$.

证明 对任何确定的 $x \in \mathcal{R}$, 定义

$$g(X_j) = \begin{cases} 1, & \text{if } X_j \leq x, \\ 0, & \text{if } X_j > x. \end{cases}$$

$j = 1, 2, \ldots$. 则 $\{g(X_j)\}$ 是独立同分布的随机序列. 由 SLLN,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} g(X_j) = Eg(X_1) = \mathbb{P}(X_1 \leq x) = F(x) \quad a.s.$$

即以概率 1 有 $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} g(x_j) = F(x)$.

记 $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)$ (经验分布函数).

例 5.20 Monte Carlo 蒙特卡洛 设需估计某复杂有界函数 $f(x)$ 在区间 $[a, b]$ 上的积分 $\int_a^b f(x)dx$.

解 从 A 中随机抽取 n 个点 $(X_1, Y_1), \ldots, (X_n, Y_n)$. 用 I_j 表示 (X_j, Y_j) 落入 B 否. 则

$$\text{以概率 1 有 } \frac{1}{n}\sum_{j=1}^{n} I_j \to EI_j = \mathbb{P}(I_j = 1) = \frac{\int_a^b f(x)dx}{c(b-a)} \quad I_j = 0$$

$$\int_a^b f(x)dx \approx c(b-a)\frac{1}{n}\sum_{j=1}^{n} I_j.$$

5.5 收敛性的一些定义

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间, X_n 和 X 是随机变量, 其分布函数分别为 $F_n(x)$ 和 $F(x)$, 即.

$$F_n(x) = \mathbb{P}(X_n \le x), \quad F(x) = \mathbb{P}(X \le x).$$

定义 5.7 (数列的收敛)

设 $\{a_i\}$ 是实数列, a 为一实数, 如果对 $\forall \varepsilon > 0$, 存在正整数 N, 使得当 $n \ge N$ 时, 有

$$|a_n - a| \le \varepsilon,$$

则称数列 $|a_n|$ 收敛于 a, 记作 $\lim_{n\to\infty} a_n = a$. ♣

定义 5.8 (convergence in distribution)

如果在 $F(x)$ 的连续点 x, 有

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

则称 X_n 依分布收敛到 X (convergence in distribution), 记作 $X_n \xrightarrow{d} X$, 或者称 F_n 弱收敛到 F (weak convergence), 记作 $F_n \to F$. ♣

定义 5.9 (almost sure convergence)

如果

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1,$$

则称 X_n 几乎处处收敛到 X (almost sure convergence), 记作 $X_n \xrightarrow{a.s.} X$, 或 $X_n \to X$ a.s. ♣

对 $\forall \varepsilon > 0$,

$$\begin{aligned}\left\{\lim_{n\to\infty} X_n = X\right\} &= \left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} \\ &\to \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{\omega : |X_k(\omega) - X(\omega)| < \varepsilon\} \\ &= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \{|X_k - X| < \varepsilon\} = \liminf_{n\to\infty} A_n,\end{aligned}$$

with $A_n = \{|X_n - X| < \varepsilon\}$.

注 $\lim_{n\to\infty} a_n = a \Leftrightarrow \forall \varepsilon > 0, \exists n, \text{s.t.} \forall k \ge n, |a_k - a| \le \varepsilon$

定义 5.10 (convergence in probability)

如果对 $\forall \varepsilon > 0$, 有

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \ge \varepsilon) = 0,$$

则称 X_n 依概率收敛到 X (convergence in probability), 记作 $X_n \xrightarrow{p} X$, 或 $X_n \to X$ in prob. ♣

对 $\forall \varepsilon > 0$, denote $A_n = \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\} = \{|X_n - X| < \varepsilon\}$.
则

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \ge \varepsilon) = \lim_{n\to\infty} \mathbb{P}(A_n^c) = \lim_{n\to\infty} [1 - \mathbb{P}(A_n)] = 1 - \lim_{n\to\infty} \mathbb{P}(A_n).$$

故 $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \ge \varepsilon) = 0 \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(A_n) = 1$.

