概率论笔记

作者：肖程哲

时间：November 20, 2022

苟日新，日日新，又日新

目录

第 1 章 概率基础

考试重点

- 古典概型与几何概型
- 条件概率与独立

1.1 概率空间

1.1.1 随机事件

定义 1.1 (样本空间)

随机试验 (trial) 可能出现的基本 结果称为**样本点** (sample point)，记为 ω；样本的全体构成**样本空间** (sample space)，记为 $\Omega = \{\omega\}$。♣

定义 1.2 (事件的古典定义)

样本点 ω 的某种集合称为（随机）**事件** (event)。♣

随机现象被抽象为集合。事件间的包含、等价、互不相容关系，对应成集合间的包含、相等、不相交；逻辑运算（且、或、非等）对应成集合论运算（交、并、补等）。

例题 1.1 设 A、B、C 是某个随机现象的三个事件, 则

- 事件“A 与 B 发生,C 不发生”可表示为:$AB\overline{C}$.
- 事件“A、B、C 中至少有一个发生”可表示为:$A \cup B \cup C$.
- 事件“A、B、C 中至少有两个发生”可表示为:$AB \cup AC \cup BC$.
- 事件“A、B、C 中恰好有两个发生”可表示为:$AB\overline{C} \cup A\overline{B}C \cup \overline{A}BC$.
- 事件“A、B、C 同时发生”可表示为:ABC.
- 事件“A、B、C 都不发生”可表示为:$\overline{A} \cap \overline{B} \cap \overline{C}$.
- 事件“A、B、C 不全发生”可表示为:$\overline{A} \cup \overline{B} \cup \overline{C}$.

性质 集合的运算性质：

- 交换律

$$A \cup B = B \cup A, \quad AB = BA \tag{1.1}$$

- 结合律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC) \tag{1.2}$$

- 分配律

$$(A \cup B) \cup C = A \cup (B \cup C), \tag{1.3}$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{1.4}$$

- 对偶律（De Morgan 公式）

$$\overline{A \cup B} = \overline{A} \cap \overline{B}, \tag{1.5}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}. \tag{1.6}$$

以上各式皆可推广到可数个集合的情况。

注 并与补是集合中最基本的运算：

- 交的运算可通过并与对立来实现（对偶律）.

- 差的运算可通过对立与交来实现（$A - B = A\overline{B}$）.

1.1.2 概率空间

为方便概率的定义，避免不可测集的出现，并不把 Ω 的一切子集作为事件。

定义 1.3 (事件域)

事件构成的全体称为**事件域** $\mathscr{F}$，是 Ω 的子集族 (collection of subsets)，应满足σ 代数 的要求：

- $\varnothing \in \mathscr{F}$，代表无事发生；
- $A \in \mathscr{F} \implies A^{\complement} \in \mathscr{F}$，即 $\mathscr{F}$ 对补集运算 (逻辑上的非) 封闭；
- $A_1, \ldots, A_n, \ldots \in \mathscr{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$，即 $\mathscr{F}$ 对可数并运算封闭。

注 可数性是为了在数学上能够恰当地处理无穷 的概念, 术语中的 σ 指的就是可数并。由对偶原理可得 σ 域同时对可数并运算封闭. 即 σ 域对逆, 并, 交, 差的可数次运算封闭.

事件域根据问题的不同要求适当选取。在概率定义没有困难时，应尽量取得大，通常以 Ω 的一切子集作为事件域。当 Ω 给定后，若某些子集必须作为事件处理，能否找到包含他们的 σ 域?

命题 1.1

若给定 Ω 的一个非空集族 $\mathscr{G}$, 必存在 Ω 上唯一的 σ 域 $\mathrm{m}(\mathscr{G})$, 满足下列性质:

- 包含 $\mathscr{G}$
- 若其他 σ 域包含 $\mathscr{G}$, 则必包含 $\mathrm{m}(\mathscr{G})$

这个 $\mathrm{m}(\mathscr{G})$ 称为包含 $\mathscr{G}$ 的最小 σ 域, 或由 $\mathscr{G}$ 扩张而成的 σ 域.

扩张, 或者称为*延拓*, 是数学中很重要的一个概念, 大抵是将某映射的定义域适当扩大, 不改变在初始定义域上的映射取值 (注意值域可能是比较抽象的集合, 配备了某些操作之后被称为空间), 同时在扩大后的定义域上仍然保持某些优良的性质. 与此相对的概念是*限制*, 即关心局部上可能更加漂亮的性质, 把初始的定义域适当缩小.

证明 由于 Σ 的一切子集构成的集类包含 $\mathscr{G}$, 所以 m 存在. 再取 Σ 上满足此条件的 σ 域之交作为 $\mathrm{m}(\mathscr{G})$ 即可.

特别地, 实数集 $\mathbb{R}$ 的子集族 $\{(-\infty, x] : x \in \mathbb{R}\}$ 生成的 σ 代数 $\mathscr{B}_{\mathbb{R}}$ 称为 $\mathbb{R}$ 上的 *Borel* 代数.

定义 1.4 (Borel 集)

设 $\mathbb{R}^1$ 为全集, 形为 $[a, b)$ 构成的集类产生的 σ 域称为**一维 Borelσ 域**, 记为 $\mathscr{B}_1$, 其中的元素称为**一维 Borel 集**

若 x, y 为任意实数, 由于:

$$\{x\} = \bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n}\right)$$

$$(x, y) = [x, y) - \{x\}$$

$$[x, y] = [x, y) + \{y\}$$

$$(x, y] = [x, y) + \{y\} - \{x\}$$

因此 $\mathscr{B}_1$ 包含一切开区间, 闭区间, 单个实数, 可列个实数, 以及他们经可列次逆, 并, 交, 差运算的集合.

定义 1.5 (概率空间)

设 Ω 为一个样本空间，$\mathscr{F}$ 为定义于其上的一个事件域。定义在事件域(非样本空间) 上的集合函数 $P : \mathscr{F} \to \mathbb{R}$ 称为**概率**的条件是:

非负性 $\quad \mathbb{P}(A) \geq 0, \forall A \in \mathscr{F}$

规范性 $\mathbb{P}(\Omega)=1$; (如果没有这条就是一般的有限测度)

可列可加性 若 $A_1,\ldots,A_n,\ldots\in\mathscr{F}$ 两两不交, 即 $A_i\cap A_j=\varnothing,\ \forall i\neq j$, 则

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

并称 $(\Omega,\mathscr{F},\mathbb{P})$ 为一个**概率空间** (probability space)

1.1.3 概率的性质

性质 概率的性质:

- $P(\varnothing)=0$;
- $P(A^{\complement})=1-P(A)$;
- 若 $A\subseteq B$ 则 $P(A)\le P(B)$;

例题 1.2 口袋中有编号为 $1,2,\cdots,n$ 的 n 个球，从中有放回地任取 m 次，求取出的 m 个球的最大号码为 k 的概率。

解 记事件 A_k 为“取出的 m 个球的最大号码为 k”。如果直接考虑事件 A_k，则比较复杂，因为“最大号码为 k”可以包括取到 1 次 k、取到 2 次 k、…、取到 m 次 k。为此我们记事件 B_i 为“取出的 m 个球的最大号码小于等于 i”。则 B_i 发生只需每次从 $1,2,\cdots,i$ 号球中取球即可。由题可知

$$P(B_i)=\frac{i^m}{n^m},\quad i=1,2,\ldots,n.$$

又因为 $A_k=B_k-B_{k-1}$，且 $B_{k-1}\subset B_k$，所以

$$\begin{aligned}P(A_k)&=P(B_k-B_{k-1})=P(B_k)-P(B_{k-1})\\&=\frac{k^m-(k-1)^m}{n^m},\quad k=1,2,\ldots,n.\end{aligned}$$

命题 1.2 ()

$$|P(AB)-P(A)P(B)|\le\frac{1}{4}$$

证明 设 $P(A)\ge P(B)$，则：

$$P(AB)-P(A)P(B)\le P(B)-P(B)P(B)=P(B)[1-P(B)]\le\frac{1}{4}$$

另一方面有：

$$\begin{aligned}P(A)P(B)-P(AB)&=P(A)[P(AB)+P(\bar{A}B)]-P(AB)\\&=P(A)P(\bar{A}B)-P(AB)[1-P(A)]\\&\le P(A)P(\bar{A}B)\le P(A)P(\bar{A})\\&=P(A)[1-P(A)]\le\frac{1}{4}\end{aligned}$$

命题 1.3 (加法公式)

基础形式:

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

一般形式:

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

特别地，若事件出现个数相同时概率相等，则可简化为:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = nP_1 - \binom{n}{2}P_2 + \binom{n}{3}P_3 - \cdots + (-1)^{n-1}P_n$$

推论 1.1

对任意 n 个事件 $A_1,\cdots,A_n$ 有

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

例题 1.3 配对问题 在一个有 n 个人参加的晚会上，每个人带了一件礼物，且假定各人带的礼物都不相同。晚会期间各人从放在一起的 n 件礼物中随机抽取一件，问至少有一个人自己抽到自己礼物的概率是多少?

解 以 A_i 记事件"第 i 个人自己抽到自己的礼物"，则所求概率为 $P(A_1\cup A_2\cup\cdots\cup A_n)$。因为

$$\begin{aligned}
&P(A_1) = P(A_2) = \cdots = P(A_n) = \frac{1}{n}\\
&P(A_1A_2) = P(A_1A_3) = \cdots = P(A_{n-1}A_n) = \frac{1}{n(n-1)}\\
&P(A_1A_2A_3) = P(A_1A_2A_4) = \cdots = P(A_{n-2}A_{n-1}A_n) = \frac{1}{n(n-1)(n-2)}\\
&\cdots\\
&P(A_1A_2\cdots A_n) = \frac{1}{n!}
\end{aligned}$$

再由概率的加法公式1.3得

$$P(A_1\cup A_2\cup\cdots\cup A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1}\frac{1}{n!}$$

显然，可列可加性可以推出有限可加性。但是一般而言，由有限可加性不能推出可列可加性。设 $A_i \in \mathscr{F}, i = 1,2,\cdots$ 且两两互不相容，则由概率的有限可加性只能推出下式成立:

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

这个等式的左边对任意 n 都不超过 1，因此右边的正项级数收敛。这样应有

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

即若希望由有限可加性推出可列可加性，则需要下式成立:

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

可写为:

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

若记

$$S_n = \sum_{i=1}^{n} A_i$$

则 $S_n \in \mathscr{F}$ 而且 $S_n \subseteq S_{n+1}$, 即 $\{S_n\}$ 是 $\mathscr{F}$ 中一个单调不减的集序列，这时将上式改写改写为：

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

定义 1.6 (下连续)

对于 $\mathscr{F}$ 上的集合函数 P，若它对 $\mathscr{F}$ 中任一单调不减的集序列 $\{S_n\}$ 均满足:

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

则称它是**下连续**的.

定理 1.1

若 P 是 $\mathscr{F}$ 上满足 $P(\Omega) = 1$ 的非负集合函数，则它具有可列可加性的充要条件为:

- 它是有限可加的;
- 它是下连续的.

证明 由上述推导可知其充分性。对于其必要性，有限可加也是易知的。设 $\{S_n\}$ 是 $\mathscr{F}$ 中一个单调不减的事件序列，即

$$\bigcup_{i=1}^{+\infty} S_i = \lim_{n\to+\infty} S_n$$

若定义 $S_0 = \varnothing$，则

$$\bigcup_{i=1}^{+\infty} S_i = \bigcup_{i=1}^{+\infty}(S_i - S_{i-1})$$

由于 $S_{i-1} \subset S_i$，显然诸 $(S_i - S_{i-1})$ 两两不相容，再由可列可加性得

$$P\left(\bigcup_{i=1}^{+\infty} S_i\right) = \sum_{i=1}^{+\infty} P(S_i - S_{i-1}) = \lim_{n\to+\infty} \sum_{i=1}^{n} P(S_i - S_{i-1})$$

由有限可加性得

$$\sum_{i=1}^{n} P(S_i - S_{i-1}) = P\left(\bigcup_{i=1}^{n}(S_i - S_{i-1})\right) = P(S_n)$$

所以

$$P(\lim_{n\to+\infty} S_n) = \lim_{n\to+\infty} P(S_n)$$

这就证得了 P 的下连续性.

同理定义上连续

定义 1.7 (上连续)

对于 $\mathscr{F}$ 上的集合函数 P，若它对 $\mathscr{F}$ 中任一单调不增的集序列 $\{B_n\}$ 均满足:

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

则称它是**上连续**的.

定理 1.2

概率是上连续的

证明 设 $\{B_n\}$ 是单调不增的事件序列，则 $\{\overline{B_n}\}$ 为单调不减的事件序列，由概率的下连续性得

$$\begin{aligned} 1-\lim_{n\to+\infty}P(B_n) &= \lim_{n\to+\infty}[1-P(B_n)] = \lim_{n\to+\infty}P(\overline{B_n}) \\ &= P\left(\bigcup_{n=1}^{+\infty}\overline{B_n}\right) = P\left(\overline{\bigcap_{n=1}^{+\infty}B_n}\right) \\ &= 1-P\left(\bigcap_{n=1}^{+\infty}B_n\right). \end{aligned}$$

由于 $\{B_n\}$ 单调不增，所以

$$\lim_{n\to+\infty}P(B_n) = P\left(\bigcap_{n=1}^{+\infty}B_n\right) = P\left(\lim_{n\to+\infty}B_n\right)$$

1.2 古典概型与几何概型

1.2.1 古典概型

古典概型的基本思想:

有限个样本点 所涉及的随机现象只有有限个样本点,譬如为 n 个,且这些事件是两两互不相容的;

等可能性 每个样本点发生的可能性相等

定义 1.8 (古典概型)

若样本空间中包含至多可数个样本点，则称其为**离散样本空间**。若定义事件域为其全体子集，且概率为：

$$P(A) = \frac{\text{事件}A\text{ 所含样本点的个数}}{\Omega\text{ 中所有样本点的个数}}$$

则称此概率空间为**古典概型**。

笔记 古典概型的大部分问题都能形象化地用摸球模型来描述，以后经常研究摸球模型的意义即在于此。

例题 1.4 生日问题 n 个人的生日全不相同的概率 p_n 是多少?

解 把 n 个人看成是 n 个球，将一年 365 天看成是 $N=365$ 个盒子，则 “n 个人的生日全不相同” 就相当于 “恰好有 $n(n\le N)$ 个盒子各有一球”，所以 n 个人的生日全不相同的概率为：

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

上式可用以下方法作近似计算：

1. 当 n 较小时，原式右边中各因子的第二项之间的乘积 $\frac{i}{365}\times\frac{j}{365}$ 都可以忽略，于是有近似公式

$$p_n \approx 1-\frac{1+2+\cdots+(n-1)}{365} = 1-\frac{n(n-1)}{730}$$

2. 当 n 较大时，因为对小的正数 x 有 $\ln(1-x)\approx -x$，所以由原式得

$$\ln p_n \approx \frac{1+2+\cdots+(n-1)}{365} = -\frac{n(n-1)}{730}$$

1.2.2 几何概型

几何概型的基本思想:

不可数样本点 如果一个随机现象的样本空间 Ω 充满某个区域，其测度 (长度、面积或体积等) 大小可用 S_Ω 表示;

测度等可能性 任意一点落在测度相同的子区域内是等可能的。

定义 1.9 (几何概型)

若将“在区域 Ω 中随机地取一点，而该点落在区域 g 中”记为事件 A_g，则其概率定义为

$$\mathbb{P}(A_g) = \frac{S_g}{S_\Omega}$$

♣

注 可用均匀分布（定义4.10）准确描述

例题 1.5 会商问题 甲乙两人约定在下午 6 时到 7 时之间在某处会面，并约定先到者应等候另一个人20 min，过时即可离去。求两人能会面的概率。

解 以 x 和 y 分别表示甲、乙两人到达约会地点的时间 (以 min 为单位)，在平面上建立 xOy 直角坐标系 (见图1.1). 因为甲、乙都是在 0 至 60 min 内等可能地到达，所以由等可能性知这是一个几何概率问题。(x, y) 的所有可能

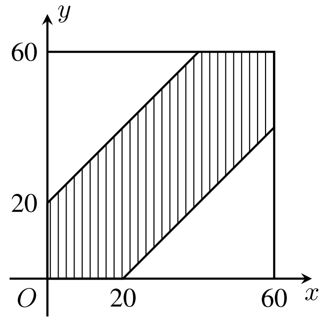


图 1.1: 会面问题中的 Ω 与 A.

取值是边长为 60 的正方形，其面积为 $S_\Omega = 60^2$。而事件 A =“两人能够会面”相当于:

$$|(|x-y) \leq 20$$

即图中的阴影部分，其面积为 $S_A = 60^2 - 40^2$，由几何概型计算公式可知

$$P(A) = \frac{S_A}{S_\Omega} = \frac{60^2 - 40^2}{60^2} = 0.5556$$

例题 1.6 蒲丰投针问题 平面上画有间隔为 $d(d > 0)$ 的等距平行线，向平面任意投掷一枚长为 $l(l < d)$ 的针，求针与任一平行线相交的概率.

解 以 x 表示针的中点与最近一条平行线的距离，又以 φ 表示针与此直线间的交角，见图1.2。易知样本空间 Ω 满足

$$0 \leq x \leq d/2, \quad 0 \leq \varphi \leq \pi$$

由这两式可以确定 $x - \varphi$ 平面上的一个矩形 Ω，这就是样本空间，其面积为 $S_\Omega = d\pi/2$。这时为了针与平行线相交 (记为事件 A)，其充要条件是

$$x \leq \frac{l}{2}\sin\varphi$$

由这个不等式表示的区域是图 1.3 中的阴影部分.

由于针是向平面任意投掷的，所以由等可能性知这是一个几何概率问题。由此得

$$P(A) = \frac{S_A}{S_\Omega} = \frac{\int_0^\pi \frac{l}{2}\sin\varphi \, d\varphi}{\frac{d}{2}\pi} = \frac{2l}{d\pi}.$$

如果 l, d 为已知，则以 π 的值代入上式即可计算得 $P(A)$ 之值。反之，如果已知 $P(A)$ 的值，则也可以利用上式去求 π，而关于 $P(A)$ 的值，可用从试验中获得的频率去近似它：即投针 N 次，其中针与平行线相交 n 次，则频率 n/N 可作为 $P(A)$ 的估计值，于是由

$$\frac{n}{N} \approx P(A) = \frac{2l}{d\pi}$$

可得

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

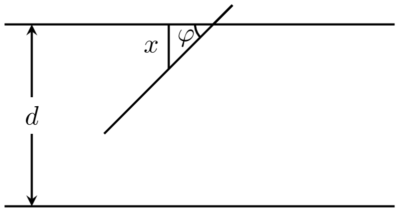


图 1.2: 蒲丰投针问题

.

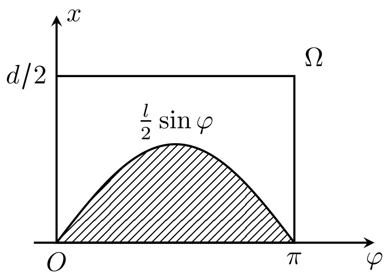


图 1.3: 蒲丰投针问题中的 Ω 和 A

.

例题 1.7 在长度为 a 的线段内任取两点将其分为三段，求它们可以构成一个三角形的概率。

解 由于是将线段任意分成三段，所以由等可能性知这是一个几何概率问题。分别用 x,y 和 $a-x-y$ 表示线段被分成的三段长度，见图1.4。则显然应该有

$$0 < x < a; \quad 0 < y < a; \quad 0 < a-(x+y) < a$$

第三个式子等价于：$0 < x+y < a$。所以样本空间为 (见图1.5)

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

Ω 的面积为

$$S_\Omega = \frac{a^2}{2}$$

又根据构成三角形的条件：三角形中任意两边之和大于第三边，得事件 A 所含样本点 (x,y) 必须满足:

$$\begin{aligned}
&0 < a-(x+y) < x+y,\\
&0 < x < y+(a-x-y),\\
&0 < y < x+(a-x-y).
\end{aligned}$$

整理得

$$\frac{a}{2} < x+y < a; \quad 0 < x < \frac{a}{2}; \quad 0 < y < \frac{a}{2}$$

所以事件 A 可用图 1.6 中的阴影部分表示。事件 A 的面积为

$$S_A = \frac{a^2}{8}.$$

由此得

$$P(A) = \frac{1}{4}.$$

1.2.3 Bertrand 奇论

在一圆内任取一条弦，问其长度超过该圆内接等边三角形的边长的概率是多少？这是一个几何概率问题，它有三种解法，具体如下：

1. 由于对称性，可只考察某指定方向的弦。作一条直径垂直于这个方向。显然，只有交直径于 1/4 与 3/4 之

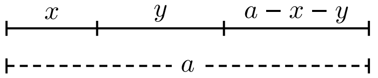


图 1.4: 长度为 a 的线段分成三段.

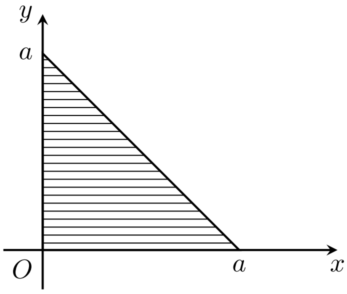


图 1.5: 线段分成三段的样本空间

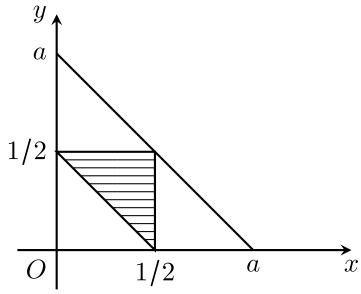


图 1.6: 构成三角形的条件

间的弦才能超过正三角形的边长（见图），如此，所求概率为 1/2。

2. 由于对称性，可让弦的一端点固定，让另一端点在圆周上作随机移动。若在固定端点作一切线，则与此切线夹角在 60° 与 120° 之间的弦才能超过正三角形的边长（见图），如此，所求概率为 1/3。
3. 圆内弦的位置被其中点唯一确定。在圆内作一同心圆，其半径仅为大圆半径的一半，则大圆内弦的中点落在小圆内，此弦长才能超过正三角形的边长（见图），如此，所求概率为 1/4。

同一问题有三种不同答案，究其原因在圆内“取弦”时规定尚不够具体，不同的“等可能性假定”导致了不同的样本空间。具体如下：其中“均匀分布”应理解为“等可能取点”。

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

可见，上述三个答案是针对三个不同样本空间引起的，它们都是正确的，贝特朗奇论引起人们注意，在定义概率时要事先明确指出样本空间是什么.

1.3 条件概率与独立

1.3.1 条件概率

定义 1.10 (条件概率)

设 $A, B \in \mathscr{F}$，且 $P(B) > 0$ 称

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

为基于于 B 的**条件概率** (probability conditional on$B$)。♣

命题 1.4

条件概率仍是一个概率测度，即若设 $P(B) > 0$，则

- $P(A|B) \geq 0, \quad A \in \mathscr{F}$,
- $P(\Omega|B) = 1$,
- 若 $\mathscr{F}$ 中的 $A_1, A_2, \cdots, A_n, \cdots$ 互不相容，则

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

♠

证明 从条件概率的定义很容易得出前两点。因为 $A_1, A_2, \cdots, A_n, \cdots$ 互不相容，所以 $A_1B, A_2B, \cdots, A_nB, \cdots$ 也

互不相容，故

$$P\left(\bigcup_{n=1}^{+\infty} A_n \middle| B\right) = \frac{P\left(\left(\bigcup_{n=1}^{+\infty} A_n\right)B\right)}{P(B)} = \frac{P\left(\bigcup_{n=1}^{+\infty}(A_nB)\right)}{P(B)}$$
$$= \sum_{n=1}^{+\infty}\frac{P(A_nB)}{P(B)} = \sum_{n=1}^{+\infty}P(A_n|B).$$

定理 1.3 (乘法法则)

设 $A, B \in \mathscr{F}$，且 $P(B) > 0$，则

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

泛化后有：

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

例题 1.8 罐子模型 设罐中有 b 个黑球、r 个红球，每次随机取出一个球，取出后将原球放回，还加进 c 个同色球和 d 个异色球。记 B_i 为“第 i 次取出的是黑球”，R_j 为“第 j 次取出的是红球”。若连续从罐中取出三个球，其中有两个红球、一个黑球。则由乘法公式我们可得

$$\begin{aligned}
P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2)\\
&= \frac{b}{b+r}\frac{r+d}{b+r+c+d}\frac{r+d+c}{b+r+2c+2d},\\
P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2)\\
&= \frac{r}{b+r}\frac{b+d}{b+r+c+d}\frac{r+d+c}{b+r+2c+2d},\\
P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2)\\
&= \frac{r}{b+r}\frac{r+c}{b+r+c+d}\frac{b+2d}{b+r+2c+2d}.
\end{aligned}$$

以上概率与黑球在第几次被抽取有关.