定义 5.11 (L_p convergence)

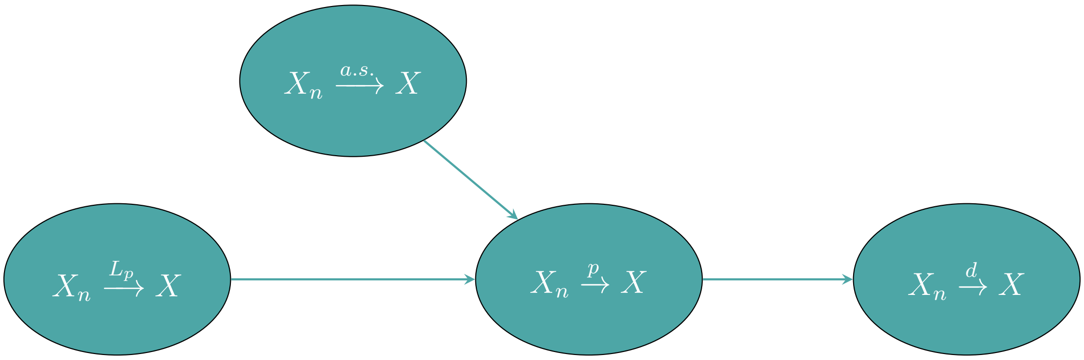
对 $p>0$, 如果

$$\lim_{n\to\infty}\mathbb{E}(|X_n-X|^p)=0,$$

则称 X_n 在 L_p 下收敛到 X, 记作 $X_n\xrightarrow{L_p}X$, 或 $X_n\to X$ in L_p.

例如, $\lim_{n\to\infty}\mathbb{E}(|X_n-X|)=0\Leftrightarrow X_n\xrightarrow{L_1}X$.

5.6 各种收敛之间的关系



定理 5.15 (a.s. implies in prob.)

如果 $X_n\xrightarrow{a.s.}X$, 则 $X_n\xrightarrow{p}X$.

证明 注意到, 对 $\forall\varepsilon>0$,

$$\left\{\lim_{n\to\infty}X_n=X\right\}\subset\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k-X|<\varepsilon\}$$

因此, 利用概率的连续性可得

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}(|X_n-X|\geq\varepsilon)&\leq\lim_{n\to\infty}\mathbb{P}(\cup_{k=n}^{\infty}|X_k-X|\geq\varepsilon)\\&=\mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}\{|X_k-X|\geq\varepsilon\}\right)\\&=1-\mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\{|X_k-X|<\varepsilon\}\right)=1-1=0\end{aligned}$$

定理 5.16 (L_p implies in prob.)

如果 $X_n\xrightarrow{L_p}X$, 则 $X_n\xrightarrow{p}X$.

证明 对 $\forall\varepsilon>0$, 利用 Markov 不等式可得.

$$\lim_{n\to\infty}\mathbb{P}(|X_n-X|\geq\varepsilon)\leq\lim_{n\to\infty}\frac{1}{\varepsilon^p}\mathbb{E}(|X_n-X|^p)=0.$$

定理 5.17 (in prob. implies in dist.)

如果 $X_n\xrightarrow{p}X$, 则 $X_n\xrightarrow{d}X$.

证明 令 $F_n(x)=\mathbb{P}(X_n\le x)$ 和 $F(x)=\mathbb{P}(X\le x)$. 首先,

$$\begin{aligned}F_n(x)&=\mathbb{P}(X_n\le x,|X_n-X|\le\varepsilon)+\mathbb{P}(X_n\le x,|X_n-X|>\varepsilon)\\&\le\mathbb{P}(X\le x+\varepsilon)+\mathbb{P}(|X_n-X|>\varepsilon)\\&=F(x+\varepsilon)+\mathbb{P}(|X_n-X|>\varepsilon)\end{aligned}$$

另一方面,

$$\begin{aligned}F_n(x)&=1-\mathbb{P}(X_n>x)\\&=1-\mathbb{P}(X_n>x,|X_n-X|\le\varepsilon)-\mathbb{P}(X_n>x,|X_n-X|>\varepsilon)\\&\ge1-\mathbb{P}(X>x-(X_n-X),|X_n-X|\le\varepsilon)-\mathbb{P}(|X_n-X|>\varepsilon)\\&\ge1-\mathbb{P}(X>x-\varepsilon)-\mathbb{P}(|X_n-X|>\varepsilon)\\&=F(x-\varepsilon)-\mathbb{P}(|X_n-X|>\varepsilon)\end{aligned}$$

因此,

$$F(x-\varepsilon)-\mathbb{P}(|X_n-X|>\varepsilon)\le F_n(x)\le F(x+\varepsilon)+\mathbb{P}(|X_n-X|>\varepsilon)$$

令 $n\to\infty$,

$$F(x-\varepsilon)\le\liminf_n F_n(x)\le\limsup_n F_n(x)\le F(x+\varepsilon).$$

如果 $F(x)$ 在 x 处连续, 则当 $\varepsilon\downarrow0$, 有 $F(x-\varepsilon)\uparrow F(x)$ 和 $F(x+\varepsilon)\downarrow F(x)$. 所以

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

即, $X_n\xrightarrow{d}X$.

一些反例

例 5.21

$$X_n\xrightarrow{d}X\not\Rightarrow X_n\xrightarrow{p}X$$

解 $\{X_n\}$ 独立同分布 $\sim\mathcal{N}(0,1)$, 则 $X_n\xrightarrow{d}X_1$. 但 $n\ne1$ 时, $X_n-X_1\sim\mathcal{N}(0,2)$, 即 $\dfrac{X_n-X_1}{\sqrt2}\sim\mathcal{N}(0,1)$, 有 $\dfrac{X_n-X_1}{\sqrt2}$ 与 X_1 同分布, 从而

$$\mathbb{P}(|X_n-X_1|>1)=\mathbb{P}\left(\left|\frac{X_n-X_1}{\sqrt2}\right|>\frac{1}{\sqrt2}\right)=\mathbb{P}\left(|X_1|>\frac{1}{\sqrt2}\right)>0$$

例 5.22

$$X_n\xrightarrow{d}C\Leftrightarrow X_n\xrightarrow{p}C,\ \text{其中}C\ \text{为常数}.$$

例 5.23

$$X_n\xrightarrow{p}X\not\Rightarrow X_n\xrightarrow{L_p}X$$

解 取概率空间 $(\Omega,\mathcal{F},\mathbb{P})=((0,1),\mathcal{B}((0,1)),\lambda)$, $\lambda(A)=length(A)$. 即一维几何概型. 定义

$$X_n(\omega)=\begin{cases}n, & \text{if }\omega\in(0,n^{-p}],\\0, & \text{if }\omega\in(n^{-p},1),\end{cases}\qquad(1)$$

且 $X\equiv0$. 则, 对 $\forall\varepsilon>0$, 当 $n\to\infty$,

$$\begin{aligned}\mathbb{P}(|X_n-0|>\varepsilon)&=\mathbb{P}(X_n>\varepsilon)=\mathbb{P}(X_n=n)\\&=\mathbb{P}(\{\omega:\omega\in(0,n^{-p}]\})=\mathbb{P}((0,n^{-p}])=n^{-p}\\&\to0\end{aligned}$$

但是 $\mathbb{E}(|X_n - 0|^p) = 1$.

例 5.24

$$X_n \xrightarrow{L_p} X \nRightarrow X_n \xrightarrow{L_q} X \quad (p < q).$$

例 5.25

$$X_n \xrightarrow{a.s.} X \nRightarrow X_n \xrightarrow{L_p} X$$

解 X_n 的定义见上页 (1). 对 $\forall \omega \in \Omega$, 都有 $X_n(\omega) \to 0 \equiv X$. 但是 $\mathbb{E}(|X_n - 0|^p) = 1$.

例 5.26

$$X_n \xrightarrow{L_p} X \nRightarrow X_n \xrightarrow{a.s.} X. \quad (thus \quad X_n \xrightarrow{p} X \nRightarrow X_n \xrightarrow{a.s.} X.)$$

解 假设 $\{X_n\}$ 独立, 且 X_n 定义如下:

$$\mathbb{P}(X_n = 0) = 1 - n^{-1}, \quad \mathbb{P}(X_n = 1) = n^{-1}.$$

则 $X_n \xrightarrow{L_1} 0$. 但是 $X_n \xrightarrow{a.s.} 0$. 令 $A_n = \{X_n = 1\}$, 则

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

由 Borel-Cantelli 引理, "$\{A_n\}$ 有无穷多个发生" 的概率为 1.