罐子模型也称为波利亚 (Polya) 模型，这个模型可以有各种变化，具体见下:

1. 当 $c = -1, d = 0$ 时，即为**不返回抽样**。此时前次抽取结果会影响后次抽取结果。但只要抽取的黑球与红球个数确定，则概率不依赖其抽出球的次序，都是一样的。此例中有

$$\begin{aligned}
P(B_1R_2R_3) = P(R_1B_2R_3) &= P(R_1R_2B_3)\\
&= \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}.
\end{aligned}$$

2. 当 $c = 0, d = 0$ 时，即为**返回抽样**。此时前次抽取结果不会影响后次抽取结果。故上述三个概率相等，且都等于

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br^2}{(b+r)^3}$$

3. 当 $c > 0, d = 0$ 时，称为**传染病模型**。此时，每次取出球后会增加下一次取到同色球的概率，或换句话说，每次发现一个传染病患者，以后都会增加再传染的概率。与前面两个一样，以上三个概率都相等，且都等于

$$\begin{aligned}
P(B_1R_2R_3) = P(R_1B_2R_3) &= P(R_1R_2B_3)\\
&= \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}.
\end{aligned}$$

从以上可以看出：在罐子模型中只要 $d = 0$，则以上三个概率都相等。即只要抽取的黑球与红球个数确定，则概率不依赖其抽出球的次序，都是一样的。但当 $d > 0$ 时，就不同了.

4. 当 $c = 0, d > 0$ 时，称为**安全模型**。此模型可解释为：每当事故发生了 (红球被取出)，安全工作就抓紧一

些，下次再发生事故的概率就会减少；而当事故没有发生时 (黑球被取出)，安全工作就放松一些，下次再发生事故的概率就会增大。在这种场合，上述三个概率分别为

$$P(B_1R_2R_3) = \frac{b}{b+r}\frac{r+d}{b+r+d}\frac{r+d}{b+r+2d},$$
$$P(R_1B_2R_3) = \frac{r}{b+r}\frac{b+d}{b+r+d}\frac{r+d}{b+r+2d},$$
$$P(R_1R_2B_3) = \frac{r}{b+r}\frac{r}{b+r+d}\frac{b+2d}{b+r+2d}.$$

定理 1.4 (全概率公式)

设可列个事件 $B_1, B_2, \cdots, B_n, \cdots$ 互不相容，且包含事件 A。即 $A \subseteq \bigcup_{i=1}^{\infty} B_i$，　$B_iB_j = \varnothing, \forall i \neq j$，则：

$$P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i), \quad \forall A \in \mathscr{F}$$

笔记 $P(A|B_i)$ 可视为事件 A 在 B_i 上的分量，$P(B_i)$ 则为其权重，加权平均后得到 $P(A)$

证明 因为 $A \subseteq \bigcup_{i=1}^{\infty} B_i$，所以

$$A = A(\bigcup_{n=1} \infty B_i) = \bigcup_{i=1} \infty (AB_i)$$

且 $AB_1, AB_2, \cdots, AB_n, \cdots$ 互不相容，所以由可加性得

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

再将 $P(AB_i) = P(B_i)P(A|B_i)$，代人上式即得全概率公式。

推论 1.2

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

例题 1.9 敏感性问题调查 学生阅读黄色书刊和观看黄色影像会严重影响学生身心健康发展。但这属个人隐私行为。现在要设计一个调查方案，从调查数据中估计出学生中阅读黄色书刊和观看黄色影像的比率 p.

敏感性问趣的调查方案，关键要使被调查者愿意作出真实回答又能保守个人秘密。经过多年研究和实践，一些心理学家和统计学家设计了一种调查方案，在这个方案中被调查者只需回答以下两个问题中的一个问题，而且只需回答"是"或"否"。

- 你的生日是否在 7 月 1 日之前？
- 你是否看过黄色书刊或影像？

这个调查方来看似简单，但为了消除被调查者的顾虑，使被调查者确信他这次调查不会泄露个人秘密，在操作上有以下关键点:

- 被调查者在没有旁人的情况下，独自一人回答问题.
- 被调查者从只有白球和红球的罐子中随机抽一只球，看过颜色后放回。若抽到白球，则回答问题 A；若抽到红球，则回答问题 B。

如此的调查方法，主要在于旁人无法知道被调查回答的是问题 A 还是问题 B，由此可以极大地消除被调查者的顾虑。现在的问题是如何分析调查的结果。很显然，我们只对问题 B 是感兴趣。

首先假设有 n 张答卷 (n 较大)，其中有 k 张回答"是"。有两个信息我们是预先知道的：

- 在参加人数较多的情况，任选一人其生日在 7 月 1 日之前的概率为 0.5.
- 罐中红球的比率 π。

由全概率公式得

$$P(\text{是}) = P(\text{白球})P(\text{是}|\text{白球}) + P(\text{红球})P(\text{是}|\text{红球})$$

将 $P(\text{红球}) = \pi, P(\text{白球}) = 1 - \pi, P(\text{是}|\text{白球}) = 0.5, P(\text{是}|\text{红球}) = p$ 代入上式右边，而上式左边用频率 k/n 代

替概率 P(是)，得

$$\frac{k}{n} = 0.5(1-\pi) + p\cdot\pi$$

由此得

$$p = \frac{k/n - 0.5(1-\pi)}{\pi}$$

定理 1.5 (Bayes 定理)

设可列个事件 $B_1, B_2, \cdots, B_n, \cdots$ 互不相容，且包含事件 A。即 $A \subseteq \bigcup_{i=1}^{\infty} B_i$，$\quad B_iB_j = \varnothing, \forall i \neq j$，则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}$$

例题 1.10 某地区居民的肝癌发病率为0.0004，现用甲胎蛋白法进行普查。医学研究表明，化验结果是存有错误的。已知患有肝癌的人其化验结果 99 % 呈阳性 (有病)，而没患肝癌的人其化验结果 99.9 % 呈阴性 (无病)。现某人的检查结果呈阳性，求其真患肝癌的概率。

解 记 B 为事件“被检查者患有肝癌”，A 为事件“检查结果呈阳性”。由题设知

$$P(B) = 0.0004, \quad P(\overline{B}) = 0.996,$$
$$P(A|B) = 0.99, \quad P(A|\overline{B}) = 0.001.$$

由贝叶斯公式得

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\overline{B})P(A|\overline{B})} \\ &= \frac{0.0004 \times 0.99}{0.0004 \times 0.99 + 0.9996 \times 0.001} \\ &= 0.284. \end{aligned}$$

条件概率的三公式中：乘法公式是求事件交的概率；全概率公式是求一个复杂事件的概率；而贝叶斯公式是求一个条件概率。在贝叶斯公式中，称 $P(B_i)$ 为 B_i 的**先验概率**，称 $P(B_i|A)$ 为 B_i 的**后验概率**。贝叶斯公式是专门用于计算后验概率的，也就是通过 A 的发生这个新信息，来对 B_i 的概率作出的修正。

1.3.2 独立性

定义 1.11 (事件的独立性)

如果 $A, B \in \mathscr{F}$ 满足

$$P(A \cap B) = P(A)P(B),$$

则称 A 与 B **独立** (independent), 记为 $A \perp\!\!\!\perp B$.

笔记 当 $P(A) > 0$ 时，$P(B|A) = P(B) \iff B \perp\!\!\!\perp A$, 由此可得到<u>$B$ 独立于 A</u> 的直观理解

性质 独立性是对称的, 即 $A \perp\!\!\!\perp B \iff B \perp\!\!\!\perp A$. 若两事件独立, 则其补集也独立.

$$\begin{array}{ccc} A & \longleftrightarrow & B \\ & \times & \\ A^{\mathsf{C}} & \leftrightarrow & B^{\mathsf{C}} \end{array}$$

证明 由概率的性质知

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

又由 A 与 B 的独立性知

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

所以

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

这表明 A 与 $\overline{B}$ 独立. 类似可证 $\bar{A}$ 与 $\overline{B}$ 独立，$\bar{A}$ 与 B 独立.

例题 1.11 有两名选手比赛射击，轮流对同一目标进行射击，甲命中目标的概率为 α，乙命中目标的概率为 β。甲先射，谁先命中谁得胜。问甲、乙两人获胜的概率各为多少？

解 记事件 A_i 为“第 i 次射击命中目标”，因为甲先射，所以事件“甲获胜”可以表示为

$$A_1 \cup \bar{A}_1\bar{A}_2A_3 \cup \bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4A_5 \cup \cdots$$

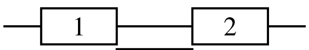
又因为各次射击是独立的，所以得

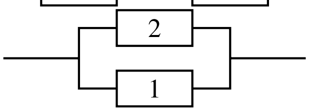
$$\begin{aligned} P(\text{甲获胜}) &= \alpha + (1-\alpha)(1-\beta)\alpha + (1-\alpha)^2(1-\beta)^2\alpha + \cdots \\ &= \alpha \sum_{i=0}^{+\infty}(1-\alpha)^i(1-\beta)^i \\ &= \frac{\alpha}{1-(1-\alpha)(1-\beta)}. \end{aligned}$$

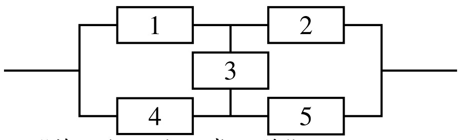
同理可得

$$\begin{aligned} P(\text{乙获胜}) &= P(\bar{A}_1A_2 \cup \bar{A}_1\bar{A}_2\bar{A}_3A_4 \cup \cdot) \\ &= (1-\alpha)\beta + (1-\alpha)(1-\beta)(1-\alpha)\beta + \cdots \\ &\beta(1-\alpha)\sum_{i=0}^{+\infty}(1-\alpha)^i(1-\beta)^i \\ &= \frac{\beta(1-\alpha)}{1-(1-\alpha)(1-\beta)}. \end{aligned}$$

例题 1.12 系统由多个元件组成，且所有元件都独立地工作. 设每个元件正常工作的概率都为 $p = 0.9$，试求以下系统正常工作的概率.

1. 串联系统 S_1： 

2. 并联系统 S_2： 

3. 5 个元件组成的桥式系统 S_3： 

解 设 S_i = “第 i 个系统正常工作”，A_i = “第 i 个元件正常工作”.

(1) 对串联系统而言，“系统正常工作”相当于“所有元件正常工作”，即 $S_1 = A_1A_2$，所以

$$P(S_1) = P(A_1A_2) = P(A_1)P(A_2) = p^2 = 0.81$$

(1) 对并联系统而言，“系统正常工作”相当于“至少一个元件正常工作”，即 $S_2 = A_1 \cup A_2$，所以

$$P(S_2) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) = p + p - p^2 = 0.99$$

或

$$\begin{aligned} P(S_2) &= 1 - P(\bar{S}_2) = 1 - P(A_1 \bar{\cup} A_2) = 1 - P(\bar{A}_1 \cap \bar{A}_2) \\ &= 1 - P(\bar{A}_1)P(\bar{A}_2) = 1 - (1-p)^2 = 0.99. \end{aligned}$$

(1) 在桥式系统中，第 3 个元件是关键，我们先用全概率公式得

$$P(S_3) = P(A_3)P(S_3|A_3) + P(\bar{A}_3)P(S_3|\bar{A}_3)$$

因为在“第 3 个元件正常工作”的条件下，系统成为先并后串系统（见图 1.7）。所以

$$P(S_3|A_3) = P((A_1 \cup A_4)(A_2 \cup A_5)) = P(A_1 \cup A_4)P(A_2 \cup A_5)$$
$$= [1-(1-p)^2]^2 = 0.9801.$$

又因为在“第 3 个元件不正常工作”的条件下，系统成为先串后并系统（见图 1.8）.

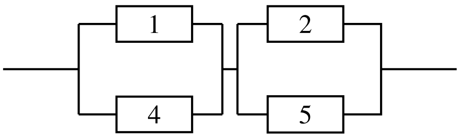


图 1.7: 先并后串系统

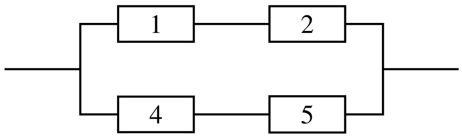


图 1.8: 先串后并系统

所以

$$P(S_3|\bar{A}_3) = P(A_1A_2| \cup A_4A_5) = 1-(1-p^2)^2 = 0.9639$$

最后得

$$P(S_3) = p[1-(1-p)^2]^2 + (1-p)[1-(1-p^2)^2]$$
$$= 0.9 \times 0.9801 + 0.1 \times 0.9639 = 0.9785.$$

定义 1.12 (相互独立)

对于事件集 $A_1, A_2, \ldots, A_n$, 若对于其中任意子集 $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ 有:

$$P(A_{i_1} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$$

则称此事件集**相互独立** (mutually independent)

注 各事件两两独立不能推出所有事件相互独立。

例题 1.13 伯恩斯坦反例 一个均匀的正四面体，其第一面染成红色，第二面染成白色，第三面染成黑色，而第四面同时染上红，白，黑三种颜色. 现在以 A, B, C 分别记投一次四面体出现红，白，黑颜色朝下的事件，则由于在四面体中有两面有红色，因此 $P(A) = \frac{1}{2}$。同理 $P(B) = P(C) = \frac{1}{2}$，同时易得 $P(AB) = P(BC) = P(AC) = \frac{1}{4}$。所以 A, B, C 两两独立，但是 $P(ABC) = \frac{1}{4} \neq P(A)P(B)P(C)$，从而 A, B, C 不相互独立。

定义 1.13 (事件域的独立性)

若 $\mathscr{G} \subset \mathscr{F}$ 与 $\mathscr{H} \subset \mathscr{F}$ 满足

$$A \perp\!\!\!\perp B, \quad \forall A \in \mathscr{G}, \ B \in \mathscr{H},$$

则称 $\mathscr{G}$ 与 $\mathscr{H}$ 独立, 记为 $\mathscr{G} \perp\!\!\!\perp \mathscr{H}$

测度论告诉我们一个重要结果: 如果 $\mathscr{G}$ 对交集运算封闭, 那么成立 $\mathscr{G} \perp\!\!\!\perp \mathscr{H} \implies \sigma(\mathscr{G}) \perp\!\!\!\perp \mathscr{H}$，其中 $\sigma(\mathscr{G})$ 是 $\mathscr{G}$ 扩张而成的一个合适的最小的 σ 代数, 定义为所有包含 $\mathscr{G}$ 的 σ 代数的交集。

类似地可以定义 n 个试验 $E_1, E_2, \cdots, E_n$ 的相互独立性：如果 E_1 的任一结果、E_2 的任一结果……E_n 的任一结果都是相互独立的事件，则称**试验** $E_1, E_2, \cdots, E_n$ **相互独立**。

定义 1.14 (Bernouli 试验)

如果 n 个独立试验是相同的，则称其为 n 重**独立重复试验**。如果在 n 重独立重复试验中，每次试验的可能结果为两个：A 或 $\bar{A}$，且每次实验事件 A 发生的概率不变，则称这种试验为 n 重 **Bernouli 试验**。

错题记录

1. （茆 1.2.4）从一副 52 张的扑克牌中任取 4 张，求下列事件的概率:

(a). 全是黑桃;
(b). 同花;
(c). 没有两张同一花色;
(d). 同色.

2. (茆 1.2.5) 考虑一元二次方程 $x^2+Bx+C=O$，其中 B,C 分别是将一颗骰子接连掷两次先后出现的点数，求该方程有实根的概率和有重根的概率
3. (茆 1.2.10) 从 n 个数 $1,2,\cdots,n$ 中任取 2 个，问其中一个小于 $k(1<k<n)$，另一个大于 k 的概率是多少?
4. (茆 1.2.13) 把 10 本书任意地放在书架上，求其中指定的三本书放在一起的概率.
5. (茆 1.2.19) n 个男孩，m 个女孩 $(m\le n+1)$ 随机地排成一排，试求任意两个女孩都不相邻的概率.
6. (茆 1.2.20) 将 3 个球随机地放入 4 个杯子中去，求杯子中球的最大个数分别为 1, 2, 3 的概率各为多少?
7. (茆 1.2.22) 将 n 个完全相同的球 (这时也称球是不可辨的) 随机地放入 N 个盒子中，试求恰好有 m 个空盒的概率
8. (茆 1.2.23) 在区间 $(0,1)$ 中随机地取两个数，求事件“两数之和小于 7/5”的概率.
9. (茆 1.3.4) 从 $0,1,2,\cdots,9$ 十个数字中任意选出三个不同的数字, 试求事件“三个数字中含 0 但不含 5”的概率
10. (茆 1.3.8) 从数字 $1,2,\cdots,9$ 中可重复地任取 n 次，求 n 次所取数字的乘积能被 10 整除的概率.
11. (茆 1.3.23) 证明:$|P(AB)-P(A)P(B)|\le\frac{1}{4}$
12. (茆 1.4.6) n 件产品中有 m 件不合格品，从中任取两件，已知两件中有一件是合格品，求另一件也是合格品的概率
13. (茆 1.4.27) 口袋中有 a 个白球，b 个黑球和 n 个红球，现从中一个一个不返回地取球. 试证白球比黑球出现得早的概率为 $\frac{a}{a+b}$，与 n 无关.
14. (茆 1.5.20) 甲、乙、丙三人进行比赛，规定每局两个人比赛，胜者与第三人比赛，依次循环，直至有一人连胜两次为止，此人即为冠军. 而每次比赛双方取胜的概率都是 1/2，现假定甲、乙两人先比，试求各人得冠军的概率.
15. (茆 1.5.21) 甲、乙两个赌徒在每一局获胜的概率都是 1/2. 两人约定谁先赢得一定的局数就获得全部赌本. 但赌博在中途被打断了，请问在以下各种情况下，应如何合理分配赌本：
    (a). 甲、乙两个赌徒都各需赢 k 局才能获胜;
    (b). 甲赌徒还需赢 2 局才能获胜，乙赌徒还需赢 3 局才能获胜;
    (c). 甲赌徒还需赢 n 局才能获胜，乙赌徒还需赢 m 局才能获胜.
16. (李 1.23) 任意从数列 $1,2,\cdots,N$ 中有放回地取出 n 个数并按大小排列成: $x_1\le x_2\le\cdots\le x_n$, 试求 $x_m=M$ 的概率。
17. (李 1.38) (赠券收集) 食品厂把印有水浒 108 将之一的画卡作为赠券装人某种儿童食品袋中，每袋一卡，试求购买 n 袋这种食品而能收齐全套画卡的概率。
18. (李 1.40) 有 w 个白球与 b 个黑球任你放人两个袋子中，让你的朋友随机抽一袋并从中摸出一只球，你将如何做以使你的朋友摸得黑球的概率最大。
19. (李 1.42) 父，母，子三人举行比赛，每局总有一人胜一人负 (没有和局)，每局的优胜者就与未参加此局的人再进行比赛，如果某人首先胜了两局，则他就是整个比赛的优胜者，由父决定第一局由哪两人参加，其中儿子实力最强，所以父为了使自己得胜的概率达到最大，就决定第一局由他与妻子先比赛，试证父的决策为最优策略 (任何一对选手中一人胜对方的概率在整个比赛中是不变的).

第 2 章 随机变量

内容提要

- ❑ 离散与连续随机变量
- ❑ 一元与多元
- ❑ cdf, pmf, pdf
- ❑ 条件分布
- ❑ 独立随机变量
- ❑ 随机变量函数的分布
- ❑ 次序随机变量

在概率论中，主要关心 X 取值于数值集合 $\mathcal{X}$ 中某个子集 B 的可能性，即希望得到 $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$。概率论不关心具体的样本点 $\omega \in \Omega$，而关注其集合 $\{\omega \in \Omega : X(\omega) \in B\}$，将其记为 $X^{-1}(B)$。由于 $\mathbb{P}$ 定义在 $\mathscr{F}$ 上，故需 $X^{-1}(B) \in \mathscr{F}$。

定义 2.1 (可测性)

设所有值得关心的 $B \subset \mathcal{X}$ 组成 $\mathscr{F}_{\mathcal{X}}$，且 $\forall B \in \mathscr{F}_{\mathcal{X}}$ 都满足 $X^{-1}(B) \in \mathscr{F}$，则称 X 为 $\mathscr{F}/\mathscr{F}_{\mathcal{X}}$ **可测的**。当 $\mathscr{F}_{\mathcal{X}}$ 不引起混淆时，简记为关于 $\mathscr{F}$ **可测**，写作 $X \in \mathscr{F}$。 ♣

由于原像保持交、并、补等集合运算, 且 $\mathscr{F}$ 是 σ 代数, 可将 $\mathscr{F}_{\mathcal{X}}$ 扩张为合适的最小的 σ 代数, 即 $\sigma(\mathscr{F}_{\mathcal{X}})$, 因此可测映射的定义不妨只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数 的情况.