5.7 常见的收敛性相关定理

定理 5.18 (Continuous mapping theorem)

设 $\{X, X_n\}$ 是随机元序列, g 连续, 则

1. $X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$;
2. $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$;
3. $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$.

♡

但是 $X_n \xrightarrow{L_p} X \nRightarrow g(X_n) \xrightarrow{L_p} g(X)$.

定理 5.19 (Slutsky's theorem)

假设 $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} c$, 则

1. $X_n + Y_n \xrightarrow{d} X + c$;
2. $X_n Y_n \xrightarrow{d} cX$;
3. $X_n / Y_n \xrightarrow{d} X/c$, $(c \neq 0)$.

♡

定理 5.20 (Delta Method)

假设 $\sqrt{n}(X_n - a) \xrightarrow{d} \mathcal{N}(0, V)$, $g : \mathbb{R} \to \mathbb{R}$ 是一个连续可导函数, 则

$$\sqrt{n}(g(X_n) - g(a)) \xrightarrow{d} \mathcal{N}(0, (g'(a))^2 \cdot V);$$

♡

定理 5.21 (Continuity theorem, 连续性定理)

设 X_n 的特征函数为 $\phi_n(t)$. X 的特征函数为 $\phi(t)$. 则 X_n 依分布收敛到 X 的充分必要条件是,

$$\lim_{n\to\infty} \phi_n(t) = \phi(t), \quad \forall t \in R$$

♡

此定理是概率论中最常用、最重要的定理之一.

定理 5.22

设随机向量 X_n 的特征函数 $\phi_n(t)$ 收敛到在 $t=0$ 处连续的函数 $\phi(t)$, 则 $\phi(t)$ 是某个随机向量 X 的特征函数, 并且对任何常数向量 a, 有 $a^T X_n \xrightarrow{d} a^T X$. ♡

推论 5.1

设随机变量 X_n 的特征函数 $\phi_n(t)$ 收敛到在 $t=0$ 处连续的函数 $\phi(t)$, 则 $\phi(t)$ 是某个随机变量 X 的特征函数, 并且有 $X_n \xrightarrow{d} X$. ♡

判定依分布收敛, 可借助特征函数的收敛性

定理 5.23

设随机向量序列 $\{X_n, n=1,2\ldots\}$ 和随机向量 X 的维度一致, 则 $X_n \xrightarrow{d} X$ 的充分必要条件是对任何常数向量 a, 有 $a^T x_n \xrightarrow{d} a^T X$. ♡

引理 5.1 (Cramér-Wold)

设 X, Y 都是 p-维随机向量. 则 X, Y 同分布的充分必要条件是对任何常数向量 $a \in \mathbb{R}^p$, 有 $a^T X, a^T Y$ 同分布. ♡

向量问题都可转化为一维

5.8 中心极限定理

定理 5.24 (Lindeberg–Lévy 中心极限定理)

设 $\{X_n\}$ 是独立同分布的随机变量序列, 其期望为 μ, 方差为 $\sigma^2 < \infty$, 记 $S_n = X_1 + \cdots + X_n$, 则

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

♡

n 充分大时提供近似分布:

$$S_n \approx \mathcal{N}\left(E(S_n), \mathrm{Var}(S_n)\right) \approx \mathcal{N}\left(n\mu, n\sigma^2\right)$$

定理 (Lindeberg–Lévy 中心极限定理 - 样本均值形式)

设 $\{X_n\}$ 是独立同分布的随机变量序列, 其期望为 μ, 方差为 $\sigma^2 < \infty$, 记 $M_n = (X_1 + \cdots + X_n)/n$, 则

$$\frac{M_n - E(M_n)}{\sqrt{\mathrm{Var}(M_n)}} \xrightarrow{d} \mathcal{N}(0,1).$$