定义 2.2 (随机变量)

为了表示因随机性而变动的量，称可测映射(measurable mapping)

$$X : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{F}_{\mathcal{X}}), \quad \omega \in \Omega \mapsto X(\omega) \in \mathcal{X}$$

为**随机元** (random element)，也称**随机变量** (random variable)。其中 $\mathscr{F}_{\mathcal{X}}$ ♣

由于只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数的情况，可将随机变量看作将原概率空间映射到新概率空间的方式。新样本空间为 $\mathcal{X}$（一般由Borel 点集 构成），新事件域为 $\mathscr{F}_{\mathcal{X}}$，概率测度等于对应原像的。

注 使用随机变量 X 时，有两个可能的含义：

- X 的 (随机) 取值
- X 的分布

例如若设 $Y = -X, X \sim U(0,1)$，则两者对某样本点的取值是不同的，但两者的分布相同。

定义 2.3 (随机向量)

若随机变量 $X_1(\omega), X_2(\omega), \cdots, X_n(\omega)$ 定义在同一概率空间$(\Omega, \mathscr{F}, \mathbb{P})$ 上, 则称

$$X(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$$

构成一个 n 维**随机向量**，亦称 n 维随机变量。 ♣

定义 2.4 (离散与连续随机变量)

若 $\mathcal{X}$ 是（至多可数的）离散点集，$\mathscr{F}_{\mathcal{X}}$ 由 $\mathcal{X}$ 的所有子集组成，则称 X 为**离散随机变量** (discrete random varible)。若 $\mathcal{X} = \mathbb{R}^n$，$\mathscr{F}_{\mathcal{X}}$ 为 $\{\prod_{i=1}^n (-\infty, x_i] : x_1, \ldots, x_n \in \mathbb{R}\}$ 生成的 Borel 代数 (最小的 σ 代数)，则称其为**连续随机变量** (continuous random varible)。 ♣

2.1 随机变量的分布

2.1.1 分布函数

定义 2.5 (概率分布)

称随机元 X 诱导的概率测度

$$Q(\bullet) = \mathbb{P}\{X \in \bullet\}, \ \bullet \in \mathscr{F}_X$$

为 X 的**概率分布** (distribution/law)。

注 对于随机变量，其取值是随机的，但其分布是固定的。

定义 2.6 (单变量分布函数)

若函数 $F : \mathbb{R} \to [0,1]$ 满足以下性质，则称为单变量分布函数：

单调性 $F(x_1) \le F(x_2), \quad \forall x_1 < x_2$

有界性 $\lim_{n\to-\infty} F(x) = 0, \quad \lim_{n\to\infty} F(x) = 1$

右连续性 $\lim_{x\to x_0^+} F(x) = F(x_0)$

命题 2.1

对每个分布 Q 都存在唯一一个分布函数 F_Q 使得 $F_Q(x) = Q[(-\infty, x]], \quad \forall x \in \mathbb{R}$ 成立。

证明

$$\begin{aligned}
&\because \forall x_1 > x_2 \in \mathbb{R}, \ (-\infty, x_1] = (-\infty, x_2] \cup (x_2, x_1] \\
&\therefore \forall x_1 > x_2 \in \mathbb{R}, \ Q[(-\infty, x_1]] = Q[(-\infty, x_2]] + Q[(x_2, x_1]] \\
&\Rightarrow F_Q(x_1) = F_Q(x_2) + \mathbb{P}\{X \in (x_2, x_1]\} \ge F_Q(x_2)
\end{aligned}$$

即 F_Q 满足单调性。

由于 $F_Q(x) = \mathbb{P}\{X \in (-\infty, x]\}$，所以 $0 \le F(x) \le 1$。由 $F(x)$ 的单调性知，对任意整数 m 和 n，有

$$\lim_{x\to-\infty} F(x) = \lim_{m\to-\infty} F(m), \ \lim_{x\to+\infty} F(x) = \lim_{n\to+\infty} F(n)$$

又由概率的可列可加性得

$$\begin{aligned}
1 = P(-\infty < X < +\infty) &= P\left(\bigcup_{i=-\infty}^{+\infty} \{i-1 < X \le i\}\right) \\
&= \sum_{i=-\infty}^{+\infty} P(i-1 < X \le i) = \lim_{\substack{n\to+\infty \\ m\to-\infty}} \sum_{i=m}^{n} P(i-1 < X \le i) \\
&= \lim_{n\to+\infty} F(n) - \lim_{m\to-\infty} F(m)
\end{aligned}$$

所以：

$$1 \le \lim_{n\to+\infty} F(n) = 1 + \lim_{m\to-\infty} F(m) \le 1 \tag{2.1}$$

$$\lim_{n\to+\infty} F(n) = 1, \ \lim_{m\to-\infty} F(m) = 0 \tag{2.2}$$

即 F_Q 满足有界性。

因为 $F(x)$ 是单调有界非降函数，所以其任一点 x_0 的右极限 $F(x_0+0)$ 必存在。右连续性的充要条件为：对任意单调下降的数列 $x_1 > x_2 > \cdots > x_n > \cdots > x_0$，当 $x_n \to x_0 (n \to +\infty)$ 时，有 $\lim_{n\to+\infty} F(x_n) = F(x_0)$。因

为

$$
\begin{aligned}
F(x_1)-F(x_0)=P(x_0<X\le x_1)&=P\left(\bigcup_{r=1}^{+\infty}\{x_{i+1}<X\le x_i\}\right)\\
&=\sum_{i=1}^{+\infty}P(x_{i+1}<X\le x_i)=\sum_{i=1}^{+\infty}F(x_i)-F(x_{i+1})\\
&=\lim_{n\to\infty}[F(x_1)-F(x_n)]=F(x_1)-\lim_{n\to\infty}F(x_n)
\end{aligned}
$$

即 F_Q 满足右连续性。

笔记 由于 $\{x_1\le x_0\}\neq\bigcup_{r=1}^{+\infty}\{x_i<X\le x_{i+1}\}$，例 $(-1,0]\neq\bigcup_{r=1}^{+\infty}(\frac{-1}{i},\frac{-1}{i+1}]=(-1,0)$。若将分布函数定义中右连续性改为左连续性，只需将命题改为 $F_Q(x)=Q[(-\infty,x)],\quad\forall x\in\mathbb{R}$ 即可。

命题 2.2

对每个分布函数 F 都存在唯一一个分布 Q_F 使得 $Q_F[(-\infty,x]]=F(x),\quad\forall x\in\mathbb{R}$ 成立。

证明 参考知乎

定理 2.1

分布函数可以唯一决定概率分布，即：

$$Q_{F_Q}=Q,\quad F_{Q_F}=F$$

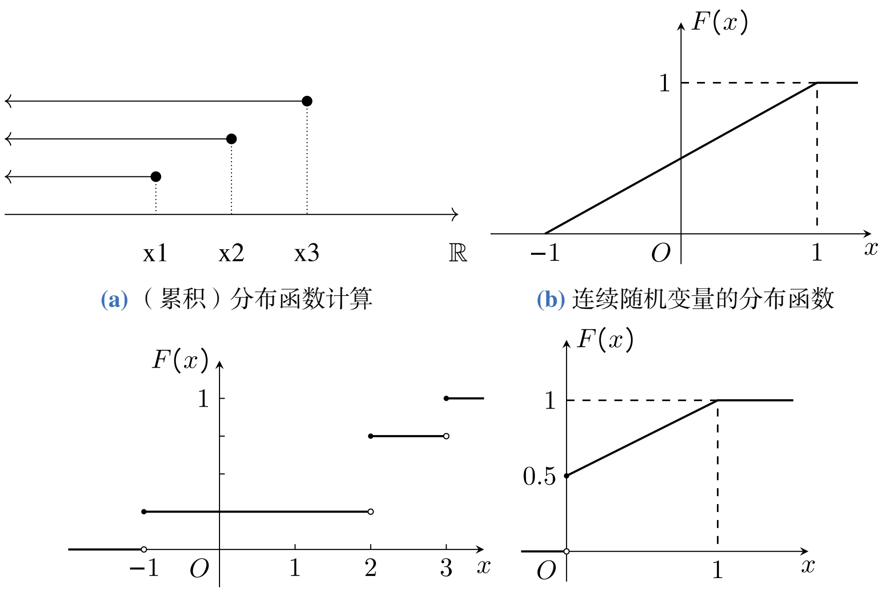
由此，概率分布与分布函数具有等同意义。所以可有以下定义：

定义 2.7 (随机变量的分布函数)

设 X 是一个随机变量，其分布由分布函数：

$$F(x)=P(X\le x),\ \forall x\in\mathbb{R}$$

唯一刻画，称为随机变量 X 的（**累积**）**分布函数** (cumulative distribution function, c.d.f.)。把随机变量 X 服从分布函数 $F(x)$，简记作 $X\sim F(x)$。



(a)（累积）分布函数计算

(b) 连续随机变量的分布函数

(c) 离散随机变量的分布函数

(d) 既非离散、又非连续的分布函数

图 2.1: 累积分布函数

利用分布函数表示概率

$$
\begin{aligned}
&P(a < X \leq b) = F(b) - F(a),\\
&P(X = a) = F(a) - F(a-0),\\
&P(X \geq b) = 1 - F(b-0),\\
&P(X > b) = 1 - F(b),\\
&P(a < x < b) = F(b-0) - F(a),\\
&P(a \leq X \leq b) = F(b) - F(a-0),\\
&P(a \leq X < b) = F(b-0) - F(a-0).
\end{aligned}
$$

定义 2.8 (随机向量的分布函数)

若 n 元函数 $F:\mathbb{R}^n \to [0,1]$ 满足:

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

则称其为随机向量 $X(\omega)$ 的**联合分布函数** (joint cdf)。

性质 多元分布函数的一些性质:

单调性 关于每个变元是单调不减函数;

有界性

$$
\begin{aligned}
F(x_1, x_2, \cdots, -\infty, \cdots, x_n) &= 0\\
F(+\infty, +\infty, \cdots, +\infty) &= 1
\end{aligned}
$$

连续性 关于每个变元右连续

$$F(x_1, x_2, \cdots, x_k + 0, \cdots, x_n) = F(x_1, x_2, \cdots, x_k, \cdots, x_n),\ \forall x_i \in \mathbb{R}$$

非负性 当 $n = 2$ 时,有

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \geq 0 \tag{2.3}$$

证明 单调性、有界性和右连续性的证明与单变量分布函数一样。对于非负性,只需证

$$P(a < X \leq b, c < Y \leq d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

为此记(见图 2.2)

$$A = \{X \leq a\},\quad B = \{X \leq b\},\quad C = \{Y \leq c\},\quad D = \{Y \leq d\}$$

考虑到

$$\{a < X \leq b\} = B - A = B \cap \overline{A},\quad \{c < Y \leq d\} = D - C = D \cap \overline{C}$$

且 $A \subset B, C \subset D$, 由此可得

$$\begin{aligned}
0 &\le P(a < X \le b, c < Y \le d) \\
&= P(B \cap \overline{A} \cap D \cap \overline{C}) \\
&= P(BD - (A \cup C)) \\
&= P(BD) - P(ABD \cup BCD) \\
&= P(BD) - P(AD \cup BC) \\
&= P(BD) - P(AD) - P(BC) + P(ABCD) \\
&= P(BD) - P(AD) - P(BC) + P(AC) \\
&= P(BD) - P(AD) - P(BC) + P(ABCD) \\
&= F(b,d) - F(a,d) - F(b,c) + F(a,c).
\end{aligned}$$

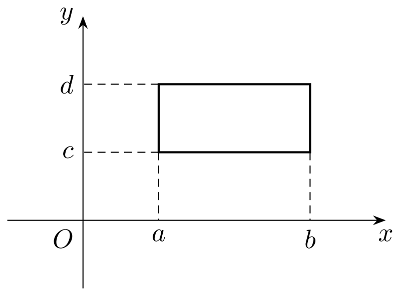


图 2.2: 二维随机变量 (X,Y) 落在矩形中的情况

非负性式2.3是二维情况特有的，不能由前三条性质推出（反例见例2.1）。同理在 $n = k$ 时，也可写出类似的等式，这是多元场合与一元场合的不同之处。

例题 2.1 设二元函数

$$G(x,y) = \begin{cases} 0, & x+y<0; \\ 1, & x+y \ge 0 \end{cases}$$

$G(x,y)$ 具有非降性、有界性和右连续性，但在正方形区域 $\{(x,y) : -1 \le x \le 1, -1 \le y \le 1\}$ 的四个顶点上，右上三个顶点位于右上半闭平面，只有左下顶点 $(-1,-1)$ 位于左下半开平面，故有：

$$G(1,1) - G(1,-1) - G(-1,1) + G(-1,-1) = 1 - 1 - 1 + 0 = -1 < 0$$

所以 $G(x,y)$ 不满足式2.3，故 $G(x,y)$ 不能成为某二维随机变量的分布函数。

定义 2.9 (概率质量函数)

若定义于至多可数点集 $\mathcal{X}$ 的函数 $p(x)$ 满足下述条件，则称其为**概率质量函数** (probability mass function, p.m.f.):

$$p(x) \ge 0$$
$$\sum_{x \in \mathcal{X}} p(x) = 1$$

当 $p(\mathbf{x})$ 是多元函数时，则称其为**联合概率质量函数** (joint probability mass function, joint p.m.f.)

设 X 为离散型随机变量，设 $\mathscr{F}_{\mathcal{X}}$ 由 $\mathcal{X}$ 的所有子集组成，此时 X 的分布由概率质量函数

$$p_X(x) = \mathbb{P}\{X = x\} = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}), \quad x \in \mathcal{X}$$

唯一刻画。其与分布函数间有如下关系：

$$F(x) = \sum_{t \le x} P(X = t) = \sum_{t \le x} p(t)$$

$$p(x) = P(X = x) = F(x) - F(x_-)$$

由于 $p(x)$ 取值也为至多可列点集，故亦称其为 X 的**概率分布列**。

若 $\mathbf{X}$ 为离散型随机向量，则只需将以上陈述改为多元情形即可。唯一不同的是由分布函数获取质量函数的方式，其应依照分布函数的非负性改为：

$$p(x, y) = P(X = x, Y = y) = F(x, y) - F(x_-, y) - F(x, y_-) + F(x_-, y_-)$$

定义 2.10 (概率密度函数)

若定义于实数集 $\mathcal{X}$ 的函数 $f(x)$ 满足下述条件，则称其为**概率密度函数** (probability density function, p.d.f.):

$$f(x) \ge 0$$

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

当 $f(\mathbf{x})$ 是多元函数时，则称其为**联合概率密度函数** (joint probability density function, joint p.d.f.)

设 X 为连续型随机变量，且其分布函数 $F(x)$ 可微，此时 X 的分布由

$$f_X(x)\Delta x = P(x < X < x + \Delta x)$$

刻画。其与分布函数间的关系为：

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

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

若 $\mathbf{X}$ 为连续型随机向量，则只需将以上陈述改为多元情形即可。唯一不同的也是由分布函数获取密度函数的方式，其应改为偏导：

$$f(\mathbf{x}) := \frac{\partial^n F(\mathbf{x})}{\partial x_1 \cdots \partial x_n}$$

离散随机变量的分布函数 $F(x)$ 总是右连续的阶梯函数，而连续随机变量的分布函数 $F(x)$ 一定是整个数轴上的连续函数。因为对任意点 x 的增量 Δx，相应分布函数的增量总有

$$F(x + \Delta x) - F(x) = \int_{x}^{x+\Delta x} p(x)\,\mathrm{d}x \longrightarrow 0, \quad (\Delta x \to 0)$$

离散随机变量 X 在其可能取值的点 $x_1, \cdots, x_n, \cdots$ 上的概率不为 0。但即使对于 $f(x) > 0$ 的 x，$P(X = x) = x\int_x^x f(t)dt = 0$，即连续型随机变量在实轴上任意一点的概率测度为零。不可能事件的概率为 0，但概率为 0 的事件不一定是不可能事件；类似地，必然事件的概率为 1，但概率为 1 的事件不一定是必然事件。概率密度函数 $f(x)$ 代表的是在此位置上单位长度的概率，可能是一个很大的值。

由于连续随机变量 X 仅取一点的概率恒为 0，从而在事件 $\{a \le X \le b\}$ 中减去 $X = a$ 或减去 $X = b$，不影响其概率，即

$$P(a \le x \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

由于在若干点上改变密度函数 $f(x)$ 的值并不影响其积分的值，从而不影响其分布函数 $F(x)$ 的值，即连续分布的密度函数不唯一。

2.1.2 边际分布与条件分布

定义 2.11 (边际分布)

对于多维随机变量 X,将其中一个分量的分布称为 X_i 的**边际分布**。对于分量 X_i,其**边际分布函数** (marginal cdf) 为:

$$F_{X_i}(x_i) = \mathbb{P}\{X_i \le x_i\} = F(\infty, \cdots, x_i, \cdots, \infty)$$

♣

定义 2.12 (边际质量函数与边际密度函数)

对于多维离散（连续）随机变量 X，将其中一个分量 X_i 对应的质量函数（密度函数）称为 X_i 的**边际质量函数（边际密度函数）**。此时，边际质量函数（边际密度函数）即可通过边际分布函数获得：

$$p_{X_i}(x_i) = F_{X_i}(x_i) - F_{X_i}(x_{i-}), \quad f_{X_i}(x_i) = \frac{\mathrm{d}F_{X_i}}{\mathrm{d}x_i}(x_i)$$

也可通过质量函数（密度函数）获得：

$$p_{X_i}(x_i) = \sum_{\mathbf{x}\in\mathcal{X}_i(x_i)} p(\mathbf{x}), \quad f_{X_i}(x_i) = \int_{\mathbf{x}\in\mathcal{X}_i(x_i)} f(\mathbf{x})\mathrm{d}\mathbf{x}$$

其中 $\mathcal{X}_i(a) = \{\mathbf{x} \in \mathcal{X} : x_i = a\}$。

♣

定义 2.13 (独立随机变量)

若随机向量 $\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega))$ 联合分布函数可分解成各分量边缘分布函数的乘积，即：

$$F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n), \quad \forall x_1, x_2, \cdots, x_n \in \mathbb{R}$$

则称随机向量 $\mathbf{X}$ 各分量相互**独立**

♣

注 对于一般的多元随机变量，其各分量边缘分布不足以描述联合分布的情况；但若其各分量独立，则可以。

定理 2.2

对于连续情况:

$$\begin{aligned} &F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \\ \Leftrightarrow &f(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) \end{aligned}$$

对于离散情况:

$$\begin{aligned} &F(x_1, x_2, \cdots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n) \\ \Leftrightarrow &p(x_1, x_2, \cdots, x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n) \end{aligned}$$

♡

注 分解函数时应注意其自变量的取值范围是否粘连。若相互粘连，则必不能分解。

设二维连续随机变量 (X, Y) 的联合密度函数为 $f(x, y)$，边际密度函数为 $p_X(x), p_Y(y)$。在离散随机变量场合，其条件分布函数为 $P(X \le x|Y = y)$。但由于连续随机变量取某个值的概率为零，即 $P(Y = y) = 0$，所以无法用条件概率直接计算 $P(X \le x|Y = y)$。一个很自然的想法是：将 $P(X \le x|Y = y)$ 看成 $h \to 0$ 时，

$P(X \le x|y \le Y \le y+h)$ 的极限（在离散情况下，此式也成立），即：

$$\begin{aligned}P(X \le x|Y = y) &= \lim_{h\to 0} P(X \le x|y \le Y \le y+h)\\ &= \lim_{h\to 0}\frac{P(X \le x, y \le Y \le y+h)}{P(y \le Y \le y+h)}\\ &= \lim_{h\to 0}\frac{\int_{-\infty}^{x}\int_{y}^{y+h} f(u,v)\mathrm{d}v\mathrm{d}u}{\int_{y}^{y+h} f_Y(v)\mathrm{d}v}\\ &= \lim_{h\to 0}\frac{\int_{-\infty}^{x}\left\{\frac{1}{h}\int_{y}^{y+h} f(u,v)\mathrm{d}v\right\}\mathrm{d}u}{\frac{1}{h}\int_{y}^{y+h} f_Y(v)dv}\end{aligned}$$

当 $f_Y(y), p(x,y)$ 在 y 处连续时，由积分中值定理可得：

$$\lim_{h\to 0}\frac{1}{h}\int_{y}^{y+h} f_Y(v)dv = f_Y(y)$$

$$\lim_{h\to 0}\frac{1}{h}\int_{y}^{y+h} f(u,v)dv = f(u,y)$$

所以

$$P(X \le x|Y = y) = \int_{-\infty}^{x}\frac{p(u,y)}{p_Y(y)}\mathrm{d}u$$

至此，可以定义随机变量的条件分布如下：

定义 2.14 (条件分布)

对一切使 $f_Y(y) > 0$ 的 y，给定 $Y = y$ 条件下 X 的**条件分布函数**和**条件密度函数**分别为

$$\begin{aligned}F(x|y) &= \int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}\mathrm{d}u,\\ f(x|y) &= \frac{f(x,y)}{f_Y(y)}.\end{aligned}$$

注 对于每一个固定的 x，$p_{Y|X}(y|x)$ 是一个关于 y 的概率质量函数；$f_{Y|X}(y|x)$ 是一个关于 y 的概率密度函数

例题 2.2 设随机变量 X 与 Y 相互独立, 且 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$. 在已知 $X + Y = n$ 的条件下, 求 X 的条件分布.

解 因为独立泊松变量的和仍为泊松变量, 即 $X + Y \sim P(\lambda_1 + \lambda_2)$, 所以

$$\begin{aligned}P(X = k|X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)}\\ &= \frac{P(X = k)P(Y = n-k)}{P(X + Y = n)}\\ &= \frac{\frac{\lambda_1^k}{k!}\mathrm{e}^{-\lambda_1}\cdot\frac{\lambda_2^{n-k}}{(n-k)!}\mathrm{e}^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}\mathrm{e}^{-(\lambda_1+\lambda_2)}}\\ &= \frac{n!}{k!(n-k)!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}\\ &= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}, \quad k = 0, 1, \ldots, n.\end{aligned}$$

即在 $X + Y = n$ 的条件下, X 服从二项分布 $b(n,p)$, 其中 $p = \lambda_1/(\lambda_1 + \lambda_2)$.

例题 2.3 设 (X, Y) 服从二维正态分布 $N(u_1, u_2, \sigma_1^2, \sigma_2^2, \rho)$，由边际分布知 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 服从正态分布。分别求 X, Y 条件分布。

解

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

这正是正态密度函数，其均值 μ_3 和方差 σ_3^2 分别为

$$
\mu_3 = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \quad \sigma_3^2 = \sigma_1^2\left(1-\rho^2\right)
$$

类似可得,在给定 $X = x$ 的条件下,Y 的条件分布仍为正态分布 $N(\mu_4, \sigma_4^2)$,其均值和方差分别为

$$
\mu_4 = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \quad \sigma_4^2 = \sigma_2^2\left(1-\rho^2\right)
$$

由此也可以看出：二维正态分布的边际分布和条件分布都是一维正态分布，这是正态分布的一个重要性质.

与概率三定理的对应：

乘法法则	$p_{XY}(x,y) = p_{Y\mid X}(y\mid x)p_X(x)$,	$f_{XY}(x,y) = f_{Y\mid X}(y\mid x)f_X(x)$
全概率公式	$p_Y(y) = \sum_x p_{Y\mid X}(y\mid x)p_X(x)$,	$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y\mid X}(y\mid x)f_X(x)dx$
Bayes 原理	$p_{X\mid Y}(x\mid y) = \frac{p_{Y\mid X}(y\mid x)p_X(x)}{\sum_x p_{Y\mid X}(y\mid x)p_X(x)}$,	$f_{X\mid Y}(x\mid y) = \frac{f_{Y\mid X}(y\mid x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y\mid X}(y\mid x)f_X(x)dx}$

例题 2.4 设在一段时间内进入某一商店的顾客人数 $X \sim P(\lambda)$，每个顾客购买某种物品的概率为 p，并且各个顾客是否购买该种物品相互独立，求进人商店的顾客购买这种物品的人数 Y 的分布列。

解 由题意知

$$
P\{X = m\} = \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda}, \quad m = 0, 1, 2, \ldots
$$

在进人商店的人数 $X = m$ 的条件下，购买某种物品的人数 Y 的条件分布为二项分布 $b(m, p)$，即

$$
P\{Y = k\mid X = m\} = \binom{m}{k}p^k(1-p)^{m-k}, \quad k = 0, 1, 2, \ldots, m
$$

由全概率公式有

$$
\begin{aligned}
P\{Y = k\} &= \sum_{m=k}^{+\infty} P\{X = m\}P\{Y = k\mid X = m\} \\
&= \sum_{m=k}^{+\infty} \frac{\lambda^m}{m!}\mathrm{e}^{-\lambda} \cdot \frac{m!}{k!(m-k)!}p^k(1-p)^{m-k} \\
&= \mathrm{e}^{-\lambda}\sum_{m=k}^{+\infty}\frac{\lambda^m}{k!(m-k)!}p^k(1-p)^{m-k} \\
&= \mathrm{e}^{-\lambda}\frac{(\lambda p)^k}{k!}\sum_{m=k}^{+\infty}\frac{[(1-p)\lambda]^{m-k}}{(m-k)!} \\
&= \frac{(\lambda p)^k}{k!}\mathrm{e}^{-\lambda p}, \quad k = 0, 1, 2, \ldots.
\end{aligned}
$$

即 Y 服从参数为 λp 的泊松分布.

这个例子告诉我们: 在直接寻求 Y 的分布有困难时, 有时借助条件分布可把困难克服了.

例题 2.5 设 $X \sim N(\mu, \sigma_1^2)$，在 $X = x$ 的条件下 $Y\mid X = x \sim N(x, \sigma_2^2)$。试求 Y 的（无条件）密度函数 $p_Y(y)$。

解 由题意知

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu)^2}{2\sigma_1^2}\right\},$$
$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-x)^2}{2\sigma_2^2}\right\}.$$

所以

$$\begin{aligned}
p_Y(y) &= \int_{-\infty}^{+\infty} p_X(x)p(y|x)\,\mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}\exp\left\{-\frac{(x-\mu)^2}{2\sigma_1^2}-\frac{(y-x)^2}{2\sigma_2^2}\right\}\mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)x^2-2\left(\frac{y}{\sigma_2^2}+\frac{\mu}{\sigma_1^2}\right)x+\frac{y^2}{\sigma_2^2}\right]\right\}\mathrm{d}x.
\end{aligned}$$

记 $c=\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$, 则上式化成

$$\begin{aligned}
p_Y(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}c^{-1}\left[x-c\left(\frac{\mu}{\sigma_1^2}+\frac{y}{\sigma_2^2}\right)\right]^2-\frac{1}{2}\frac{(y-\mu)^2}{\sigma_1^2+\sigma_2^2}\right\}\mathrm{d}x \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\sqrt{2\pi c}\exp\left\{-\frac{(y-\mu)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}\right\} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\exp\left\{-\frac{(y-\mu)^2}{2\left(\sigma_1^2+\sigma_2^2\right)}\right\}.
\end{aligned}$$

这表明 Y 仍服从正态分布 $N(\mu,\sigma_1^2+\sigma_2^2)$.