♡

n 充分大时提供近似分布:

$$M_n \approx \mathcal{N}\left(E(M_n), \mathrm{Var}(M_n)\right) \approx \mathcal{N}\left(\mu, \sigma^2/n\right)$$

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{(X_k - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k$$

证明

令 $Y_k = (X_k - \mu)/\sigma$. 则 $\{Y_k\}$ 独立同分布, 其期望为 0, 方差为 1. 用 $\phi(t) = \mathbb{E}e^{itY_1}$ 表示 Y_1 的特征函数. 注意到

$$\phi'(0) = i\mathbb{E}Y_1 = 0, \quad \phi''(0) = i^2 \mathbb{E}Y_1^2 = -1.$$

利用 Taylor 展开, 可得

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + o(t^2).$$

所以 $\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k-\mu)$ 的特征函数为

$$\begin{aligned}\phi_n(t) &= \mathbb{E}\exp(it(Y_1+\cdots+Y_n)/\sqrt{n})\\ &= [\phi(t/\sqrt{n})]^n = \left(1-\frac{t^2}{2n}+o(t^2/n)\right)^n\\ &\to e^{-t^2/2}, \quad (n\to\infty)\end{aligned}$$

由连续性定理 (Continuity theorem) 可知, 结论成立.

例 5.27 设人群中疫苗有效的比例为 μ. 随机抽 n 个人. X_i 表示疫苗对第 i 个人是否有效. 通过样本中有效比例 $M_n = (X_1+\cdots+X_n)/n$ 认识 μ, 需多少样本量能以 95% 概率保证误差率小于 3%?

解 求适当的 n 使 $\mathbb{P}(|M_n-\mu|<0.03)\geq 0.95$.

$$\begin{aligned}&= \mathbb{P}\left(\left|\frac{M_n-\mu}{\sigma/\sqrt{n}}\right| < \frac{0.03}{\sigma/\sqrt{n}}\right) \quad \sigma=\sqrt{\mu(1-\mu)}\\ &\approx \mathbb{P}\left(|Z|<\frac{0.03\sqrt{n}}{\sigma}\right) \quad Z\sim\mathcal{N}(0,1)\\ &\geq \mathbb{P}(|Z|<0.06\sqrt{n}) = 0.95 \quad \sigma\leq\frac{1}{2}\end{aligned}$$

取 $n=\left(\frac{1.96}{0.06}\right)^2\approx 1068$ 即可.

定理 5.25 (Lindeberg-Feller, CLT (选修))

设 $\{X_n\}$ 是独立的 r.v.s., 则: 方差序列 $\{\sigma_k^2 := var(X_k)\}$ 满足

$$B_n^2 = \sum_{k=1}^{n} var(X_k) \to \infty, \quad \frac{\sigma_n^2}{B_n^2}\to 0 \quad (3)$$

并且中心极限定理

$$\frac{1}{B_n}\sum_{k=1}^{n}(X_k-EX_k) \xrightarrow{d} \mathcal{N}(0,1)$$

成立的充分必要条件是 Lindeberg 条件成立, 即: 对 $\forall\varepsilon>0$,

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\mathbb{E}\{(X_k-EX_k)^2 I_{\{|X_k-EX_k|>\varepsilon B_n\}}\} = 0.$$

♡

条件 (3) $\Leftrightarrow \lim_{n\to\infty}\max_{1\leq k\leq n}\frac{\sigma_k^2}{B_n^2}=0$ (Feller 条件)

5.8.1 中心极限定理的离散修正

定理 5.26 (de Moivre-Laplace, CLT, 1716)

设 $S_n\sim B(n,p)$, n 充分大, k 和 m 是非负整数, 则

$$\mathbb{P}(k\leq S_n\leq m)\approx\Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

其修正形式【近似更为精确】:

$$\mathbb{P}(k\leq S_n\leq m)\approx\Phi\left(\frac{m+0.5-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right)$$

更重要的是: 修正形式可以计算单点概率 $\mathbb{P}(S_n = k)$.

该修正方法对其他离散型随机变量同样适用!