2.2 随机变量的函数

在统计学中，常需要转化原始数据以获取其中信息，由此引出了研究随机变量的函数的需要。

定理 2.3 (随机变量的函数的分布)

设 $Y=g(X)$ 是随机向量 $X=(X_1,X_2,\cdots,X_n)$ 的函数，则 Y 的分布由 X 的分布通过下式决定：

$$\mathbb{P}\{Y\in B\}=\mathbb{P}\{X\in A\},\quad A=\{\omega|g(X(\omega))\in B\}$$

此法是其他方法的基础，但使用不便，常用于离散随机变量。

例题 2.6 离散卷积公式 已知离散随机变量 X,Y 的联合概率质量函数为 $p(x,y)$, 求 $Z=X+Y$ 的分布

解

$$p_Z=P(Z=z)=P(X+Y=z)=\sum_{x=-\infty}^{\infty}p(x,z-x)$$

若 X,Y 独立，则 $p_Z(z)=\sum_{-\infty}^{\infty}p_X(x)p_Y(z-x)dx$，为 p_X 与 p_Y 的卷积。

定理 2.4 ()

若随机变量 X,Y 独立，则其变换 $Z=g(X),W=h(Y)$ 也独立。

泛化情况：若随机向量 $\{X\}_n$ 各分量独立，则其变换 $\{Y\}_n=g(\{X\}_n)$ 各分量也独立。

证明

2.2.1 分布函数法

通过下式获取随机变量的函数的分布函数：

$$F_Y(y)=\begin{cases}\int_{A_y} f_X(\mathbf{x})\mathrm{d}\mathbf{x}\\ \sum_{\mathbf{x}\in A_y} p_X(\mathbf{x})\end{cases}\qquad A_y=\{\mathbf{x}|g(\mathbf{x})\le y\}$$

对每一个变换分别运用上式则可得到向量函数的分布.

例题 2.7 已知随机变量 X 的概率密度函数 $f_X(x)$ 与分布函数 $F_X(x)$, 求 $Y=X^2$ 的分布

解 易知，若 $y<0$，则 $F_Y(y)=0, f_Y(y)=0$；若 $y\ge 0$，则：

$$\begin{aligned}F_Y(y)=P(Y\le y)&=P(-\sqrt{y}<X\le\sqrt{y})+P(X=\sqrt{y})\\&=F_X(\sqrt{y})-F_X(-\sqrt{y})+0\\ f_Y(y)=\frac{d}{dy}F_Y(y)&=\frac{d}{dy}F_X(\sqrt{y})-\frac{d}{dy}F_X(\sqrt{y})\\&=f_X(\sqrt{y})\frac{1}{2\sqrt{y}}-f_X(-\sqrt{y})\frac{-1}{2\sqrt{y}}\\&=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})]\end{aligned}$$

例题 2.8 连续变量之和 已知随机变量 X,Y 的联合概率密度函数 $f(x,y)$, 求 $Z=X+Y$ 的分布

解

$$\begin{aligned}F_Z(z)&=P(Z\le z)=P(X+Y\le z)\\&=\iint_{x+y\le z} f(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-x} f(x,y)\mathrm{d}y\mathrm{d}x\\&=\int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x,v-x)\mathrm{d}x\mathrm{d}v,\quad y=v-x\\ f_Z(z)&=\frac{d}{dz}F_Z(z)=\int_{-\infty}^{\infty} f(x,z-x)\mathrm{d}x\end{aligned}$$

若 X,Y 独立，则 $f_Z(z)=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$，为 f_X 与 f_Y 的卷积，记为 $F_X(x|\theta_X)*F_Y(y|\theta_Y)$，与例2.6类似

例题 2.9 连续变量之商 已知随机变量 X,Y 的联合概率密度函数 $f(x,y)$, 求 $Z=\frac{Y}{X}$ 的分布

解

$$Q_z=\{(x,y):y/x\le z\}=\{(x,y):x<0,y\ge zx\}\cup\{(x,y):x>0,y\le zx\}$$

$$\begin{aligned}F_Z(z)&=\iint_{Q_z} f(x,y)dxdy=\int_{-\infty}^{0}\int_{xz}^{\infty}+\int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)dydx\\&=\int_{-\infty}^{0}\int_{z}^{-\infty}+\int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx\quad(\text{set}y=xv)\\&=\int_{-\infty}^{0}\int_{-\infty}^{z}(-x)f(x,xv)dvdx+\int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx\\&=\int_{-\infty}^{z}\int_{-\infty}^{\infty}|x|f(x,xv)dxdv\\ f_Z(z)&=\frac{d}{dz}F_Z(z)=\int_{-\infty}^{\infty}|x|f(x,xz)dx\end{aligned}$$

2.2.2 Copula

定义 2.15

设连续型 实值随机变量 X 有分布函数 F, 易见 F 在 $\overline{\mathbb{R}} = [-\infty, +\infty]$ 上从 0 递增到 1. 定义相应的**分位数函数** (quantile function) 为

$$F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in [0, 1].$$

注 当 F 严格递增时, 这与一般的反函数定义相同.

定理 2.5

设连续型 实值随机变量 X 分布函数为 F，则 $F(X) \sim \mathrm{U}(0,1)$。若设 $U \sim \mathrm{U}([0,1])$，则 $F^{-1}(U) \overset{d}{=} X$, 其中 $\overset{d}{=}$ 表示分布相同 (equal in distribution)。

证明

$$\mathbb{P}\{F(X) \leq p\} = \mathbb{P}\{X \leq F^{-1}(p)\} = F(F^{-1}(p)) = p, \quad \forall p \in [0, 1].$$

注 设随机变量 $Y = h(U)$，然而 $h^{-1}(y)$ 未必等于其分布函数 $F_Y(y)$，只有当 $h^{-1}(y)$ 满足分布函数条件时两者才相等。

2.2.3 概率密度函数法

定理 2.6 (单变量函数的概率函数变换)

设连续随机变量 X 的概率密度函数为 $f_X(x)$，若随机变量 $Y = g(X)$，其中 g 为可微函数，且严格单调，则当 $y = g(x)$ 有定义时，其概率密度函数为：

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \mathbb{1}_{(a,b)}(x)$$

其中 $a = \min\{g(-\infty), g(+\infty)\}, b = \max\{g(-\infty), (+\infty)\}$。
若 g 为分段单调函数，则计算单调分段 的以上结果，再进行相加。

证明 不妨先设 $g(x)$ 是严格单调增函数，这时它的反函数 $h(y)$ 也是严格单调增函数，且 $h(y) > 0$。记 $a = g(-\infty), b = g(+\infty)$，这意味着 $y = g(x)$ 仅在区间 (a, b) 取值，于是：

- 当 $y < a$ 时，
$$F_Y(y) = P(Y \leq y) = 0$$
- 当 $y > b$ 时，
$$F_Y(y) = P(Y \leq y) = 1$$
- 当 $a \leq y \leq b$ 时，
$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P(X \leq h(y)) = \int_{-\infty}^{h(y)} p(x)\,\mathrm{d}x. \end{aligned}$$

由此得 Y 的密度函数为

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \mathbb{1}_{(a,b)}(x)$$

同理可证当 $g(x)$ 是严格单调减函数时，结论也成立。但此时 $h'(y) < 0$，须加绝对值符号，这时 $a = g(+\infty), b = g(-\infty)$。

定理 2.7 (多变量函数的概率函数变换)
设连续随机向量 $\mathbf{X}$ 的概率密度函数为 $f_{\mathbf{X}}(\mathbf{x})$. 令 $\mathbf{Y}=\mathbf{g}(\mathbf{X})$, 其中 $\mathbf{g}$ 为双射, 定义其逆函数为:

$$\mathbf{g}^{-1}(\mathbf{y})=\mathbf{w}(\mathbf{y})=\mathbf{x}$$

若 $\mathbf{w}$ 存在连续偏导数, 则当 $\mathbf{Y}=\mathbf{g}(\mathbf{X})$ 有定义时:

$$f_{\mathbf{Y}}(\mathbf{y})=f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))\left|\frac{\partial\mathbf{w}}{\partial\mathbf{y}}\right|$$

否则为 0

注 若 $\mathbf{Y}$ 的维数 k 小于 $\mathbf{X}$ 的维数 n, 可增补 $n-k$ 维的函数 $\mathbf{Z}=\mathbf{h}(\mathbf{X})$, 使得 $(\mathbf{Y},\mathbf{Z})$ 满足条件, 再通过对 $\mathbf{Z}$ 积分获取 $\mathbf{Y}$ 的概率密度函数.

例题 2.10 连续变量之积 已知随机变量 X,Y 的联合概率密度函数 $f(x,y)$, 求 $Z=XY$ 的分布

解 令 $V=Y$，则：

$$X=\frac{Z}{V}\equiv w_1(z,v)$$
$$Y=V\equiv w_2(z,v)$$
$$J=\begin{vmatrix}\frac{1}{v} & -\frac{z}{v^2}\\ 0 & 1\end{vmatrix}=\frac{1}{v}$$

所以:

$$f_{Z,V}(z,v)=f_{X,Y}(\frac{z}{v},v)\left|\frac{1}{v}\right|$$
$$f_Z(z)=\int_{-\infty}^{\infty}f_{X,Y}(\frac{z}{v},v)\left|\frac{1}{v}\right|\mathrm{d}v$$

命题 2.3
若随机向量 $\mathbf{X}=(X_1,\cdots,X_n)$ 各分类相互独立, 则其各分量的函数 $\mathbf{Y}=(g_1(X_1),\cdots,g_n(X_n))$ 也相互独立

证明

2.2.4 矩母函数法

2.2.5 次序统计量

定义 2.16
设 $X_1,X_2,...,X_n$ 为随机变量，将其按大小排序后记为 $X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$，称 $x_{(i)}$ 为该样本的第 i 个**次序统计量**。并且：

- **最小次序统计量**定义为：$X_{(1)}=\min\{x_1,...,x_n\}$
- **最大次序统计量**定义为：$X_{(n)}=\max\{x_1,...,x_n\}$
- **极差**定义为：$R=X_{(n)}-X_{(1)}$
- 第 i 个**间差**定义为：$S_i=X_{(i)}-X_{(i-1)}$

注 虽然 $X_1,X_2,...,X_n$ 独立，但其次序统计量一般不独立

例题 2.11 若 $X_1,X_2,...,X_n$ 独立同分布，求 $X_{(1)}$ 与 $X_{(n)}$ 的分布

解

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \le x) = P(X_1 \le x, \dots, X_n \le x) \\ &= P(X_1 \le x) \cdots P(X_n \le x) \\ &= [F(x)]^n \end{aligned}$$

$$f_{X_{(n)}}(x) = \frac{d}{dx} F_{X_{(n)}}(x) = n f(x) [F(x)]^{n-1}$$

$$\begin{aligned} 1 - F_{X_{(1)}}(x) &= P(X_{(1)} > x) = P(X_1 > x, \dots, X_n > x) \\ &= P(X_1 > x) \cdots P(X_n > x) \\ &= [1 - F(x)]^n \end{aligned}$$

$$f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = n f(x) [1 - F(x)]^{n-1}$$

命题 2.4

若 $X_1, X_2, \dots, X_n$ 独立同分布，第 k 个**次序统计量**的概率密度函数为：

$$f_{X_{(k)}} = \frac{n!}{(k-1)!(n-k)!} f(x) [F(x)]^{k-1} [1 - F(x)]^{n-k}$$

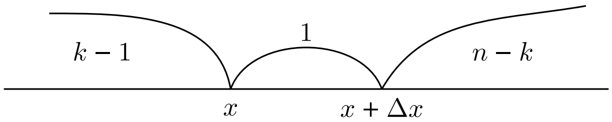


图 2.3: $X_{(k)}$ 取值的示意图

命题 2.5

若 $X_1, X_2, \dots, X_n$ 独立同分布，则次序统计量 $(x_{(i)}, x_{(j)})(i < j)$ 的联合分布密度函数为：

$$f_{ij}(y, z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(y)]^{i-1} [F(z) - F(y)]^{j-i-1} [1 - F(z)]^{n-j} f(y) f(z),\ y \le z$$

证明 对正 $\Delta y, \Delta z$ 以及 $y < z$, 事件 “$x_{(i)} \in (y, y + \Delta y], x_{(j)} \in (z, z + \Delta z]$” 可以表示为 “容量为 n 的样本 $x_1, \dots, x_n$ 中有 $i-1$ 个观测值小于等于 y, 一个落入区间 $(y, y + \Delta y]$, $j - i - 1$ 个落入区间 $(y + \Delta y, z]$, 一个落入区间 $(z, z + \Delta z]$, 而余下 $n - j$ 个大于 $z + \Delta z$”(见图 2.4).

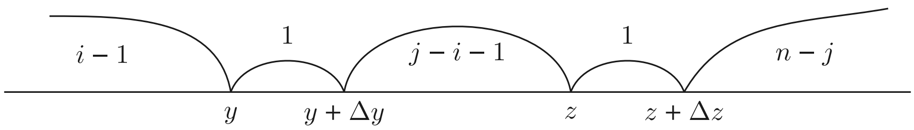


图 2.4: $x_{(i)}$ 与 $x_{(j)}$ 取值的示意图

例题 2.12 若 $X_1, X_2, \dots, X_n$ 独立同分布, 求 $R = X_{(n)} - X_{(1)}$ 的分布

解

$$f_{X_{(1)}X_{(n)}}(s, t) = n(n-1) f(s) f(t) [F(t) - F(s)]^{n-2} \mathbb{I}(s \le t)$$

$$f_R(r) = \mathbb{I}(r > 0) \int_{-\infty}^{\infty} f_{X_{(1)}X_{(n)}}(s, s + r) ds$$

错题记录

1. (茆 2.6.5) 设随机变量 X 服从 $(-\pi/2, \pi/2)$ 上的均匀分布，求随机变量 $Y = \cos X$ 的密度函数 $p_Y(y)$.

2. （茆 3.1.1）100 件产品中有 50 件一等品，30 件二等品，20 件三等品。从中抽取 5 件，以 X、Y 分别表示取出的 5 件中一等品、二等品的件数，在以下情况下求 (X,Y) 的联合分布列：(1) 不放回抽取；（2）有放回抽取。
3. （茆 3.1.7）设二维随机变量 (X,Y) 的联合密度函数为
$$p(x,y)=\begin{cases}4xy, & 0<x<1,0<y<1,\\ 0, & \text{其他}.\end{cases}$$
试求
(1) $P(X<Y)$;
(2) (X,Y) 的联合分布函数.
4. （茆 3.1.12）设二维随机变量 (X,Y) 的联合密度函数为
$$p(x,y)=\begin{cases}x^2+\frac{xy}{3}, & 0<x<1,0<y<2;\\ 0, & \text{其他}.\end{cases}$$
求 $P(X+Y)\geq 1$.
5. （茆 3.1.16）设二维随机变量 (X,Y) 的联合分布函数为 $F(x,y)$，试用 $F(x,y)$ 表示 $P(X=a,Y>b)$
6. （茆 3.3.4）设随机变量 X,Y 独立且都服从几何分布，即 $P(X=k)=(1-p)^{k-1}p,k=1,2,\dots$ 求随机变量 $Z=\max\{X,Y\}$ 的分布列.
7. （茆 3.3.14）设二维随机变量 (X,Y) 在矩形
$$G=\{(x,y)|0\leq x\leq 2,0\leq y\leq 1\}$$
上服从均匀分布, 试求边长分别为 X 和 Y 的矩形面积 Z 的密度函数.
8. （茆 3.3.17）设 $X,Y\overset{\text{i.i.d.}}{\sim}N(0,1)$，试证明:$U=X^2+Y^2$ 与 $V=\frac{X}{Y}$ 相互独立。

第 3 章 随机变量的数值特征

考试重点

- 重期望公式
- 重方差公式
- 切比雪夫不等式

定义 3.1 (期望)

对于实值随机向量 $X:(\Omega,\mathscr{F},\mathbb{P})\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ 和 (可测) 函数 $g:\mathbb{R}^n\to\mathbb{R}$，若

$$\mathbb{E}[g(X)]=\int_{\mathbb{R}}g(x)\mathrm{d}F_X(x)$$

存在，则称其为 $g(X)$ 的**期望** (expectation)。

注 当 $F_X(x)$ 在 x_0 初连续可导时，$\mathrm{d}F_X(x_0)=f_X(x_0)\mathrm{d}x$；当 x_0 为其间断点时时，$\mathrm{d}F_X(x_0)=p_X(x_0)\delta(x_0)\mathrm{d}x$。

命题 3.1 (期望的线性性质)

由积分的线性性质可得：均值为随机变量的线性映射，即

$$\mathbb{E}(a+\sum_{i=1}^{n}b_ig(X_i))=a+\sum_{i=1}^{n}b_i\mathbb{E}(g(X_i))$$

命题 3.2 (独立变量的期望)

若 X,Y 独立，则

$$\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$$
$$\mathbb{E}(g(X)h(Y))=\mathbb{E}(g(X))\mathbb{E}(h(Y))$$

注 由于 $\mathbb{E}(X/Y)=\mathbb{E}(X)\mathbb{E}(\frac{1}{Y})$, 而 $\mathbb{E}(\frac{1}{Y})\neq\frac{1}{\mathbb{E}(Y)}$, 所以 $\mathbb{E}(X/Y)\neq\mathbb{E}(X)/\mathbb{E}(Y)$

3.1 均值与方差

定义 3.2

1. 当 $g(x)=x$ 时，$\mathbb{E}[g(x)]=\mathbb{E}[X]$ 称作 X 的**均值** (mean)，记为 μ_X
2. 当 $g(x)=(x-\mu_X)^2$ 时，$\mathbb{E}[g(x)]=\mathbb{E}[(X-\mathbb{E}[X])^2]$ 称作 X 的**方差** (variance)，记为 σ_X^2。其平方根称作 X 的**标准差** (standard deviation)，记为 σ_X
3. 当 $g(x,y)=(x-\mu_X)(y-\mu_Y)$ 时，$\mathbb{E}[g(X,Y)]=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$ 称作 X 与 Y 的**协方差** (covariance)，记为 $\mathrm{Cov}(X,Y)$ 或 σ_{XY}

3.1.1 均值

随机变量的均值可看作其加权平均，权重为其 pdf 或 pmf，结果为其函数图像上的形心。从大数定律 (第5.2节) 的角度看，也可解释为其长期均值。

例题 3.1 在一个人数为 N 的人群中普查某种疾病，为此要抽验 N 个人的血。如果将每个人的血分别检验，则共需检验 N 次。为了能减少工作量，一位统计学家提出一种方法：按 k 个人一组进行分组，把同组 k 个人的血样混合后检验，如果这混合血样呈阴性反应，就说明此 k 个人的血都呈阴性反应，此 k 个人都无此疾病，因而这 k

个人只要检验 1 次就够了，相当于每个人检验 $1/k$ 次，检验的工作量明显减少了。如果这混合血样呈阳性反应，就说明此 k 个人中至少有一人的血呈阳性反应，则再对此 k 个人的血样分别进行检验，因而这 k 个人的血要检验 $1+k$ 次，相当于每个人检验 $1+1/k$ 次，这时增加了检验次数。假设该疾病的发病率为 p，且得此疾病相互独立。试问此种方法能否减少平均检验次数?

解 令 X 为该人群中每个人需要的验血次数，则 X 的分布列为 所以每人平均验血次数为

x	$1/k$	$1+1/k$
P	$(1-p)^k$	$1-(1-p)^k$

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

由此可知，只要选择 k 使

$$1-(1-p)^k+\frac{1}{k}<1,\ \text{即},\ (1-p)^k>\frac{1}{k}$$

就可减少验血次数，而且还可适当选择 k 使其达到最小。

例题 3.2 凑一法求解二项分布的均值 求解二项分布（定义4.3）的均值时，可通过将其凑成概率质量函数之和（为 1）的形式，称此为**凑一法** (sum to one, STO)：

$$\begin{aligned}E(X)&=\sum_{i=1}^{n}x\binom{n}{x}p^x(1-p)^{n-x}\\&=np\sum_{i=1}^{n}\binom{n-1}{x-1}p^{x-1}(1-p)^{n-1-(x-1)}\\&=np\end{aligned}$$

例题 3.3 微分法求解几何分布均值

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

命题 3.3

设 X 为仅取非负整数的离散随机变量，若其数学期望存在，则

$$E(X)=\sum_{k=1}^{+\infty}P(X\geq k)$$

3.1.2 方差

方差为随机变量距其均值的均方偏差，刻画了 X 的变动程度。随机变量的均值与标准差的单位和其本身的相同，方差的为其平方。如果随机变量 X 的数学期望存在, 其方差不一定存在；而当 X 的方差存在时，则 $E(X)$ 必定存在，其原因在于 $|x|\leq x^2+1(0\leq x^2-2|x|+1)$ 总是成立的（若 $E(X^2)$ 收敛，则 $E(|X|)$ 收敛）。

命题 3.4

$$\mathrm{Var}(X)=\mathbb{E}[(X-\mu_X)^2]=E(X^2)-\mu_X$$

例题 3.4 凑一法求解二项分布的方差

$$
\begin{aligned}
E(X(X-1)) &= \sum_{i=1}^{n} x(x-1)\binom{n}{x}p^{x}(1-p)^{n-x} \\
&= n(n-1)p^{2}\sum_{i=1}^{n}\binom{n-2}{x-2}p^{x-2}(1-p)^{n-1-(x-1)} \\
&= n(n-1)p^{2} \\
\mathrm{Var}(X) &= E(X(X-1)) + E(X) - (E(X))^{2} \\
&= n(n-1)p^{2} + np - n^{2}p^{2} \\
&= np(1-p)
\end{aligned}
$$

例题 3.5 微分法求解几何分布方差

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

命题 3.5

$$
\mathrm{Var}(a+bX) = b^{2}\,\mathrm{Var}(X)
$$

命题 3.6

$$
\mathrm{Var}(a+\sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} b_i^{2}\,\mathrm{Var}(X_i) + \mathbf{b}^{\mathrm{T}}\Sigma\mathbf{b}
$$

其中 Σ 为协方差矩阵,$\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$

证明 由方差的定义知

$$
\begin{aligned}
\mathrm{Var}(X \pm Y) &= E\{(X \pm Y) - E(X \pm Y)\}^{2} \\
&= E\{(X - E(X)) \pm (Y - E(Y))\}^{2} \\
&= E\{(X - E(X))^{2} + (Y - E(Y))^{2} \pm 2(X - E(X))(Y - E(Y))\} \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)
\end{aligned}
$$

推论 3.1

若 $X_1, \cdots, X_n$ 相互不相关，则：

$$
\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\mathrm{Var}(X_i)
$$

定理 3.1 (Chebyshev 不等式)

设随机变量 X 的均值与方差分别为:μ,σ^2, 则:

$$\mathbb{P}(|X-\mu|>t)\le\frac{\sigma^2}{t^2}$$

证明 设 $F(x)$ 为 X 的分布函数, 令 $R=\{x:|x-\mu|>t\}$

$$\begin{aligned}\mathbb{P}(|x-\mu|>t)&=\int_R 1\cdot \mathrm{d}F(x)\le\int_R\frac{(x-\mu)^2}{t^2}\mathrm{d}F(x)\\&\le\int_{-\infty}^{\infty}\frac{(x-\mu)^2}{t^2}\mathrm{d}F(x)=\frac{\sigma^2}{t^2}\end{aligned}$$

再由协方差的运算性质（命题3.9），通过数学归纳法可推出泛化的结果。

注 若令 $t=k\sigma$, 则 $\mathbb{P}(|X-\mu|>k\sigma)\le\frac{1}{k^2}$, 即标准差可代表随机变量偏离均值的概率单位距离.

推论 3.2

$$\mathrm{Var}(X)=0\Leftrightarrow P(X=\mu)=1$$

预处理随机变量有两个常用变换:

- **中心化** (centralization)$X\mapsto X-\mathbb{E}X$;
- **标准化** (standardization)$X\mapsto\dfrac{X-\mathbb{E}X}{\sqrt{\mathrm{Var}(X)}}$.

3.1.3 协方差

协方差代表了 X 与 Y 之间的联合变化倾向，或者说他们间的相关程度，但其间未必有 因果关系。

- 若 $\mathrm{Cov}(X,Y)>0$，则称两变量**正相关**；
- 若 $\mathrm{Cov}(X,Y)<0$，则称两者**负相关**；
- 若 $\mathrm{Cov}(X,Y)=0$，则称两者**不相关**。

注 不相关是指 X 与 Y 之间没有线性关系，但 X 与 Y 之间可能有其他的函数关系，譬如平方关系、对数关系等。

命题 3.7

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

命题 3.8

独立是不相关的**充分条件**，但不是必要条件。

例题 3.6 设随机变量 $X\sim N(0,\sigma^2)$，令 $Y=X^2$，则 X 与 Y 不独立。此时 X 与 Y 的协方差为

$$\mathrm{Cov}(X,Y)=\mathrm{Cov}(X,X^2)=E(X\cdot X^2)-E(X)E(X^2)$$

由于正态分布 $N(0,\sigma^2)$ 的奇数阶原点矩均为零，即 $E(X)=E(X^3)=0$，故上式也等于零。

例子表明：“独立”必导致“不相关”；而“不相关”不一定导致“独立”。因为独立性是用分布定义的，而不相关只是用矩定义的。

命题 3.9

协方差的运算性质：

- 协方差 $\mathrm{Cov}(X,Y)$ 的计算与 X,Y 的次序无关

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

- 任意随机变量 X 与常数 a 的协方差为零

$$\mathrm{Cov}(X,a)=0$$

- 对任意常数 a,b, 有

$$\mathrm{Cov}(aX,bY)=ab\,\mathrm{Cov}(X,Y)$$

- 设 X,Y,Z 是任意三个随机变量, 则

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

推论 3.3

线性变换后的协方差:

$$\mathrm{Cov}(a+bX,c+dY)=bd\,\mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}(a+\sum_{i=1}^{n}b_iX_i,c+\sum_{j=1}^{m}d_jY_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}b_id_j\,\mathrm{Cov}(X_i,Y_i)=\mathbf{b}^{\mathrm{T}}\Sigma\mathbf{d}$$

定义 3.3 (协方差矩阵)

对于 n 维随机向量 $\mathbf{X}=(X_1,X_2,\ldots,X_n)'$，将

$$E[(\mathbf{X}-E(\mathbf{X}))(\mathbf{X}-E(\mathbf{X}))']$$
$$=\begin{pmatrix}\mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \mathrm{Var}(X_n)\end{pmatrix}$$

称为该随机向量的**方差—协方差阵**，简称**协方差阵**，记为 $\Sigma=\mathrm{Cov}(\mathbf{X})$。

定理 3.2

n 维随机向量的协方差阵 $\mathrm{Cov}(\mathbf{X})=(\mathrm{Cov}(X_i,X_j))_{n\times n}$ 是一个对称的非负定矩阵。

证明 因为 $\mathrm{Cov}(X_i,X_j)=\mathrm{Cov}(X_j,X_i)$，所以对称性是显然的。下证非负定性。因为对任意 n 维实向量 $\mathbf{c}=$

$(c_1, \cdots, c_n)'$，有

$$\begin{aligned}
\mathbf{c}' \operatorname{Cov}(\mathbf{X})\mathbf{c} &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \operatorname{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E\{[c_i(X_i - E(X_i))][c_j(X_j - E(X_j))]\} \\
&= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}[c_i(X_i - E(X_i))][c_j(X_j - E(X_j))]\right\} \\
&= E\left\{\left[\sum_{i=1}^{n} c_i(X_i - E(X_i))\right]\left[\sum_{j=1}^{n} c_j(X_j - E(X_j))\right]\right\} \\
&= E\left[\sum_{i=1}^{n} c_i(X_i - E(X_i))\right]^2 \geq 0
\end{aligned}$$

所以矩阵 $\operatorname{Cov}(\mathbf{X})$ 是非负定的。

3.1.4 相关系数

定义 3.4

设 (X, Y) 是一个二维随机变量, 且 $\operatorname{Var}(X) > 0, \operatorname{Var}(Y) > 0$。则称

$$\operatorname{Corr}(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

为 X 与 Y 的（线性）**相关系数** (correlation coefficient)，可简记为 ρ_{XY}。

相关系数是相应标准化变量的协方差。若记 X 与 Y 的数学期望分别为 μ_X, μ_Y，其标准化变量为:

$$X^* = \frac{X - \mu_X}{\sigma_X}, \quad Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

则有

$$\operatorname{Cov}(X^*, Y^*) = \operatorname{Cov}\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y} = \operatorname{Corr}(X, Y)$$

命题 3.10

平移与缩放随机变量都不影响其相关系数的绝对值，但其是否变号取决于缩放系数的正负，即:

$$\operatorname{Corr}(a + bX, c + dY) = \operatorname{Corr}(X, Y)\operatorname{sign}(bd)$$

例题 3.7 二维正态分布 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 的相关系数是 ρ。

解 由上述可知，平移与缩放变量不影响两者间相关系数，故不妨将变量标准化:

$$X^* = \frac{X - \mu_X}{\sigma_X}, \quad Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

则其密度函数变为:

$$f_{X^*, Y^*}(x, y) = \sigma_X \sigma_Y f_{X,Y}(\sigma_X x + \mu_X, \sigma_Y y + \mu_Y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right)$$

$$E(X^* Y^*) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) \mathrm{d}x\mathrm{d}y$$

做变换:

$$\begin{cases} u = \frac{x - \rho y}{\sqrt{1-\rho^2}} \\ v = y \end{cases} \Longrightarrow \begin{cases} x = \sqrt{1-\rho^2}u + \rho v \\ y = v \end{cases} \Longrightarrow \mathrm{d}x\,\mathrm{d}y = \sqrt{1-\rho^2}\,\mathrm{d}u\,\mathrm{d}v$$

由此得

$$\mathrm{Cov}(X,Y)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(uv\sqrt{1-\rho^2}+\rho v^2)\exp\left\{-\frac{u^2+v^2}{2}\right\}\mathrm{d}u\,\mathrm{d}v$$

上式右端积分可以分为两个积分，其中

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}uv\exp\left(-\frac{u^2+v^2}{2}\right)\mathrm{d}u\,\mathrm{d}v=0,$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}v^2\exp\left(-\frac{u^2+v^2}{2}\right)\mathrm{d}u\,\mathrm{d}v=2\pi.$$

从而

$$\mathrm{Cov}(X^*,Y^*)=E(X^*Y^*)=\frac{1}{2\pi}\cdot\rho\cdot 2\pi=\rho$$

$$\mathrm{Corr}(X,Y)=\mathrm{Corr}(X^*,Y^*)=\frac{\mathrm{Cov}(X,Y)}{1\cdot 1}=\rho$$

引理 3.1 (施瓦茨不等式)

对任意二维随机变量 (X,Y)，若 X 与 Y 的方差都存在，且记 $\sigma_X^2=\mathrm{Var}(X),\sigma_Y^2=\mathrm{Var}(Y)$，则有

$$[\mathrm{Cov}(X,Y)]^2\le\sigma_X^2\sigma_Y^2$$

证明 不妨设 $\sigma_X^2>0$，因为当 $\sigma_X^2=0$ 时，则 X 几乎处处为常数，因而其与 Y 的协方差亦为零，从而上式两端皆为零，结论成立. 在 $\sigma_X^2>0$ 成立下，考虑 t 的如下二次函数:

$$g(t)=E[t(X-E(X))+(Y-E(Y))]^2=t^2\sigma_X^2+2t\cdot\mathrm{Cov}(X,Y)+\sigma_Y^2$$

由于上述的二次三项式非负，平方项系数 σ_X^2 为正，所以其判别式小于或等于零，即

$$[2\,\mathrm{Cov}(X,Y)]^2-4\sigma_X^2\sigma_Y^2\le 0$$

移项后即得施瓦茨不等式.

命题 3.11

$-1\le\rho_{XY}\le 1$。取等号的充要条件是 X 与 Y 间几乎处处有线性关系，即存在 $a(\neq 0)$ 与 b，使得:

$$P(Y=aX+b)=1$$

上式换成 $X=cY+d$ 也一样

证明 **充分性**：若 $Y=aX+b$，则有

$$\mathrm{Var}(Y)=a^2\,\mathrm{Var}(X),\mathrm{Cov}(X,Y)=a\,\mathrm{Cov}(X,X)=a\,\mathrm{Var}(X)$$

代入相关系数的定义中得

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{a\,\mathrm{Var}(X)}{|a|\,\mathrm{Var}(X)}=\begin{cases}1, & a>0;\\ -1, & a<0.\end{cases}$$

必要性：因为

$$\mathrm{Var}\left(\frac{X}{\sigma_X}\pm\frac{Y}{\sigma_Y}\right)=2[1\pm\mathrm{Corr}(X,Y)]$$

所以当 $\mathrm{Corr}(X,Y)=1$ 时，有

$$\mathrm{Var}\left(\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y}\right)=0$$

由此得（推论3.2）

$$P\left(\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y}=c\right)=1$$

$$P\left(Y=\frac{\sigma_Y}{\sigma_X}X-c\sigma_Y\right)=1$$

即当 $\mathrm{Corr}(X,Y)=1$ 时,Y 与 X 几乎处处线性正相关.

当 $\mathrm{Corr}(X,Y)=-1$ 时, 由同理得

$$P\left(Y=-\frac{\sigma_Y}{\sigma_X}X+c\sigma_Y\right)=1$$

即当 $\mathrm{Corr}(X,Y)=-1$ 时,Y 与 X 几乎处处线性负相关.

3.2 不等式

3.2.1 Chebyshev-Markov 型不等式

定理 3.3 (Markov 不等式)

设 $g(x)$ 为随机变量 X 取值的集合上的非负不减函数，且 $E(g(X))$ 存在，则

$$P(X>\varepsilon)\le\frac{E(g(X))}{g(\varepsilon)},\quad\forall\varepsilon>0$$

证明

$$\begin{aligned}P(X>\varepsilon)&=\int_\varepsilon^{+\infty}\mathrm{d}F(x)\le\int_\varepsilon^{+\infty}\frac{g(X)}{g(\varepsilon)}\mathrm{d}F(x)\\&\le\int_{-\infty}^{+\infty}\frac{g(X)}{g(\varepsilon)}\mathrm{d}F(x)=\frac{E(g(X))}{g(\varepsilon)}\end{aligned}$$

推论 3.4

若 $g(X)\ge0$，则

$$\mathbb{E}[g(X)]=0\Longleftrightarrow\mathbb{P}\{g(X)=0\}=1\Longleftrightarrow g(X)\overset{\text{a.s.}}{=}0$$

命题 3.12 (Chebyshev 不等式)

$$P\{|X-E(X)|\ge x\}\le\frac{\mathrm{Var}(X)}{x^2}$$

证明 令 Markov 不等式（定理3.3）中的 $g(x)=x^2$ 则可得：

$$P\{|X-E(X)|\ge x\}=P\{[X-E(X)]^2\ge x^2\}\le\frac{E[X-E(X)]^2}{x^2}=\frac{\mathrm{Var}(X)}{x^2}$$

命题 3.13 (Markov 不等式)

$$P\{|X|\ge x\}\le\frac{E(|X|^r)}{x^r},\quad\forall r>0,x>0$$

证明 令 Markov 不等式（定理3.3）中的 $g(x)=x^r$ 则可得：

$$P\{|X|\ge x\}=P\{|X|^r\ge x^r\}\le\frac{E(|X|^r)}{x^r}$$

命题 3.14 (Chebyshev 不等式的推广)

$$P\{X - E(X) \geq x\} \leq \frac{\sigma^2 + a^2}{(x+a)^2}, \quad \forall x, a$$

证明 令 Markov 不等式（定理3.3）中的 $g(x) = x$ 则可得：

$$\begin{aligned} P\{X - E(X) \geq x\} &= P\{X - E(X) + a \geq x + a\} \\ &\leq P\{[X - E(X) + a]^2 \geq (x+a)^2\} \\ &\leq \frac{E[X - E(X) + a]^2}{(x+a)^2} = \frac{\sigma^2 + a^2}{(x+a)^2} \end{aligned}$$

例题 3.8 单边 Chebyshev 不等式 设随机变量 X 的均值为 μ，方差为 σ^2，则对于任意 $x > 0$ 有：

$$P\{X - \mu \geq x\} \leq \frac{\sigma^2}{x^2 + \sigma^2}$$

解 令 Chebyshev 不等式的推广形式（命题3.14）中的 $a = \frac{\sigma^2}{x}$，则

$$\frac{\sigma^2 + a^2}{(x+a)^2} = \frac{\sigma^2 + \frac{\sigma^4}{x^2}}{x^2 + 2\sigma^2 + \frac{\sigma^4}{x^2}} = \frac{\sigma^2}{x^2 + \sigma^2}$$

命题 3.15 (Chebyshev 不等式的离散情形)

设离散随机变量 X 的取值范围为正自然数，且 $P\{X - k\}$ 非增，则

$$P\{X = k\} < \frac{2}{k^2} E(X)$$

证明

$$\begin{aligned} E(X) &= \sum_{i=1}^{\infty} iP\{X = i\} \geq \sum_{i=1}^{k} iP\{X = i\} \\ &\geq \sum_{i=1}^{k} iP\{X = k\} = P\{X = k\}\frac{k(k+1)}{2} \\ &\geq \frac{k^2}{2} P\{X = k\} \end{aligned}$$

3.2.2 Cauchy-Schwarz 不等式

命题 3.16 (Cauchy-Schwarz 不等式)

$$E(|XY|) \leq [E(X^2)]^{\frac{1}{2}}[E(Y^2)]^{\frac{1}{2}}$$

证明 令 $g(t) = E[|X| + t|Y|]^2$，易知其恒大于 0。又

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

则其判别式

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

命题 3.17 (Holder 不等式)

若 p, q 满足 $p, q > 1$，$\frac{1}{p} + \frac{1}{q} = 1$，则

$$E(|XY|) \leq [E(X^p)]^{\frac{1}{p}}[E(Y^q)]^{\frac{1}{q}}$$

3.2.3 Jensen 不等式

定理 3.4 (Jensen 不等式)

若 $g: R \to R$ 为下凸 (convex) 函数，则 $g[\mathbb{E}(X)] \leq \mathbb{E}[g(X)]$；若 g 为上凸 (concave) 函数，则 $\mathbb{E}[g(X)] \leq g[\mathbb{E}(X)]$。

证明 若 g 为下凸函数，则

$$g(x) \geq g(x_0) + g'(x - x_0)$$

令 $x_0 = E(X)$，得

$$g(x) \geq g(E(X)) + g'(x - E(X))$$

再对两边取期望，得

$$E(g(X)) \geq g(E(X)) + E[g'(X - E(X))] = g(E(X))$$

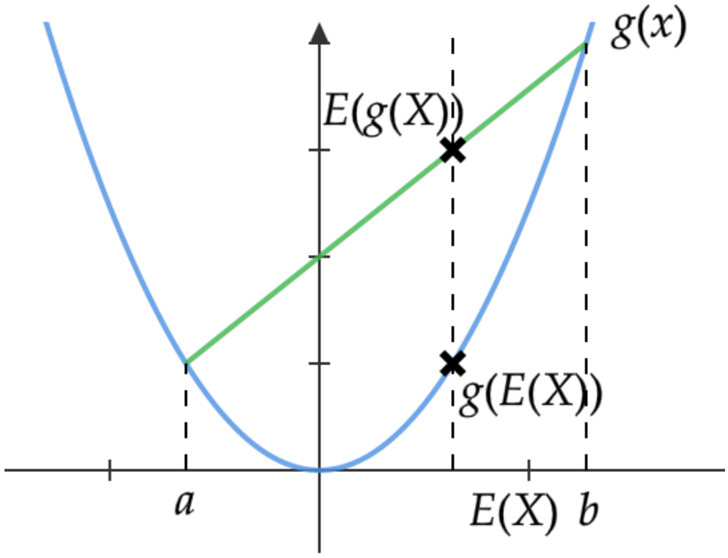


图 3.1: 设随机变量 X 遵循两点分布，取值为 $\{a, b\}$，此为其变换后各值

引理 3.2

$$(E|X|^r)^{\frac{1}{r}} \leq (E|X|^s)^{\frac{1}{s}}, \quad \forall 0 < r \leq s$$

证明 令 $g(x) = x^{\frac{s}{r}}$ 且 $Y = |X|^r$ 得

$$E^{\frac{s}{r}}(Y) \leq E(Y^{\frac{s}{r}})$$

即

$$E^{\frac{s}{r}}(|X|^r) \leq E(|X|^s)$$

3.2.4 最小一乘法

命题 3.18

设随机变量 X 的均值 为 μ，则对任意常数 c，有

$$E(X-\mu)^2 \leq E(X-c)^2$$

证明 将 $E(X-c)^2 = E(X-\mu+\mu-c)^2$ 展开即可

命题 3.19

设随机变量 X 的**中位数**为 m，则对任意常数 c，有

$$E|X-m| \leq E|X-c|$$

证明

命题 3.20

若 $x_1, \cdots, x_n$ 是总体 X 的一组样本，其中 x_{med} 为样本中位数，$\overline{x}$ 是样本均值。则对任意常数 c 有：

- $\sum_{i=1}^{n}(x_i-\overline{x})^2 \leq \sum_{i=1}^{n}(x_i-c)^2$
- $\sum_{i=1}^{n}|x_i-x_{med}| \leq \sum_{i=1}^{n}|x_i-c|$

3.3 条件期望

定义 3.5

若 $h(Y)$ 在给定 $X=x$ 下的条件分布 (定义2.14) 的数学期望存在，则定义其为**条件期望**如下:

$$E(h(Y)|X=x) = \begin{cases} \sum_y h(y)p_{Y|X}(y|x), & \text{离散情况} \\ \int_{-\infty}^{+\infty} h(y)f_{Y|X}(y|x)dy & \text{连续情况} \end{cases}$$

注 条件期望 $\mathbb{E}_{Y|X}(Y|x)$ 是关于给定变量 x 的函数，不随对应变量 Y 本身变动，但其单位与对应变量 Y 相同，可看作一条在 (X,Y) 平面的曲线。

命题 3.21

若随机变量 X, Y 独立，则:

$$\mathbb{E}_{Y|x}(Y|x) = \mathbb{E}_Y(Y)$$

证明 由定理2.2可知，若随机变量 X, Y 独立，则

$$p_{Y|X}(y|x) = p_Y(y)$$
$$f_{Y|X}(y|x) = f_Y(y)$$

所以 $\mathbb{E}_{Y|x}(Y|x) = \mathbb{E}_Y(Y)$

由直观感受亦可知：若 X, Y 独立，则 X 不通过任何与 Y 相关的信息，其条件期望亦当与原期望相同。

若令 $g(x) = \mathbb{E}(h(Y)|X=x)$，则 $g(X)$ 是随机变量 X 的变换，也是随机变量，记为 $\mathbb{E}(h(Y)|X)$。

定理 3.5 (重期望公式)

$$\mathbb{E}_Y[h(Y)] = \mathbb{E}_X[\mathbb{E}_{Y|X}(h(Y)|X)]$$

证明 设二维连续随机变量 (X,Y) 的联合密度函数为 $f(x,y)$。利用 $f(x,y)=f_{y|x}(y|x)f_X(x)$，可得：

$$
\begin{aligned}
\mathbb{E}_Y[h(Y)] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(y)f_{y|x}(y|x)f_X(x)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty} h(y)f_{y|x}(y|x)\,\mathrm{d}y\right\} f_X(x)\,\mathrm{d}x. \\
&= \int_{-\infty}^{+\infty} \mathbb{E}_{Y|X}(h(Y)|x)f_X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{+\infty} g(x)f_X(x)\,\mathrm{d}x \\
&= \mathbb{E}_X[g(X)] = \mathbb{E}_X[\mathbb{E}_{Y|X}(h(Y)|X)]
\end{aligned}
$$

离散场合可类似证明.

命题 3.22 (随机个随机变量和的数学期望)

设 $X_1,X_2,\ldots,$ 为一列独立同分布的随机变量，随机变量 N 只取正整数值，且 N 与 $\{X_n\}$ 独立，则：

$$E\left(\sum_{i=1}^{N} X_i\right) = \mu_X E(N)$$

证明

$$
\begin{aligned}
E\left(\sum_{i=1}^{N} X_i\right) &= E\left[E\left(\sum_{i=1}^{N} X_i|N\right)\right] \\
&= \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{N} X_i|N=n\right)P(N=n) \\
&= \sum_{n=1}^{+\infty} E\left(\sum_{i=1}^{n} X_i\right)P(N=n) \\
&= \sum_{n=1}^{+\infty} n\mu_X P(N=n) \\
&= \mu_X \sum_{n=1}^{+\infty} nP(N=n) \\
&= E(X_i)E(N)
\end{aligned}
$$

例题 3.9 设一天内到达某商场的顾客数 N 是仅取非负整数值的随机变量，且 $E(N)=35000$。又设进入此商场的第 i 个顾客的购物金额为 X_i，可以认为 X_i 是独立同分布的随机变量，且 $E(X_i)=82$(元)。假设 N 与 X_i 相互独立是合理的，则此商场一天的平均营业额为：

$$E\left(\sum_{i=0}^{N} X_i\right) = E(X_i)E(N) = 82\times 35000 = 287(\text{万元})$$

命题 3.23 (联合期望公式)

$$E_{X,Y}[h(X,Y)] = E_X E_{Y|X}[h(X,Y)|X] = E_Y E_{X|Y}[h(X,Y)|Y]$$

定理 3.6 (重方差公式)

随机变量 Y 的方差可作如下分解：

$$\mathrm{Var}_Y(Y) = \mathrm{Var}_X[E_{Y|X}(Y|X)] + E_X[\mathrm{Var}_{Y|X}(Y|X)]$$

证明

$$
\begin{aligned}
E_X[\mathrm{Var}_{Y|X}(Y|X)] &= E_X[E_{Y|X}(Y^2|X) - E_{Y|X}^2(Y|X)] \\
&= E_Y(Y^2) - E_X[E_{Y|X}^2(Y|X)] \\
\mathrm{Var}_X[E_{Y|X}(Y|X)] &= E_X[E_{Y|X}^2(Y|X)] - E_X^2[E_{Y|X}(Y|X)] \\
&= E_X[E_{Y|X}^2(Y|X)] - E_Y^2[Y] \\
E_X[\mathrm{Var}_{Y|X}(Y|X)] + \mathrm{Var}_X[E_{Y|X}(Y|X)] &= E_Y(Y^2) - E_Y^2(Y) \\
&= \mathrm{Var}_Y(Y)
\end{aligned}
$$

推论 3.5

$$\mathrm{Var}_Y(Y) \geq E_X[\mathrm{Var}_{Y|X}(Y|X)]$$

当且仅当 $E_{Y|X}(Y|X) = E_Y(Y)$ 时取等号

推论 3.6

$$\mathrm{Var}_Y(Y) \geq \mathrm{Var}_X[E_{Y|X}(Y|X)]$$

当且仅当 $\mathrm{Var}_{Y|X}(Y|X) = 0$，即 $\mathbb{E}_{Y|X}(Y|X) = Y$ 时取等号

3.4 矩母函数与特征函数

3.4.1 矩

定义 3.6

对于随机变量 X，定义其 k 阶**矩** (moment) 为 $E(X^k)$，记为 μ_k；定义其 k 阶**中心矩** (central moment) 为 $E((X-\mu_X)^k)$，记为 υ_k。

由于 $|X|^{k-1} \leq |X|^k + 1$，故 k 阶矩存在时，$k-1$ 阶矩也存在，从而低于 k 的各阶矩都存在. 易知矩与中心矩间存在以下关系:

$$
\begin{aligned}
\upsilon_k &= \sum_{i=0}^{k} \binom{k}{i} \mu_i (-\mu_X)^{k-i} \\
\mu_k &= \sum_{i=0}^{k} \binom{k}{i} \upsilon_i (\mu_X)^{n-i} j
\end{aligned}
$$

笔记 对于 μ_k 做变换 $E(X-\mu_X+\mu_X)^k$

例题 3.10 设随机变量 $X \sim N(0, \sigma^2)$，则

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

在 k 为奇数时，上述被积函数是奇函数，故

$$\mu_k = 0, k = 1, 3, 5, \cdots$$

在 k 为偶数时，上述被积函数是偶函数，再利用变换 $z = u^2/2$，可得

$$\begin{aligned}\mu_k &= \sqrt{\frac{2}{\pi}}\sigma^k 2^{(k-1)/2}\int_0^{+\infty} z^{(k-1)/2}\mathrm{e}^{-z}\,\mathrm{d}z = \sqrt{\frac{2}{\pi}} sigma^k 2^{(k-1)/2}\Gamma\left(\frac{k+1}{2}\right)\\ &= \sigma^k(k-1)(k-3)\cdots 1. \quad k = 2,4,6,\cdots.\end{aligned}$$

故 $N(0,\sigma^2)$ 分布的前四阶原点矩为

$$\mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = 0, \quad \mu_4 = 3\sigma^4$$

又因为 $E(X) = 0$，所以有原点矩等于中心矩，即 $\mu_k = \nu_k, k = 1,2,\cdots$.

3.4.2 矩母函数

定义 3.7

对于随机变量 X, 若下式期望存在：

$$M_X(t) = \mathbb{E}(e^{tX})$$

则称其为**矩母函数** (moment generating function, mgf)。

注 此表达式等价于对概率质量函数或密度函数作 Laplace 变换, 当 t 取某些特定值时, 可能不存在.(若 $t = 0$ 则永远存在)

定理 3.7

若当 t 属于一个包含零点的开区间时, 矩母函数一直存在, 则其唯一对应一个概率分布.