例 5.28 设 $S_n \sim B(36, 0.5)$, 求概率 $\mathbb{P}(S_n \le 21)$.

解 其精确的概率为:

$$\mathbb{P}(S_n \le 21) = \sum_{k=0}^{21} \binom{36}{k} 0.5^{36} = 0.8785.$$

利用中心极限定理, 若端点不经修正, 则近似概率为:

$$\mathbb{P}(S_n \le 21) \approx \Phi\left(\frac{21 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{21 - 18}{3}\right) = 0.8413.$$

修正之后的近似为:

$$\mathbb{P}(S_n \le 21) \approx \Phi\left(\frac{21.5 - np}{\sqrt{np(1-p)}}\right) = \Phi(1.17) = 0.8789995.$$

另: $\mathbb{P}(S_n = 19) = 0.1251$, 正态近似 $\mathbb{P}(S_n = 19) \approx 0.124$

5.9 多元正态分布

定义 5.12 (Multivariate normal distribution (MVN) 定义)

当 Σ 正定时, $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$ 有联合密度函数

$$f(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{det(\Sigma)}} exp\left[-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})\right]$$

$\vec{X}$ 的特征函数

$$\phi_{\vec{X}}(\vec{t}) = E\exp(i\vec{t}^T \vec{X}) = \exp\left[i\vec{t}^T \vec{\mu} - \frac{1}{2}\vec{t}^T \Sigma \vec{t}\right]$$

从中可以看出, $\vec{X}$ 的期望和协方差矩阵唯一决定了 $\vec{X}$ 的特征函数. 由于随机向量的特征函数与分布函数是相互唯一决定的, 所以 $\vec{X}$ 的分布由 $\vec{\mu}$ 和 Σ 唯一决定.

定义 5.13 (multivariate normal distribution, MVN 等价定义)

设 $\vec{\mu} = (\mu_1, \mu_2, ..., \mu_n)^T$ 是 n 维常数列向量, $\mathbf{B}$ 是 $n \times m$ 常数矩阵, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_m$ 是相互独立且服从标准正态分布的随机变量. 如果

$$\vec{X} = \vec{\mu} + \mathbf{B}\vec{\varepsilon}$$

其中 $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)^T$, 且矩阵 $\mathbf{B}\mathbf{B}^T$ 满秩, 就称 $\vec{X} = (X_1, X_2, ..., X_n)^T$ 服从 n-维正态分布, 记作 $X \sim \mathcal{N}(\vec{\mu}, \mathbf{B}\mathbf{B}^T)$.

证明 因为 Σ 是正定的, 所以存在可逆方阵 B 使得 $\Sigma = BB^T$, 且 $\vec{X} = \vec{\mu} + B\vec{\varepsilon}$, 其中 $\vec{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. 易得 $\vec{\varepsilon}$ 的密度函数为

$$f_{\vec{\varepsilon}}(\vec{y}) = \frac{1}{(\sqrt{2\pi})^n} exp\left(-\frac{1}{2}\vec{y}^T \vec{y}\right)$$

注意到

$$\{\vec{X} = \vec{x}\} = \{\vec{\varepsilon} = B^{-1}(\vec{x} - \vec{\mu})\}$$

且下面的映射是可逆的

$$\vec{y} = B^{-1}(\vec{x} - \vec{\mu})$$

其雅克比行列式为

$$\left|\frac{\partial \vec{y}}{\partial \vec{x}}\right| = |B^{-1}| = \frac{1}{\sqrt{det(\Sigma)}}$$

故

$$\begin{aligned} f(\vec{x}) &= \frac{1}{(\sqrt{2\pi})^n} exp\left\{-\frac{1}{2}[B^{-1}(\vec{x}-\vec{\mu})]^T[B^{-1}(\vec{x}-\vec{\mu})]\right\}\left|\frac{\partial y}{\partial x}\right| \\ &= \frac{1}{(\sqrt{2\pi})^n\sqrt{det(\Sigma)}} exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T\Sigma^{-1}(\vec{x}-\vec{\mu})\right] \end{aligned}$$