定理 3.8

若当 t 属于一个包含零点的开区间时, 矩母函数一直存在, 则:

$$M_X^{(k)}(0) = E(X^k)$$

注 借此可方便地计算各阶矩, 故称为矩母函数. 反过来, 若已知各阶矩, 通过 Tayler 展开 $M_X(t) = \sum_{k=0}^{\infty}\frac{M_X^{(k)}(0)}{k!}t^k$ 可还原矩母函数, 进而得出概率分布.

命题 3.24

若 a, b 为常数, 则

$$M_{a+bX}(t) = e^{at}M_X(bt)$$

定理 3.9

若 X, Y 独立, 则

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

泛化情况: 若 $X_1,\cdots,X_n$ 相互独立, 则

$$M_{X_1+\cdots+X_n} = \prod_{i=1}^{n} M_{X_i}(t)$$

3.4.3 联合特征函数

定义 3.8

对于随机变量 $X_1,\cdots,X_n$, 若下式期望存在:

$$M_{X_1\cdots X_n}(t_1,\cdots,t_n)=\mathbb{E}(e^{t_1X_1+\cdots+t_nX_n})$$

则称其为**联合矩母函数** (joint moment generating function, joint mgf)。

注 此处为多元 函数

命题 3.25

$$M_{X_i}(t_i)=M_{X_1\cdots X_n}(0,\cdots,t_i,\cdots,0)$$

定理 3.10

当且仅当:

$$M_{X_1\cdots X_n}(t_1,\cdots,t_n)=\prod_{i=1}^{n}M_{X_i}(t_i)$$

时,$X_1,\cdots,X_n$ 相互独立

注 与累计函数、密度函数、质量函数 的情况类似，变量相互独立等价于联合函数可拆分为边缘函数的乘积

定理 3.11

$$\frac{\partial^{r_1+\cdots+r_n}}{\partial t_1^{r_1}\cdots\partial t_n^{r_n}}M_{X_1\cdots X_n}(t_1,\cdots,t_n)=E(X_1^{r_1}\cdots X_n^{r_n})$$

3.4.4 特征函数

由于有时矩母函数可能不存在，为避免此缺陷，构造出与之特性类似的特征函数。

定义 3.9

对于随机变量 X, 定义其**特征函数** (characat function, chf) 为:

$$\varphi_X(t)=\mathbb{E}(e^{itX})=\mathbb{E}(\cos(tX))+i\mathbb{E}(\sin(tX))$$

对于随机变量 $X_1,\cdots,X_n$, 定义其**联合特征函数** (joint characat function, joint chf) 为:

$$\varphi_{X_1\cdots X_n}(t_1,\cdots,t_n)=\mathbb{E}(e^{i(t_1X_1+\cdots+t_nX_n)})$$

注 此表达式等价于对概率质量函数或密度函数作 Fourier 变换

命题 3.26 (特征函数的性质)

1. $|\varphi(t)|\le\varphi(0)=1$，即随机变量的特征函数总是存在。
2. 且若矩母函数存在，则其与特征函数之间满足关系：$\varphi_X(t)=M_X(it)$
3. $\varphi(-t)=\overline{\varphi(t)}$
4. 若 $Y=aX+b$, 其中 a,b 是常数, 则

$$\varphi_Y(t)=\mathrm{e}^{ibt}\varphi_X(at). \tag{3.1}$$

5. 独立随机变量和的特征函数为特征函数的积, 即设 X 与 Y 相互独立, 则

$$\varphi_{X+Y}(t)=\varphi_X(t)\cdot\varphi_Y(t). \tag{3.2}$$

6. 若 $E(x^l)$ 存在, 则 X 的特征函数 $\varphi(t)$ 可 l 次求导, 且对 $1 \le k \le l$, 有

$$\varphi^{(k)}(0) = i^k E(X^k). \tag{3.3}$$

证明

1.

$$\begin{aligned}|\varphi(t)| &= \left|\int_{-\infty}^{+\infty} \mathrm{e}^{itx} p(x)\,\mathrm{d}x\right| \le \int_{-\infty}^{+\infty} \left|\mathrm{e}^{itx}\right| p(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = \varphi(0) = 1.\end{aligned}$$

2.

$$\varphi(-t) = \int_{-\infty}^{+\infty} \mathrm{e}^{-itx} p(x)\,\mathrm{d}x = \overline{\int_{-\infty}^{+\infty} \mathrm{e}^{itx} p(x)\,\mathrm{d}x} = \overline{\varphi(t)}$$

3.

$$\varphi_Y(t) = E(\mathrm{e}^{it(aX+b)}) = \mathrm{e}^{ibt} E(\mathrm{e}^{iatX}) = \mathrm{e}^{ibt}\varphi(at)$$

4. 因为 X 与 Y 相互独立, 所以 e^{itX} 与 e^{itY} 也是独立的, 从而有

$$E\left(\mathrm{e}^{it(X+Y)}\right) = E\left(\mathrm{e}^{itX}\mathrm{e}^{itY}\right) = \varphi_X(t)\cdot\varphi_Y(t)$$

5. 因为 $E\left(X^l\right)$ 存在, 即

$$\int_{-\infty}^{+\infty} |x|^l p(x)\,\mathrm{d}x < +\infty$$

于是含参变量 t 的广义积分 $\int_{-\infty}^{+\infty} \mathrm{e}^{itx} p(x)\,\mathrm{d}x$ 可以对 t 求导 l 次, 于是对 $0 \le k \le l$, 有

$$\varphi^{(k)}(t) = \int_{-\infty}^{+\infty} i^k x^k \mathrm{e}^{itx} p(x)\,\mathrm{d}x = i^k E\left(X^k \mathrm{e}^{itX}\right)$$

令 $t = 0$ 即得

$$\varphi^{(k)}(0) = i^k E\left(X^k\right)$$

例题 3.11 柯西分布的特征函数 对于柯西分布特征函数的求解，不妨先设 $\mu = 0$，则 $f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2+x^2}$，则其特征函数为：

$$\varphi(t) = \frac{a}{\pi}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{itx}}{x^2+\sigma^2}\mathrm{d}x = \frac{2a}{\pi}\int_{0}^{+\infty} \frac{\cos tx}{x^2+\sigma^2}\mathrm{d}x$$

因为当 $t > 0$ 时，

$$\int_{0}^{+\infty} \frac{\cos tx}{x^2+\sigma^2}\mathrm{d}x = \frac{\pi}{a}\mathrm{e}^{-at}$$

所以当 $t > 0$ 时，$\varphi(t) = \mathrm{e}^{-\sigma t}$。当 $t < 0$ 时，$\varphi(t) = \overline{\varphi(-t)} = \mathrm{e}^{\sigma t}$。即 $\varphi(t) = \mathrm{e}^{-\sigma|t|}$。

对于一般情况，则 $Cau(\mu,\sigma) = Cau(0,\sigma) - \mu$

$$\varphi(t) = \exp(i\mu t)\exp(-\sigma|t|) = \exp(i\mu t - \sigma|t|)$$

定理 3.12 (逆转公式)

设 $F(x)$ 和 $\varphi(t)$ 分别为随机变量 X 的分布函数和特征函数, 则对 $F(x)$ 的任意两个连续点 $x_1 < x_2$, 有

$$F(x_2) - F(x_1) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-itx_1} - \mathrm{e}^{-itx_2}}{it}\varphi(t)\,\mathrm{d}t. \tag{3.4}$$

定理 3.13 (唯一性定理)

随机变量的分布函数由其特征函数唯一决定.

证明 对 $F(x)$ 的每一个连续点 x, 当 y 沿着 $F(x)$ 的连续点趋于 $-\infty$ 时, 由逆转公式得

$$F(x) = \lim_{y\to-\infty}\lim_{T\to+\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-ity}-\mathrm{e}^{-itx}}{it}\varphi(t)\,\mathrm{d}t$$

而分布函数由其连续点上的值惟一决定 (非连续点的情况可由右连续的性质推出), 故结论成立.

定理 3.14

特征函数可通过以下逆变换得到分布:

离散情况:

$$p_X(x) = \lim_{T\to\infty}\int_{-T}^{T}e^{-itx}\varphi_X(t)\mathrm{d}t$$

连续情况:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi_X(t)\mathrm{d}t$$

注 此即傅里叶逆变换。

证明 记 X 的分布函数为 $F(x)$, 由逆转公式知

$$\begin{aligned}p(x) &= \lim_{\Delta x\to0}\frac{F(x+\Delta x)-F(x)}{\Delta x}\\ &= \lim_{\Delta x\to0}\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{-itx}-\mathrm{e}^{-it(x+\Delta x)}}{it\cdot\Delta x}\varphi(t)\,\mathrm{d}t.\end{aligned}$$

再次利用不等式 $|\mathrm{e}^{ia}-1|\le|a|$, 就有

$$\left|\frac{\mathrm{e}^{-itx}-\mathrm{e}^{-it(x+\Delta x)}}{it\cdot\Delta x}\right|\le 1$$

又因为 $\int_{-\infty}^{+\infty}|\varphi(t)|<+\infty$, 所以可以交换极限号和积分号, 即

$$\begin{aligned}p(x) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\lim_{\Delta x\to0}\frac{\mathrm{e}^{-itx}-\mathrm{e}^{-it(x+\Delta x)}}{it\cdot\Delta x}\varphi(t)\,\mathrm{d}t\\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-itx}\varphi(t)\,\mathrm{d}t.\end{aligned}$$

3.5 估计与预测

3.5.1 delta 法

3.5.2 预测

3.6 熵与信息

3.6.1 费尔希信息量

3.7 其他特征

3.7.1 变异系数

定义 3.10 (变异系数)

设随机变量 X 的二阶矩存在，则称比值

$$C_v(X) = \frac{\sqrt{\mathrm{Var}(X)}}{E(X)} = \frac{\sigma(X)}{E(X)}$$

为 X 的**变异系数**

因为变异系数是以其数学期望为单位去度量随机变量取值波动程度的特征数，标准差的量纲与数学期望的量纲是一致的，所以变异系数是一个无量纲的量。

3.7.2 分位数

定义 3.11 (分位数)

设连续随机变量 X 的分布函数为 $F(x)$，密度函数为 $p(x)$. 对任意 $p \in (0,1)$，称满足条件

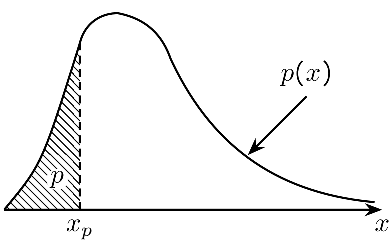
$$F(x_p) = \int_{-\infty}^{x_p} p(x)\,\mathrm{d}x = p$$

的 x_p 为此分布的 p **分位数**，又称**下侧 p 分位数**.
称满足条件

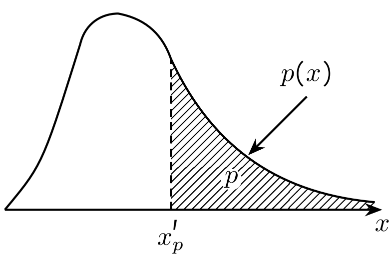
$$1 - F(x_p') = \int_{x_p'}^{+\infty} p(x)\,ddx = p$$

的 x_p' 为此分布的**上侧 p 分位数**

分位数 x_p 是把密度函数下的而积分为两块，左侧面积恰好为 p（见图 3.2(a)）。上侧分位数 x_p' 也是把密度函数下的面积分为两块，但右侧面积恰好为 p（见图 3.2(b)）.



(a)（下侧）分位数



(b) 上侧分位数

图 3.2: 分位数与上侧分位数的区别

分位数与上侧分位数是可以相互转换的，其转换公式如下.

$$x_p' = x_{1-p};\quad x_p = x_{1-p}'. \tag{3.5}$$

定义 3.12 (中位数)

设连续随机变量 X 的分布函数为 $F(x)$，密度函数为 $p(x)$. 称 $p = 0.5$ 时的 p 分位数 $x_{0.5}$ 为此分布的**中位数**，即 $x_{0.5}$ 满足

$$F(x_{0.5}) = \int_{-\infty}^{0.5} p(x)\,\mathrm{d}x = 0.5$$

♣

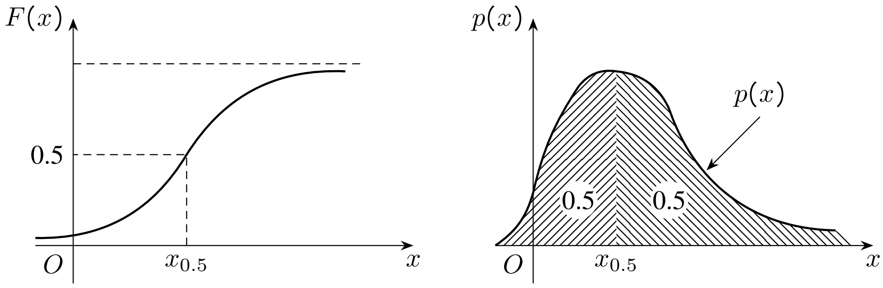


图 3.3: 连续随机变量的中位数

3.7.3 偏度系数与峰度系数

定义 3.13 (偏度系数)

设随机变量 X 的三阶矩存在，则称比值

$$\beta_1 = \frac{E[X - E(X)]^3}{[E(X - E(X))^2]^{3/2}} = \frac{\nu_3}{(\nu_2)^{3/2}}$$

为 X 的分布的**偏度系数**，简称**偏度**.

♣

偏度系数可以描述分布的形状特征，其取值的正负反映的是

- 当 $\beta_1 > 0$ 时，分布为正偏或右偏
- 当 $\beta_1 = 0$ 时，分布关于其均值 $E(X)$ 对称
- 当 $\beta_1 < 0$ 时，分布为负偏或左偏

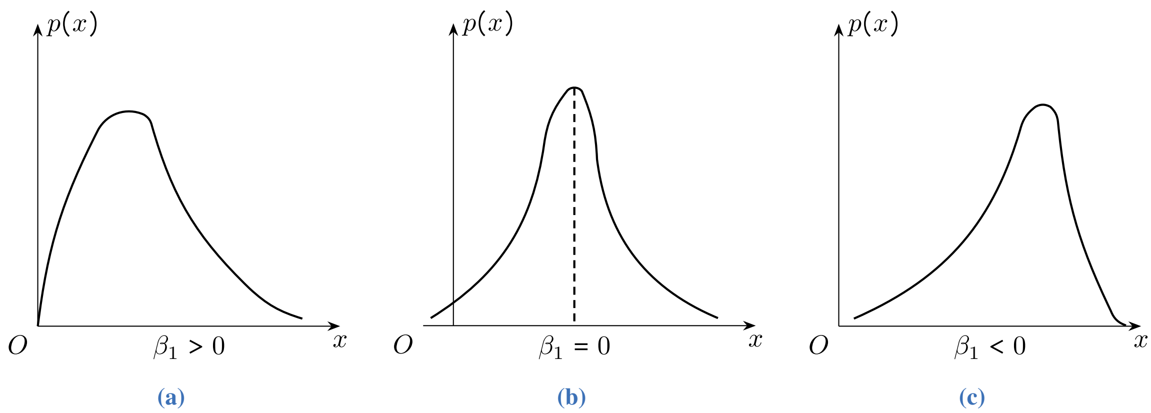


图 3.4: 三种不同偏度的分布

譬如，正态分布 $N(\mu, \sigma^2)$ 是关于其均值 $E(X) = \mu$ 是对称的，所以正态分布的偏度 $\beta_1 = 0$.

定义 3.14 (峰度系数)

设随机变量 X 的四阶矩存在，则称比值

$$\beta_2 = \frac{E[X - E(X)]^4}{[E(X - E(X))^2]^2} - 3 = \frac{\nu_4}{(\nu_2)^2} - 3$$

为 X 的分布的**峰度系数**，简称**峰度**.

峰度系数也是用于描述分布的形状特征，但峰度系数与偏度系数的差别是：偏度系数刻画的是分布的对称性，而峰度系数刻画的是分布的蜂峭性.

峰度系数把正态分布的峰峭性作为标准，因为正态分布 $N(\mu, \sigma^2)$ 的四阶中心矩为 $\nu_4 = 3\sigma^4$，所以其峰度系数为

$$\beta_2 = \frac{\nu_4}{(\nu_2)^2} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0$$

这说明：任一正态分布的峰度 $\beta_2 = 0$.

可见，这里谈论的“峰度”不是指密度函数的峰值高低，那么“峰度”的含义到底是什么呢? 或者换句话说，我们应该如何刻画密度函数的峰峭性呢? 我们知道从图形上看密度函数曲线下的面积等于 1，若随机变量取值较集中，则其密度函数的峰值必高无疑，所以密度函数峰值的高低含有随机变量取值的集中程度. 为了消除这个因素，我们不妨考察“标准化”后的分布的峰峭性，即用

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

的四阶原点矩 $E[(X^*)^4]$ 考察密度函数的峰值，再考虑到任一标准正态分布的四阶原点矩等于 3，所以就有了以上峰度系数的定义.

综上所述，一个分布的峰度系数 β_2 反映了以下情况：

- 当 $\beta_2 < 0$ 时，则标准化后的分布形状比标准正态分布更平坦，称为**低峰度**.
- 当 $\beta_2 = 0$ 时，则标准化后的分布形状与标准正态分布相当.
- 当 $\beta_2 > 0$ 时，则标准化后的分布形状比标准正态分布更尖峭，称为**高峰度**.

错题记录

1. (茆 2.2.7) 对一批产品进行检查，如查到第 a 件全为合格品，就认为这批产品合格；若在前 a 件中发现不合格品即停止检查，且认为这批产品不合格。设产品的数量很大，可认为每次查到不合格品的概率都是 p。问每批产品平均要查多少件?
2. (茆 2.2.17) 设随机变量 X 的概率密度函数为
$$p(x) = \begin{cases} \frac{1}{2}\cos\frac{x}{2}, & 0 \le x \le \pi; \\ 0, & \text{其他}. \end{cases}$$
对 X 独立重复观察 4 次，Y 表示观察值大于 $\pi/3$ 的次数，求 Y^2 的数学期望。
3. (茆 2.2.20) 设连续随机变量 X 的分布函数为 $F(x)$，且数学期望存在，证明
$$E(X) = \int_0^{+\infty} [1 - F(x)]\,\mathrm{d}x - \int_{-\infty}^0 F(x)\,\mathrm{d}x$$
4. (茆 2.2.21) 设 X 是非负连续随机变量，若 $E(X^n)$ 存在，证明：
   (a). $E(X) = \int_0^\infty P(X > x)\mathrm{d}x$
   (b). $E(X^n) = \int_0^\infty nx^{n-1}P(X > x)\mathrm{d}x$
5. (茆 2.2.22) 甲、乙两人进行象棋比赛，每局甲胜的概率为 p，乙胜的概率为 $q = 1 - p$。比赛进行到有一人连胜两局为止，求平均比赛局数。

6. （茆 2.2.23）设随机变量 X 的分布函数为

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

试求 $E(X)$。
7. （茆 3.4.2）求掷 n 颗骰子出现点数之和的数学期望与方差.
8. （茆 3.4.4）设在区间 $(0,1)$ 上随机地取 n 个点, 求相距最远的两点间的距离的数学期望.
9. （茆 3.4.10）设随机变量 X 与 Y 独立同分布，且 $E(X)=\mu,\mathrm{Var}(X)=\sigma^2$，试求 $E(X-Y)$。
10. （茆 3.4.13）系统由 n 个部件组成. 记 X_i 为第 i 个部件能持续工作的时间, 如果 $X_1,\ldots,X_n$ 独立同分布, 且 $X_i\sim Exp(\lambda)$，如果至少有一个部件在工作, 系统就工作求系统持续工作的平均时间。
11. （茆 3.4.17）一商店经销某种商品, 每周进货量 X 与顾客对该种商品的需求量 Y 是相互独立的随机变量, 且都跟从区间 $(10,20)$ 上的均匀分布. 商店每售出一单位商品可得利润 1000 元; 若需求量超过了进货量, 则可从其他商店调剂供应, 这时每单位商品获利润为 500 元. 试求此商店经销该种商品每局的平均利润.
12. （茆 3.4.21）掷一颗骰子两次，求其点数之和与点数之差的协方差。
13. （茆 3.4.26）设随机变量 X 和 Y 的数学期望分别为 -2 和 2，方差分别为 1 和 4，而它们的相关系数为 -0.5。试根据车比晓夫不等式，估计 $P(|X+Y|\geq 6)$ 的上限。
14. （茆 3.4.29）已知随机变量 X 与 Y 的相关系数为 ρ，求 $X_1=aX+b$ 与 $Y_1=cY+d$ 的相关系数，其中 a,b,c,d 均为常数。
15. （茆 3.4.32）设二维随机变量 $(X,Y)\sim N(0,0,1,1,\rho)$
    (a). 求 $E[\max(X,Y)]$
    (b). 求 $X-Y$ 与 XY 的协方差及相关系数
16. （茆 3.4.34）设二维随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}\frac{6}{7}\left(x^2+\frac{xy}{2}\right), & 0<x<1,0<y<2;\\ 0, & \text{其他.}\end{cases}$$

求 X 与 Y 的协方差及相关系数.
17. （茆 3.4.36）设二维随机变量 (X,Y) 的联合密度函数如下，试求 (X,Y) 的协方差矩阵。

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

18. （茆 3.4.38）设随机向量 (X_1,X_2,X_3) 满足条件

$$\begin{aligned}&aX_1+bX_2+cX_3=0,\\ &E(X_1)=E(X_2)=E(X_3)=d,\\ &\mathrm{Var}(X_1)=\mathrm{Var}(X_2)=\mathrm{Var}(X_3)=\sigma^2.\end{aligned}$$

其中 a,b,c,d,σ^2 均为常数，求相关系数 $\rho_{12},\rho_{23},\rho_{31}$.
19. （茆 3.4.39）设随机向量 X 与 Y 都只能取两个值，试证：X 与 Y 的独立性与不相关性是等价的。
20. （茆 3.4.42）设随机向量 (X_1,X_2,X_3) 的相关系数为 $\rho_{12},\rho_{23},\rho_{31}$，证明

$$\rho_{12}^2+\rho_{23}^2+p_{31}^2\leq 1+2\rho_{12}\rho_{23}\rho_{31}$$

21. （茆 3.4.46）设二维随机向量 (X,Y) 服从二维正态分布，且

$$E(X)=E(Y)=0,\quad E(XY)<0.$$

证明: 对任意正常数 a,b 有

$$P(X\geq a,Y\geq b)\leq P(X\geq a)P(Y\geq b)$$

22. (茆 3.4.47-52)
23. (茆 3.5.4)设随机变量 X 与 Y 独立同分布，试在以下情况下求 $P(X=k|X+Y=m)$
    - X 与 Y 都服从参数为 p 的几何分布
    - X 与 Y 都服从参数为 (n,p) 的二项分布.
24. (茆 3.5.9)设随机变量 X 服从 $(1,2)$ 上的均匀分布，在 $X=x$ 的条件下，随机变量 Y 的条件分布是参数为 x 的指数分布，证明：XY 服从参数为 1 的指数分布。
25. (茆 4.2.5)设 $X\sim N(\mu,\sigma^2)$，试用特征函数的方法求 X 的 3 阶及 4 阶中心矩。
26. (茆 4.2.12)设连续随机变量 X 的密度函数为 $p(x)$，试证：$p(x)$ 关于原点对称的充要条件是它的特征函数是实的偶函数。
27. (李 5.13)

第 4 章 常见分布

考试重点

- ❑ 各分布的特征
- ❑ 各分布联系与转换
- ❑ 多维正态分布

4.1 离散分布

4.1.1 均匀分布

定义 4.1 (离散均匀分布)

若随机变量只能在 $a_1, \cdots, a_n$ 中取值，并且对应的概率相同，则称其遵循**均匀分布** (Uniform distribution)，记为 $X \sim U(a_1, \cdots, a_n)$。♣

注 此即古典概型。

离散均匀分布的特征：

参数 $a_i \in \mathbb{R}$

概率质量函数 $p(x) = \begin{cases} \frac{1}{m} & x = a_1, a_2, \cdots, a_n \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = \frac{\sum_{i=1}^m e^{a_i t}}{m}$

均值 $\mu = \frac{\sum_{i=1}^m a_i}{m} = \overline{a}$

方差 $\sigma^2 = \frac{\sum_{i=1}^m (a_i - \overline{a})^2}{m}$

实例 丢一个均匀的骰子

4.1.2 两点分布

定义 4.2 (两点分布)

若随机变量只能取 0 或 1，并且对应的概率分别为 $1-p$ 与 p，则称其遵循**两点分布** (Bernouli distribution)（也称 0-1 分布、两点分布分布），记为 $X \sim B(1, p)$。♣

两点分布的特征：

参数 $p \in [0, 1]$

概率质量函数 $p(x) = \begin{cases} p^x (1-p)^{1-x} & x \in \{0, 1\} \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = pe^t + 1 - p$

均值 $\mu = p$

方差 $\sigma^2 = p(1-p)$

实例 丢一次硬币，p 代表某一面出现的概率

注 若 A 是一个事件，出现概率为 p_A，则指示随机变量 I_A（若 A 出现记为 1，否则为 0）遵循 Bernoulli 分布，即 $I_A \sim B(p_A)$

4.1.3 二项分布

定义 4.3 (二项分布)

若进行 n 次独立的 Bernoulli 实验（定义1.14）$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} B(1,p)$，则这些随机变量之和 $Y = X_1 + \cdots + X_n$ 遵循**二项分布** (binomial distribution)，记为 $Y \sim B(n,p)$。♣

二项分布的特征：

参数 $p \in [0,1], n \in \mathbb{N}_+$

概率质量函数 $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \mathbb{N}^n \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = (pe^t + 1 - p)^n$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（定理3.9）
- 定义

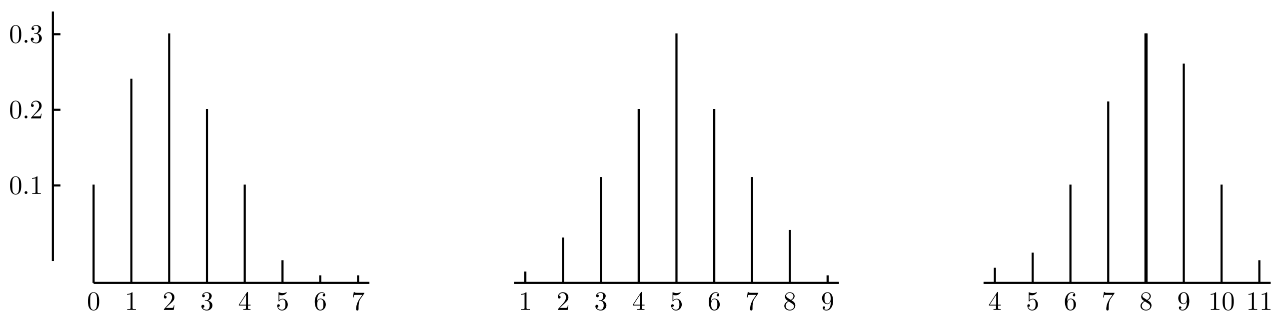
均值 $\mu = np$，求解方法有：

- 定义（例3.2）
- 矩母函数在 $t = 0$ 的一阶导
- 变量之和的均值等于各变量均值之和

方差 $\sigma^2 = np(1-p)$，求解方法有：

- 定义（例3.4）
- 矩母函数在 $t = 0$ 的二阶导，$\mathrm{Var}(X) = E(X^2) - (E(X))^2$
- 独立变量 之和的方差等于各变量均值之和

实例 丢 n 次硬币，p 代表某一面出现的概率，出现此面的次数



(a) $b(10, 0, 2)$ 的线条图 (右偏)　(b) $b(10, 0, 5)$ 的线条图 (对称)　(c) $b(10, 0, 8)$ 的线条图 (左偏)

图 4.1: 二项分布 $b(n,p)$ 的线条图

从上图可以看出：

- 位于均值 np 附近概率较大；
- 随着 p 的增加，分布的峰逐渐右移.

命题 4.1

二项分布之和仍是二项分布。若 $X_1+\cdots+X_k$ 独立且 $X_i \sim B(n_i,p)$，则 $Y = X_1+\cdots+X_k \sim B(n_1+\cdots+n_k,p)$

证明

$$M_{X_1+\cdots+X_k}(t) = \prod_{i=1}^{k}(pe^t+1-p)^{n_i} = (pe^t+1-p)^{\sum_{i=1}^{k} n_i}$$

4.1.4 泊松分布

定理 4.1 (Poisson 逼近)

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

证明 记 $np_n = \lambda_n$，记 $p_n = \lambda_n/n$，我们可得

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

对固定的 k 有

$$\lim_{n\to+\infty}\lambda_n = \lambda,\quad \lim_{n\to+\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k} = e^{-\lambda},\quad \lim_{n\to+\infty}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) = 1$$

从而

$$\lim_{n\to+\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda},\quad \forall k\in\mathbb{N}$$

定义 4.4 (Poisson 分布)

若随机变量 X 的概率分布列满足以下形式：

$$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda},\ k\in\mathbb{N}$$

则称其遵循 **Poisson 分布**，记为 $X\sim P(\lambda)$.

注 由定理4.1可看出，Poisson 分布可作为二项分布的近似。λ 的涵义

Poisson 分布的特征：

参数 $\lambda > 0$

概率质量函数 $p(k) = \begin{cases}\frac{\lambda^k}{k!}e^{-\lambda} & x\in\mathbb{N}^n \\ 0 & \text{其他}\end{cases}$

矩母函数 $M(t) = e^{\lambda(e^t-1)}$，求解方法有：

- 利用公式
- 凑一法

均值 $\mu = \lambda$

方差 $\sigma^2 = \lambda$

实例 公共汽车站来到的乘客数

命题 4.2

Poisson 分布之和仍是 Poisson 分布。若 $X_1 + \cdots + X_k \overset{\text{i.i.d.}}{\sim} P(\lambda_i)$，则 $Y = X_1 + \cdots + X_k \sim P(\lambda_1 + \cdots + \lambda_k)$

证明

$$M_Y(t) = \prod_{i}^{k} M_{X_i}(t) = \exp\{(e^t - 1)\sum_{i=1}^{k}\lambda_i\}$$

引理 4.1

若 $f(x)$ 是连续函数（或单调函数），且对一切 x, y（或一切 $x, y \geq 0$）成立：

$$f(x)f(y) = f(x+y)$$

则 $f(x)$ 为**指数函数**。

证明

$$\because f(1) = [f(\frac{1}{n})]^n$$
$$且 a = f(1) = [f(\frac{1}{2})]^2 \geq 0$$
$$\therefore f(\frac{1}{n}) = a^{\frac{1}{n}}$$
$$\therefore f(\frac{m}{n}) = [f(\frac{1}{n})]^m = a^{\frac{m}{n}}$$

即此命题对一切有理数成立。又实数具有连续性，故此命题对一切实数成立。

定义 4.5 (泊松过程)

若某一随机过程满足以下特征，则称其为**泊松过程**：

平稳性 随机事件 A 在 $[t_0, t_0 + t)$ 中的发生次数只与时间间隔长度 t 有关而与时间起点 t_0 无关。将在长度为 t 的时间区间中发生 K 次事件 A 的概率记为 $P_k(t)$。

独立增量性 在 $[t_0, t_0 + t)$ 中发生 K 次事件 A 的概率与时刻 t_0 以前发生的事件独立。

普通性 在充分小的时间间隔中，最多发生一次事件 A。即，若记 $\psi(t) = \sum_{k=2}^{\infty} P_k(t) = 1 - P_0(t) - P_1(t)$，则有 $\lim_{t\to 0}\psi(t)/t = 0$

对 $\Delta t > O$，考虑 $[O, t + \Delta t)$ 中发生 K 次事件 A 的概率 $P_k(t + \Delta t)$，由独立增量性及全概率公式可得：

$$P_k(t + \Delta t) = P_k(t)P_0(\Delta r) + P_{k-1}P_1(\Delta t) + \cdots + P_0(t)P_k(\Delta t), k \geq 0$$

（对 $n \geq 1$，假定 $P_{-n}(t) = 0$）

特别地

$$P_0(t + \Delta t) = P_0(t)P_0(\Delta t)$$

由引理4.1知

$$P_0(t) = a^t, a \geq 0$$

若 $a = O$，则 $P_0(t) \equiv 0$，说明在不管怎么短的时间间隔内事件 A 都发生，因此在有限时间间隔中将发生无穷多个次事件 A，这种情形不在我们的考虑之列。此外，因 $P_0(t)$ 是概率，故应有 $a \geq 1$，而当 $a = 1$ 时，$P_0(t) \equiv 1$，

表明事件 A 永不发生，也不是我们感兴趣的情形，所以应有 $0 < a < 1$，从而存在 $\lambda > 0$，使

$$P_0(t) = e^{-\lambda t}$$

因此当 $\Delta t \to 0$ 时，有

$$P_0(\Delta t) = \mathrm{e}^{-\lambda\Delta t} = 1 - \lambda\Delta t + o(\Delta t)$$

$$P_1(\Delta t) = 1 - P_0(\Delta t) - \psi(\Delta t) = \lambda\Delta t + o(\Delta t)$$

$$\sum_{l=2}^{\infty} P_{k-l}(t)P_l(\Delta t) \leqslant \sum_{l=2}^{\infty} P_l(\Delta t) = \psi(\Delta t) = o(\Delta t)$$

所以：

$$P_k(t+\Delta t) = P_k(t)(1-\lambda\Delta t) + P_{k-1}\lambda\Delta t + o(\Delta t), k \geq 1$$

因此：

$$P_k'(t) = \frac{P_k(t+\Delta t) - P_k(t)}{\Delta t} = \lambda[P_{k-1} - P_k(t)]$$

由于 $P_0(t) = \mathrm{e}^{-\lambda t}$，故有 $P_1'(t) = \lambda[\mathrm{e}^{-\lambda t} - P_1(t)]$，可解得 $P_0(t) = \lambda t\mathrm{e}^{-\lambda t}$，依次可递推可解得：

$$P_k(t) = \frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}, k \in \mathbb{N}$$

正是参数为 λt 的泊松分布。

4.1.5 几何分布

定义 4.6

若进行无限次独立的 Bernouli 实验，则第一次出现“1”的实验次数遵循**几何分布** (geometric distribution)，记为 $X \sim G(p)$。

几何分布的特征：

参数 $p \in [0,1]$

概率质量函数 $p(x) = \begin{cases} (1-p)^{x-1}p & x \in \mathbb{N}_+ \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = \frac{pe^t}{1-(1-p)e^t}, t < -\ln(1-p)$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（参见定理3.9）
- 定义

均值 $\mu = \frac{1}{p}$，求解方法有：

- $E(X) = \sum_{k=1}^{\infty} P(X \geq k)$（定理3.3）
- 矩母函数在 $t = 0$ 的一阶导
- 定义（例3.3）

方差 $\sigma^2 = \frac{1-p}{p^2}$，求解方法有：

- 定义（例3.5）
- 矩母函数在 $t = 0$ 的二阶导，$\mathrm{Var}(X) = E(X^2) - (E(X))^2$

实例 买彩票时，中奖所需购买张数。

命题 4.3

离散分布中有且只有几何分布具备无记忆性：

$$P\{X>m+n|X>m\}=P\{X>n\},\quad \forall s>0,t>0$$

证明

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

$$\therefore P\{X>m+n|X>m\}=P\{X>n\}=\frac{(1-p)^{m+n}}{(1-p)^m}$$

$$=(1-p)^n=P\{X>n\}$$

即几何分布具备无记忆性。

设 X 是取正整数值的随机变量，满足：在已知 $X>k$ 的条件下，$X=k+1$ 的概率与 k 无关。令 $p=P\{X=k+1|X>k\}$，并记 $q_k=P\{X>k\}$，以及 $p_k=P\{X=k\}$，那么 $p_{k+1}=q_k-q_{k+1}$，且

$$p=\frac{p_{k+1}}{q_k}$$

即

$$\frac{q_{k+1}}{q_k}=1-p$$

注意到 $q_0=1$，那么 $q_k=(l-p)^k$，因此

$$p_k=p(l-p)^{k-l},\quad k\in N_+$$

正是几何分布

4.1.6 负二项分布

定义 4.7

若进行无限次独立的 Bernouli 实验，则第 r 次出现“1”的实验次数遵循**负二项分布** (negative binomial distribution)（也称为帕斯卡分布），记为 $X\sim NB(r,p)$。

负二项分布的特征：

参数 $p\in[0,1]$

概率质量函数 $p(x)=\begin{cases}\binom{x-1}{r-1}p^r(1-p)^{x-r} & x\in\mathbb{N}_r\\ 0 & \text{其他}\end{cases}$

矩母函数 $M(t)=\frac{p^r e^{rt}}{(1-(1-p)e^t)^r},t<-\ln(1-p)$，求解方法有：

- 独立变量 之和的矩母函数等于各变量矩母函数之积（参见定理3.9）
- 定义凑一法
- 利用公式 $\sum_{x=0}^{\infty}\binom{n+x-1}{x}t^x=\frac{1}{(1-t)^n},-1<t<1$

均值 $\mu=\frac{r}{p}$，求解方法有：

- 定义凑一法
- 矩母函数在 $t=0$ 的一阶导
- 独立几何分布之和 (定理4.4)

方差 $\sigma^2=\frac{r(1-p)}{p^2}$，求解方法有：

- 定义（凑一法）
- 矩母函数在 $t=0$ 的二阶导，$\mathrm{Var}(X)=E(X^2)-(E(X))^2$
- 独立几何分布之和 (定理4.4)

实例 买彩票时，中奖 r 次所需购买张数。

命题 4.4

几何分布之和是负二项分布。若 $X_1+\cdots+X_r$ 独立且 $X_i\sim G(p)$，则 $Y=X_1+\cdots+X_r\sim NB(r,p)$

证明

命题 4.5

负二项分布之和仍是负二项分布。若 $X_1+\cdots+X_k$ 独立且 $X_i\sim NB(r_i,p)$，则 $Y=X_1+\cdots+X_k\sim B(r_1+\cdots+r_k,p)$

证明

4.1.7 多项分布

定义 4.8 (多项分布)

若进行 n 次独立的实验，每次实验有 r 种结果，每种结果对于概率分别为 $p_1,\cdots,p_r$。令 X_i 代表得出结果 i 的次数，则随机向量 $\mathbf{X}=(X_1,\cdots,X_r)$ 遵循**多项分布** (multinomial distribution)，记为 $\mathbf{X}\sim Multinomial(n,p_1,\cdots,p_r)$。

多项分布的特征：

参数 $p_i\in[0,1],\sum_{i=1}^r p_i=1$

概率质量函数 $p(x)=\begin{cases}\binom{n}{x_1\cdots x_r}p_1^{x_1}\cdots p_r^{x_r} & x\in\mathbb{N}^n\,\&\,\sum_{i=1}^r x_i=n\\ 0 & \text{其他}\end{cases}$

矩母函数 $M(t_1,\cdots,t_r)=(p_1e^{t_1}+\cdots+p_re^{t_r})^n$，求解方法有：

- 利用公式
- 凑一法

边缘分布 $X_i\sim B(n,p_i)$

均值 $E(X_i)=np_i$

方差 $\mathrm{Var}(X_i)=np_i(1-p_i)$

协方差 $\mathrm{Cov}(X_i,X_j)=-p_ip_j$，求解方法有：

- 矩母函数求解 $E(X_iX_j)$
- 凑一法求解 $E(X_iX_j)$

实例

注 多项分布是二项分布的泛化。

4.1.8 超几何分布

定义 4.9

设有 N 个产品，其中有 M 个不合格品. 若从中不放回地随机抽取 n 个，则其中含有的不合格品的个数 X 服从**超几何分布** (Hypergeometric distribution)，记为 $X \sim h(n, N, M)$

超几何分布的特征：

参数	$M, n \le N \quad n, N, M \in N_+$
概率质量函数	$p(k) = \begin{cases} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, & k = 0, 1, \cdots, r \\ 0 & \text{其他} \end{cases}$
矩母函数	存在，但没有简单的表达
均值	$\mu = n\frac{M}{N}$（利用凑一法，例3.2）
方差	$\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}$
实例	从一个有限总体中进行不放回抽样

注 若 $\frac{n}{N} \to 0$ 时，$\frac{M}{N} \to p$，则超几何分布可近似为二项分布。

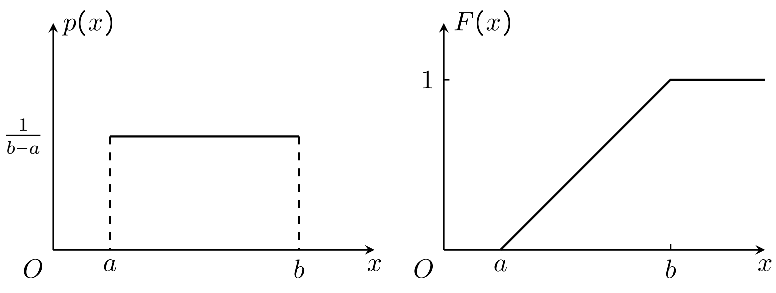
$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \approx \binom{n}{k} p^k (1-p)^{n-k}, \quad p = \frac{M}{N}$$

4.2 连续分布

4.2.1 均匀分布

定义 4.10

若随机变量 X 在 $[a, b]$ 中任一区域的概率与其测度成正比，则称 X 服从区间 (a, b) 上的**均匀分布** (Uniform distribution)，记作 $X \sim U(a, b)$。



(a) 密度函数 $p(x)$　(b) 分布函数 $F(x)$

图 4.2: (a, b) 上的均匀分布

均匀分布的特征：

参数	$a < b, \quad a, b \in \mathbb{R}$
概率质量函数	$p(x) = \frac{1}{b-a}\mathbb{1}_{\{a<x<b\}}(x)$

分布函数	$F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \le x < b; \\ 1, & x \ge b. \end{cases}$
矩母函数	$M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}, \quad t \in \mathbb{R}$
均值	$\mu = \frac{a+b}{2}$
方差	$\sigma^2 = \frac{(b-a)^2}{12}$

注 常利用 $U(0,1)$ 生成特定分布的伪随机数（参见定理2.5）

4.2.2 指数分布

定义 4.11

若随机变量 X 的密度函数（见图4.3）为

$$p(x) = \lambda e^{-\lambda x}\mathbb{1}(x \ge 0)$$

则称 X 服从**指数分布**（Exponential distribution），记作 $X \sim Exp(\lambda)$，其中参数 $\lambda > 0$。指数分布的分布函数为

$$F(x) = (1 - e^{-\lambda x})\mathbb{1}(x \ge 0)$$

♣

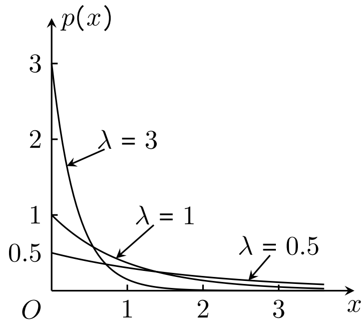


图 4.3: 参数为 λ 的指数分布密度函数

指数分布的特征：

矩母函数	$M(t) = \frac{\lambda}{\lambda-t}, \quad t < \lambda$
均值	$\mu = \frac{1}{\lambda}$
方差	$\sigma^2 = \frac{1}{\lambda^2}$
实例	物品使用寿命，等待时间

注 由其均值可看出，对于等待时间模型，$\frac{1}{\lambda}$ 代表平均等待时间，即（时间/次）；λ 代表平均发生频率，即（次/时间）。

定理 4.2 (指数分布的无记忆性)

连续分布中，有且只有指数分布具备无记忆性：

$$P(X > s + t | X > s) = P(X > t), \quad \forall s > 0, t > 0$$

♡

证明 由

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

可知，指数分布具备无记忆性。

设 X 为非负随机变量，其分布函数为 $F(x)$，记 $G(x)=P\{X\geq x\}$。若其具备无记忆性，则 $G(s+t)=G(s)G(t)$。由于 $G(x)$ 是单调函数，有由引理4.1得，$G(x)$ 为指数函数：

$$G(x)=a^x,\ x\geq 0$$

由于 $G(x)$ 是概率，故 $0<a<1$，可以写成 $a=\mathrm{e}^{-\lambda}$，$\lambda>0$。因此

$$F(x)=1-G(x)=1-a=\mathrm{e}^{-\lambda x},\ x\geq 0$$

即 X 服从指数分布。

命题 4.6

如果某事件在任何长为 t 的时间 $[0,t]$ 内发生的次数 $N(t)$ 服从参数为 λt 的泊松分布，则相继两次事件之间的时间间隔 T 服从参数为 λ 的指数分布。

证明 设 $N(t)\sim P(\lambda t)$，即

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

注意到两次事件发生的时间间隔 T 是非负随机变量；且事件 $\{T\geq t\}$ 说明在 $[0,t]$ 内事件没有故障，即 $\{T\geq t\}=\{N(t)=0\}$，由此得

- 当 $t<0$ 时，有 $F_T(t)=P(T\leq t)=0$；
- 当 $t\geq 0$ 时，有 $F_T(t)=P(T\leq t)=1-P(T>t)=1-P\big(N(t)=0\big)=1-e^{-\lambda t}$

所以 $T\sim Exp(\lambda)$，相继两次事件之间的时间间隔 T 服从参数为 λ 的指数分布。

命题 4.7

指数分布的最小统计量仍是指数分布：

$$X_{(1)}\sim P(n\lambda),\quad X_1,\cdots,X_n\overset{\text{i.i.d.}}{\sim}P(\lambda)$$

证明

$$f_{X_{(1)}}(y)=n[1-F_X(y)]^{n-1}f_X(y)=n\mathrm{e}^{-(n-1)\lambda}\lambda\mathrm{e}^{-\lambda}=n\lambda\mathrm{e}^{-n\lambda}$$

4.2.3 伽马分布

定义 4.12 (伽马函数)

称

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x,\quad \alpha>0$$

为**伽马函数** (gamma function)。

Gamma 函数的特征：

- $\Gamma(1)=1$
- $\Gamma(\frac{1}{2})=\sqrt{\pi}$
- $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$（分部积分可证）
- $\Gamma(n)=(n-1)!,\quad n\in\mathbb{N}$
- $\Gamma(\frac{\alpha}{2})=\frac{\sqrt{\pi}(\alpha-1)!}{2^{\alpha-1}(\frac{\alpha-1}{2})!},\quad \alpha=2k+1,k\in\mathbb{N}$

定义 4.13 (伽玛分布)

若随机变量 X 的密度函数为

$$p(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}\mathbb{1}_{\{x\geq 0\}}(x)$$

则称 X 服从**伽玛分布**，记作 $X\sim\Gamma(\alpha,\lambda)$，其中称 $\alpha>0$ 为形状参数，$\lambda>0$ 为尺度参数。

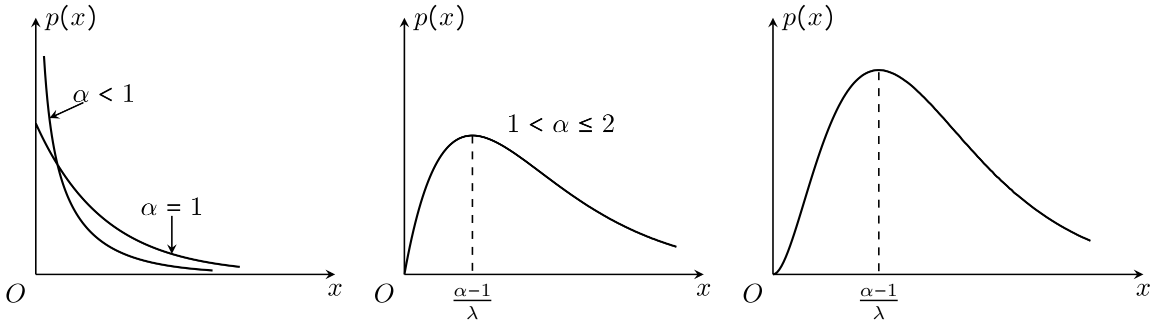


图 4.4: λ 固定、不同 α 的伽玛密度曲线

从图4.4中可以看出：

- $0<\alpha<1$ 时，$p(x)$ 是严格下降函数，且在 $x=0$ 处有奇异点；
- $\alpha=1$ 时，$p(x)$ 是严格下降函数，且在 $x=0$ 处 $p(0)=\lambda$；
- $1<\alpha\leq 2$，$p(x)$ 是单峰函数，向上凸、后下凸；
- $2<\alpha$ 时，$p(x)$ 是单峰函数，先下凸、中间上凸、后下凸. 且 α 越大，$p(x)$ 越近似于正态分布.