$\vec{\varepsilon}$ 的特征函数

$$\begin{aligned} \phi_{\vec{\varepsilon}}(\vec{t}) &= E[\exp(i\vec{t}^T\vec{\varepsilon})] = E[\exp(it_1\varepsilon_1+\cdots+it_n\varepsilon_n)] = E\left[\prod_{i=1}^n \exp(it_i\varepsilon_i)\right] \\ &= \prod_{i=1}^n E[\exp(it_i\varepsilon_i)] = \prod_{j=1}^n \exp\left(\frac{-t_j^2}{2}\right) = \exp\left(-\frac{\vec{t}^T\vec{t}}{2}\right) \end{aligned}$$

其中 $\vec{t} = (t_1, t_2, ..., t_n)^T \in \mathbb{R}^n$.

$\vec{X}$ 的特征函数

$$\begin{aligned} \phi_{\vec{X}}(\vec{t}) &= E\exp(i\vec{t}^T\vec{X}) = E\exp[i(\vec{t}^T\vec{\mu} + t^T\mathbf{B}\vec{\varepsilon})] \\ &= \exp(i\vec{t}^T\vec{\mu})E\exp[i(\vec{t}^T\mathbf{B})\vec{\varepsilon}] \\ &= \exp\left[i\vec{t}^T\vec{\mu} - \frac{1}{2}(\vec{t}^T\mathbf{B})(\vec{t}^T\mathbf{B})^T\right] \\ &= \exp\left[i\vec{t}^T\vec{\mu} - \frac{1}{2}\vec{t}^T\mathbf{B}\mathbf{B}^T\vec{t}\right] \end{aligned}$$

定理 5.27

如果 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, 则对任意常数矩阵 $\mathbf{A}$ 和常数向量 $\vec{b}$, 只要 $\vec{b} + \mathbf{A}\vec{X}$ 有意义, $\mathbf{A}\Sigma\mathbf{A}^T$ 满秩, 则 $\vec{Y} = \vec{b} + \mathbf{A}\vec{X} \sim \mathcal{N}\left(\vec{b} + \mathbf{A}\vec{\mu}, \mathbf{A}\Sigma\mathbf{A}^T\right)$.

♡

证明 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma) \Rightarrow$ 存在 $\mathbf{B}$ 是 $n \times m$ 常数矩阵; $\varepsilon_1, \varepsilon_2, ..., \varepsilon_m$ 是相互独立且服从标准正态分布的随机变量, 使得 $\vec{X} = \vec{\mu} + \mathbf{B}\vec{\varepsilon}$, 其中 $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)^T$

$\vec{Y} = \vec{b} + \mathbf{A}\vec{X} = \left(\vec{b} + \mathbf{A}\vec{\mu}\right) + (\mathbf{AB})\vec{\varepsilon}$, 即 $\vec{Y}$ 服从多元正态分布. 计算其均值和协方差即可得.

定理 5.28 (重要判定法则)

$\vec{X} = (X_1, X_2, ..., X_n)^T \sim \mathcal{N}(\vec{\mu}, \Sigma)$ 的充要条件是对任何 $\vec{a} = (a_1, a_2, ..., a_n)^T \in \mathbb{R}^n$,

$$Y := \vec{a}^T\vec{X} \sim \mathcal{N}(\vec{a}^T\vec{\mu}, \vec{a}^T\Sigma\vec{a}).$$

♡

证明 $(\Rightarrow)$ Y 的特征函数

$$\phi_Y(t) = E\exp(itY) = E\exp[i(t\vec{a}^T)\vec{X}] = \exp\left[it(\vec{a}^T\vec{\mu}) - \frac{1}{2}t^2\vec{a}^T\Sigma\vec{a}\right] \quad (1)$$

所以 $Y \sim \mathcal{N}(\vec{a}^T\vec{\mu}, \vec{a}^T\Sigma\vec{a})$.

$(\Leftarrow)$ 在 (1) 中取 $t = 1$, 得

$$E\exp(i\vec{a}^T\vec{X}) = \phi_Y(1) = E\exp(iY) = \exp\left[i\vec{a}^T\vec{\mu} - \frac{1}{2}\vec{a}^T\Sigma\vec{a}\right].$$

故 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$.

定理 5.29 (独立性判定)

设 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, 如果

$$\vec{X} = \begin{pmatrix} \vec{X}_1 \\ \vec{X}_2 \end{pmatrix}, \quad \vec{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix},$$

且 $\vec{X}_1, \vec{\mu}_1$ 和方阵 Σ_{11} 的行数相同, 则 $\vec{X}_1$ 和 $\vec{X}_2$ 独立, 而且

$$\vec{X}_1 \sim \mathcal{N}(\vec{\mu}_1, \Sigma_{11}), \quad \vec{X}_2 \sim \mathcal{N}(\vec{\mu}_2, \Sigma_{22})$$

♡

证明 $\vec{X}$ 的特征函数

$$\begin{aligned}
\phi(\vec{t}) = \phi(\vec{t}_1, \vec{t}_2) &= \exp\left[i\vec{t}^T\vec{\mu} - \frac{1}{2}\vec{t}^T\Sigma\vec{t}\right] \\
&= \exp\left[i\vec{t}_1^T\vec{\mu}_1 - \frac{1}{2}\vec{t}_1^T\Sigma_{11}\vec{t}_1 + i\vec{t}_2^T\vec{\mu}_2 - \frac{1}{2}\vec{t}_2^T\Sigma_{22}\vec{t}_2\right] \\
&= \exp\left[i\vec{t}_1^T\vec{\mu}_1 - \frac{1}{2}\vec{t}_1^T\Sigma_{11}\vec{t}_1\right] \times \exp\left[i\vec{t}_2^T\vec{\mu}_2 - \frac{1}{2}\vec{t}_2^T\Sigma_{22}\vec{t}_2\right] \\
&= \phi_1(\vec{t}_1)\phi_2(\vec{t}_2)
\end{aligned}$$

定理 5.30 (独立性判定)

如果 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, 则 $(X_1, X_2, ..., X_n)$ 相互独立的充要条件是

$$\Sigma = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2).$$

♡

定理 5.31 (条件分布)

设 $\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma)$, $det(\Sigma) > 0$ 和分块矩阵

$$\vec{X} = \begin{pmatrix} \vec{X}_1 \\ \vec{X}_2 \end{pmatrix}, \quad \vec{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

其中 $\vec{X}_1, \vec{\mu}_1$ 和方阵 Σ_{11} 的行数相同, 则在条件 $\vec{X}_1 = \vec{x}_1$ 下, $\vec{X}_2$ 服从多元正态分布 $\mathcal{N}\left(\vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$.

♡

证明 令

$$\begin{pmatrix} \vec{Y}_1 \\ \vec{Y}_2 \end{pmatrix} = \begin{pmatrix} I_{11} & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{22} \end{pmatrix} \begin{pmatrix} \vec{X}_1 - \vec{\mu}_1 \\ \vec{X}_2 - \vec{\mu}_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}\right)$$

故 $\vec{Y}_1$ 与 $\vec{Y}_2$ 独立, 注意到 $\vec{Y}_2 = \vec{X}_2 - \vec{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\vec{Y}_1$,
所以

$$\vec{X}_2 = \vec{Y}_2 + \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{X}_1 - \vec{\mu}_1)$$

从而

$$\begin{aligned}
&P(\vec{X}_2 \le \vec{x}_2 \mid \vec{X}_1 = \vec{x}_1) \\
&= P(\vec{Y}_2 + \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1) \le \vec{x}_2 \mid \vec{X}_1 = \vec{x}_1) \\
&= P(\vec{Y}_2 + \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1) \le \vec{x}_2 \mid \vec{Y}_1 = \vec{x}_1 - \vec{\mu}_1) \\
&= P(\vec{Y}_2 + \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1) \le \vec{x}_2)
\end{aligned}$$

又 $\vec{Y}_2 \sim \mathcal{N}(\mathbf{0}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$, 故

$$\begin{aligned}
\vec{X}_2 \mid_{\vec{X}_1 = \vec{x}_1} &\overset{d}{=} \vec{Y}_2 + \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1) \\
&\sim \mathcal{N}\left(\vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).
\end{aligned}$$