Gamma 分布的特征：

矩母函数	$M(t)=(\frac{\lambda}{\lambda-t})^\alpha,\quad t<\lambda$
均值	$\mu=\frac{\alpha}{\lambda}$
方差	$\sigma^2=\frac{\alpha}{\lambda^2}$

命题 4.8

Γ 分布是指数分布的泛化，即 $\Gamma(1,\lambda)=E(\lambda)$。进一步有：若随机变量 $X_1,\cdots,X_k,\overset{\text{i.i.d.}}{\sim}E(\lambda)$，则 $Y=X_1+\cdots+X_k\sim\Gamma(k,\lambda)$。更进一步有：若随机变量 $X_1',\cdots,X_k',\overset{\text{i.i.d.}}{\sim}\Gamma(\alpha_i,\lambda)$，则 $Y'=X_1'+\cdots+X_k'\sim\Gamma(\alpha_1+\cdots+\alpha_k,\lambda)$。

证明

$$M_Y(t)=\prod_{i=1}^k M_{X_i}(t_i)=\prod_{i=1}^k\frac{\lambda}{\lambda-t}=(\frac{\lambda}{\lambda-t})^k$$

$$M_{Y'}(t)=\prod_{i=1}^k M_{X_i'}(t_i)=\prod_{i=1}^k(\frac{\lambda}{\lambda-t})^{\alpha_i}=(\frac{\lambda}{\lambda-t})^{\alpha_1+\cdots+\alpha_k}$$

命题 4.9

若令随机变量 $X\sim\Gamma(\alpha,\lambda)$，则 $cX\sim\Gamma(\alpha,\frac{\lambda}{c}),\quad c>0$

证明

$$M_{cX}(t)=M_X(ct)=(\frac{\lambda}{\lambda-ct})^\alpha=(\frac{\frac{\lambda}{c}}{\frac{\lambda}{c}-t})^\alpha$$

命题 4.10

若令随机变量 $X \sim \Gamma(\alpha, \lambda)$，则 $\mu_k = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)}$，$\quad 0 < k$，且则 $\mu_{-k} = \frac{\lambda^k \Gamma(\alpha-k)}{\Gamma(\alpha)}$，$\quad 0 < k < \alpha$

证明

$$\mu_k = M_X^{(k)}(0) = \frac{1}{\lambda^k}\alpha(\alpha+1)\cdots(\alpha+k-1) = \frac{\Gamma(\alpha+k)}{\lambda^k\Gamma(\alpha)}$$
$$\mu_{-k} = M_X^{(-k)}(0) = \lambda^k \frac{1}{\alpha-1}\cdots\frac{1}{\alpha-k} = \frac{\lambda^k\Gamma(\alpha-k)}{\Gamma(\alpha)}$$

命题 4.11

4.2.4 贝塔分布

定义 4.14

称

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

为**贝塔函数**，其中参数 $a > 0, b > 0$.

命题 4.12

贝塔函数具有如下性质：

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

证明 在贝塔函数的积分中令 $y = 1-x$，即得

$$\beta(a,b) = \int_1^0 (1-y)^{a-1}y^{b-1}(-\,\mathrm{d}y) = \int_0^1 (1-y)^{a-1}y^{b-1}\,\mathrm{d}y = \beta(b,a)$$

由伽玛函数的定义知

$$\Gamma(a)\Gamma(b) = \int_0^{+\infty}\int_0^{+\infty} x^{a-1}y^{b-1}\mathrm{e}^{-(x+y)}\,\mathrm{d}x\,\mathrm{d}y$$

作变量变换 $x = uv, y = u(1-v)$，其雅可比行列式 $J = -u$，故

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

定义 4.15

若随机变量 X 的密度函数为

$$p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{\{0<x<1\}}(x)$$

则称 X 服从**贝塔分布**，记作 $X \sim Be(a,b)$，其中 $a > 0, b > 0$ 都是形状参数。

从图 4.5 可以看出：

- $a < 1, b < 1$ 时，$p(x)$ 是下凸的单峰函数.

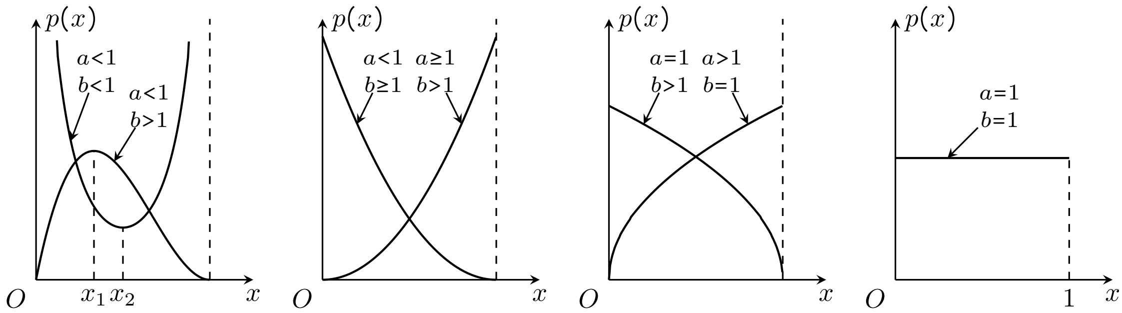


图 4.5: 贝塔密度函数曲线

- $a > 1, b > 1$ 时，$p(x)$ 是上凸的单峰函数.
- $a < 10, b \geq 1$ 时，$p(x)$ 是下凸的单调减函数.
- $a \geq 1, b < 1$ 时，$p(x)$ 是下凸的单调增函数.
- $a = 1, b = 1$ 时，$p(x)$ 是常函数，且 $\underline{Be(1,1) = U(0,1)}$.

Beta 分布的特征：

矩母函数	$M(t) = 1 + \sum_{k=1}^{\infty}(\prod_{r=0}^{k-1} \frac{a+r}{a+b+r})\frac{t^k}{k!}, \quad t < \lambda$
均值	$\mu = \frac{a}{a+b}$
方差	$\sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}$
实例	不合格品率、机器的维修率、市场的占有率、射击的命中率等各种比率

命题 4.13

令独立随机变量 $X_1 \sim \Gamma(\alpha_1, \lambda), X_2 \sim \Gamma(\alpha_2, \lambda)$，则 $\frac{X_1}{X_1+X_2} \sim \beta(\alpha_1, \alpha_2)$

证明

4.2.5 柯西分布

定义 4.16

若随机变量 X 的密度函数为

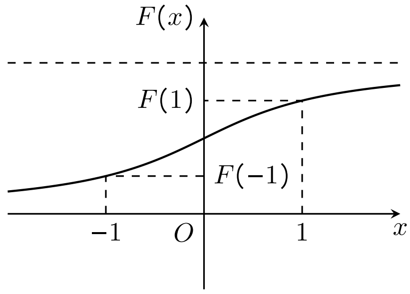
$$f(x) = \frac{\sigma}{\pi}\frac{1}{\sigma^2 + (x-\mu)^2}, x \in \mathbb{R}$$

则称 X 服从**柯西分布** (Cauchy distribution)，记作 $X \sim Cau(\mu, \sigma)$。其中参数 $\mu \in \mathbb{R}, \sigma > 0$。

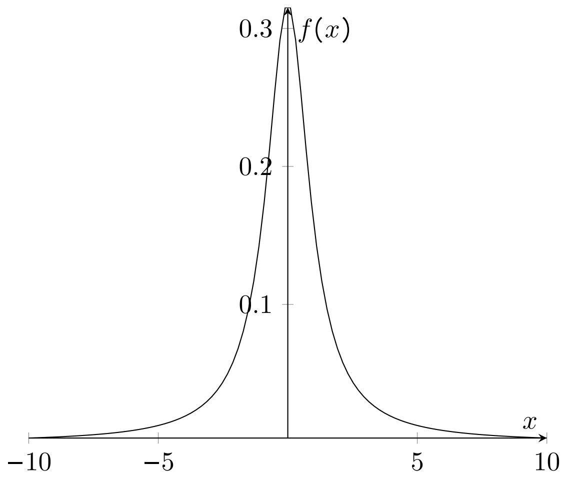
柯西分布的特征：

累积函数	$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(\mu + \sigma x)$
矩母函数	除 $t = 0$ 外不存在
特征函数	$\phi_X(t) = \exp(i\mu t - \sigma\|t\|)$（见例3.11）
均值	不存在
方差	不存在

注 此类因极端值概率密度较高而导致均值、方差不存在的分布称为重尾分布。



(a) 柯西分布累积函数



(b) 柯西分布密度函数

图 4.6: 柯西分布

命题 4.14

$C(0,1) = t_1$

命题 4.15

若随机变量 X, Y i.i.d. $\sim N(0,1)$，则 $\frac{X}{Y} \sim C(0,1)$

证明

命题 4.16

若随机变量 $X \sim C(\mu,\sigma)$，则其线性变换 $a,b \in \mathbb{R}, aX+b \sim C(a\mu+b, |a|\sigma)$。

证明

$$\phi_{aX+b}(t) = e^{ibt}\phi_X(at) = \exp(ibt + ia\mu t - \sigma|at|)$$

命题 4.17

若独立随机变量 $X_1,\cdots,X_k$ 满足 $X_i \sim C(\mu_i,\sigma_i)$，则其和 $Y = X_1+\cdots+X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i)$

证明

$$\phi_{X_1+\cdots+X_n} = \prod_{i=1}^n \phi_{X_i}(t) = \exp(it\sum_{i=1}^k \mu_i - |t|\sum_{i=1}^k \sigma_i)$$

推论 4.1

若随机变量 $X_1,\cdots,X_k$ i.i.d. $\sim N(\mu,\sigma^2)$，则其平均 $\overline{X} = (X_1+\cdots+X_k)/k \sim C(\mu,\sigma)$

注 此处的 σ 不代表方差，所以不遵守 $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{k}$ 的关系。

4.3 正态分布及其导出分布

4.3.1 正态分布

定义 4.17

若随机变量 X 的密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}$$

则称 X 服从**正态分布** (Normal distribution)，记作 $X \sim N(\mu, \sigma^2)$。其中参数 $\mu \in \mathbb{R}, \sigma > 0$。特别地，将 $N(0,1)$ 称为**标准正态分布** (standard Normal distribution)，其密度函数记为 $\varphi(x)$，分布函数记为 $\Phi(x)$。

正态分布的密度函数 $f(x)$ 的图形是一条钟形曲线，左右关于 μ 对称，衰减速度由 σ 决定，$\mu \pm \sigma$ 是该曲线的拐点。

- 如果固定 σ，改变 μ 的值，则图形沿 x 轴平移，而不改变其形状，即正态密度函数的位置由参数 μ 所确定，因此亦称 μ 为**位置参数**。
- 如果固定 μ，改变 σ 的值，则 σ 愈小，曲线呈高而瘦；σ 愈大，曲线呈矮而胖，即正态密度函数的尺度由参数 σ 所确定，因此称 σ 为**尺度参数**。

正态分布的特征：

矩母函数 $\quad M(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}), \quad t \in \mathbb{R}$

均值 $\quad \mu = \mu$

方差 $\quad \sigma^2 = \sigma^2$

命题 4.18

若随机变量 $X \sim N(\mu, \sigma^2)$，则其线性变换 $a, b \in \mathbb{R}, aX + b \sim N(a\mu + b, a^2\sigma^2)$。特别的，常对正态随机变量作标准化：$\frac{X-\mu}{\sigma} \sim N(0,1)$

证明

$$M_{aX+b}(t) = e^{bt} M_X(at) = \exp(bt + a\mu t + \frac{a^2\sigma^2 t^2}{2})$$

命题 4.19

若<u>独立</u>随机变量 $X_1, \cdots, X_k$ 满足 $X_i \sim N(\mu_i, \sigma_i^2)$，则其和 $Y = X_1 + \cdots + X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2)$

证明

$$M_{X_1+\cdots+X_n} = \prod_{i=1}^n M_{X_i}(t) = \exp(\sum_{i=1}^k \mu_i t + \sum_{i=1}^k \frac{\sigma_i^2 t^2}{2})$$

推论 4.2

若随机变量 $X_1, \cdots, X_k$ i. i. d. $\sim N(\mu, \sigma^2)$，则其平均 $\overline{X} = (X_1 + \cdots + X_k)/k \sim N(\mu, \frac{\sigma^2}{k})$

定理 4.3

若随机变量 $X_1, \cdots, X_k$ i.i.d $\sim N(\mu, \sigma^2)$，并且定义其样本均值为 $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$，其样本方差为 $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$，则有以下关系：

1. $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$，即 $\sqrt{n}(\overline{X_n} - \mu)/\sigma \sim N(0,1)$
2. 随机变量 $\overline{X_n}$ 与随机向量 $(X_1 - \overline{X_n}, \cdots, X_n - \overline{X_n})$ 独立
3. $\overline{X_n}, S_n^2$ 独立
4. $\frac{(n-1)S_n^2}{\sigma} \sim \chi_n$
5. $\frac{\sqrt{n}(\overline{X_n} - \mu)}{S_n} \sim t_{n-1}$

证明

4.3.2 卡方分布

定义 4.18

设 $Z_1, \ldots, Z_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $X = Z_1^2 + \ldots + Z_n^2$ 的分布称为 n **自由度的卡方分布** (Chi-square distribution)，记为 $X \sim \chi_n^2$.

卡方分布的特征：

参数 $n \in \mathbb{N}_+$

概率密度函数 $p(x) = \begin{cases} \frac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{y}{2}} & y > 0 \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t) = (\frac{1}{1-2t})^{\frac{n}{2}}$

均值 $\mu = n$

方差 $\sigma^2 = 2n$

实例

该密度函数的图像是一个只取非负值的偏态分布，见图 4.7。

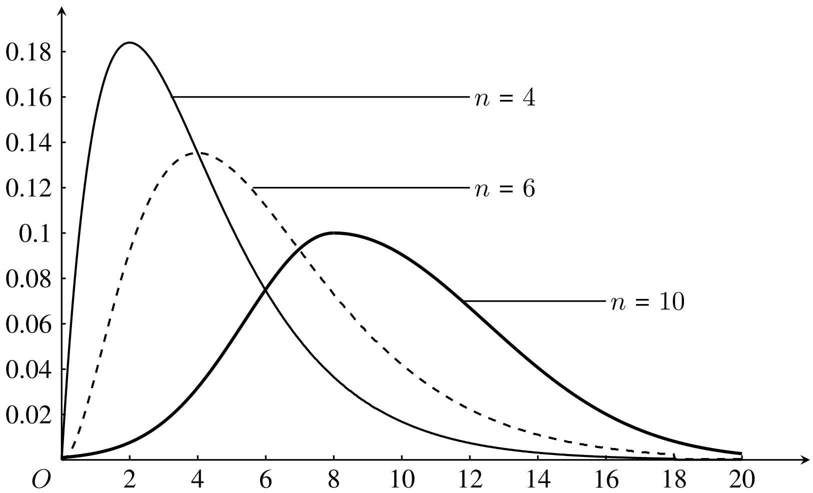


图 4.7: $\chi^2(n)$ 分布的密度函数

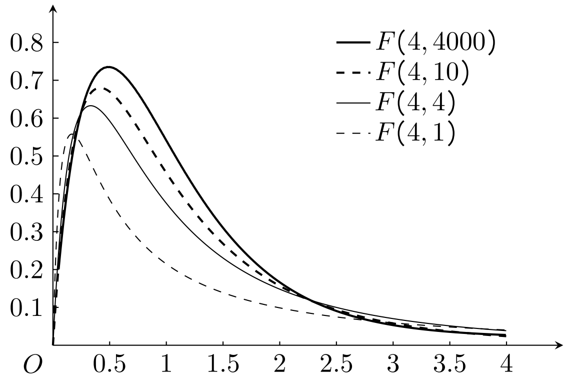


图 4.8: F 分布的密度函数，是一个只取非负值的偏态分布

命题 4.20

若 $X \sim N(0,1)$，则 $X^2 \sim Ga(1/2, 1/2)$

证明

$$
\begin{aligned}
f_{X^2}(y) &= \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] \\
&= 2\frac{1}{2\sqrt{y}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{y}{2}} \\
&= (\frac{1}{2})^{\frac{1}{2}-1}/\Gamma(\frac{1}{2})y^{\frac{1}{2}-1}\mathrm{e}^{-\frac{1}{2}y}
\end{aligned}
$$

推论 4.3

若随机变量 $X_1, \cdots, X_k, \mathrm{i.i.d.} \sim \chi^2_{n_i}$，则 $Y = X_1 + \cdots + X_k \sim \chi^2_{n_1+\cdots+n_k}$。

证明

4.3.3 F 分布

定义 4.19

设独立随机变量 $X_1 \sim \chi^2(m), X_2 \sim \chi^2(n)$，则称 $F = \frac{X_1/m}{X_2/n}$ 的分布是自由度为 m 与 n 的 **F 分布**，记为 $F \sim F(m,n)$，其中 m 称为分子自由度，n 称为分母自由度。

F 分布的特征：

参数 $m, n \in \mathbb{N}_+$

概率密度函数

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

矩母函数 不存在

均值 $\mu = \frac{n}{n-2}$

方差 $\sigma^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

实例

命题 4.21

F 分布的密度函数为

证明 首先我们导出 $Z=\frac{X_1}{X_2}$ 的密度函数, 若记 $p_1(x)$ 和 $p_2(x)$ 分别为 $\chi^2(m)$ 和 $\chi^2(n)$ 的密度函数, 根据独立随机变量商的分布的密度函数公式2.9。Z 的密度函数为

$$\begin{aligned}
p_Z(z) &= \int_0^{=\infty} x_2 p_1(zx_2) p_2(x_2)\,\mathrm{d}x_2 \\
&= \frac{z^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{\frac{m+n}{2}}} \int_0^{+\infty} x_2^{\frac{m+n}{2}-1} \mathrm{e}^{-\frac{x_2}{2}(1+z)}\,\mathrm{d}x_2.
\end{aligned}$$

运用变换 $u=\frac{x_2}{2}(1+z)$, 可得

$$p_Z(z) = \frac{z^{\frac{m}{2}-1}(1+z)^{\frac{m+n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} u^{\frac{m+n}{2}-1}\mathrm{e}^{-u}\,\mathrm{d}u.$$

最后的定积分为伽马函数 $\Gamma\left(\frac{m+n}{2}\right)$, 从而

$$p_Z(z) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}, \quad z>0.$$

第二步, 我们导出 $F=\frac{n}{m}Z$ 的密度函数, 对 $y>0$, 有

$$\begin{aligned}
p_F(y) &= p_Z\left(\frac{m}{n}y\right)\cdot\frac{m}{n} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}y\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}y\right)^{-\frac{m+n}{2}}\cdot\frac{m}{n} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} y^{\frac{m}{2}-1}\left(1+\frac{m}{n}y\right)^{-\frac{m+n}{2}}.
\end{aligned}$$

注 由 F 分布的构造知, 若 $F\sim F(m,n)$, 则有 $1/F\sim F(n,m)$

4.3.4 t 分布

定义 4.20

设随机变量 $Z\sim N(0,1)$ 与 $U\sim\chi_n^2$ 独立, 则称 $t=\frac{Z}{\sqrt{U/n}}$ 的分布为自由度为 n 的 **t 分布**, 记为 $t\sim t_n$.

t 分布的密度函数的图像是一个关于纵轴对称的分布 (图 4.9), 与标准正态分布的密度函数形状类似, 只是峰比标准正态分布低一些, 尾部的概率比标准正态分布的大一些.

t 分布的特征:

参数	$n\in\mathbb{N}_+$
概率密度函数	$p(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}}$
矩母函数	不存在
均值	$\mu=0, n>1$
方差	$\sigma^2=\frac{n}{n-2}, n>2$
实例	

注

- 自由度 $n=1$ 的 t 分布就是标准柯西分布，它的均值不存在;

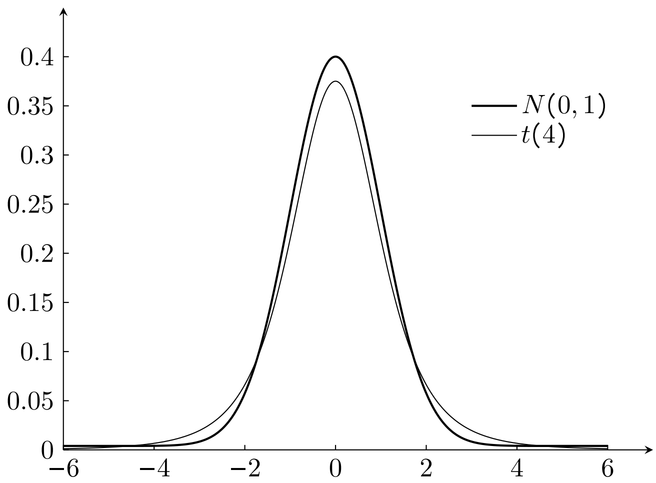


图 4.9: t 分布与 $N(0,1)$ 的密度函数，t 分布尾更重

- 自由度 $n \to \infty$ 时，t_n 趋近于 $N(0,1)$

命题 4.22

t 分布的密度函数为

证明 由标准正态密度函数的对称性知, X_1 与 $-X_1$ 有相同分布, 从而 t 与 $-t$ 有相同分布. 这意味着: 对任意实数 y 有

$$P(0<t<y)=P(-y<-t<y)=P(-y<-t<0).$$

于是

$$P(0<t<y)=\frac{1}{2}P(t^2<y^2).$$

由 F 变量构造可知, $t^2=\frac{X_1^2}{X_2^2/n}\sim F(1,n)$, 将上式两边关于 y 求导可微 t 分布的密度函数为

$$\begin{aligned}p_t(y)=yp_F(y^2)&=\frac{\Gamma\left(\frac{1+n}{2}\right)\left(\frac{1}{n}\right)^{\frac{1}{n}}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(y^2\right)^{\frac{1}{2}-1}\left(1+\frac{1}{n}y^2\right)^{-\frac{1+n}{2}}y\\&=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{y^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<y<+\infty.\end{aligned}$$

4.3.5 多元正态分布

定义 4.21

若 n 维随机向量 $\mathbf{X}=(X_1,X_2,\ldots,X_n)'$ 的联合密度函数为：

$$f(\mathbf{x})=f(x_1,\ldots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

其中 Σ 为对称非负定矩阵，则称 $\mathbf{X}$ 满足 n 元正态分布，记为 $\mathbf{X}\sim N(\mu,\Sigma)$

当 $n=2$ 时，若取数学期望向量和协方差矩阵分别为

$$\mathbf{a}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\quad \Sigma=\begin{pmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho\\ \sigma_1\sigma_2\rho & \sigma_2^2\end{pmatrix}$$

代入，则可得到二元正态密度函数。

4.4 各分布间关系

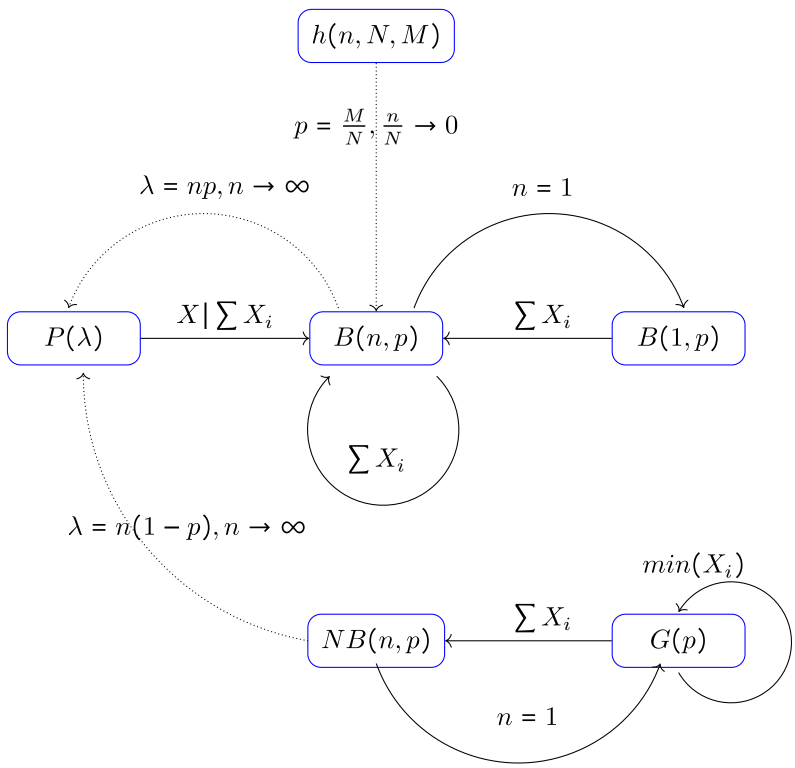


图 4.10: 离散分布间的关系

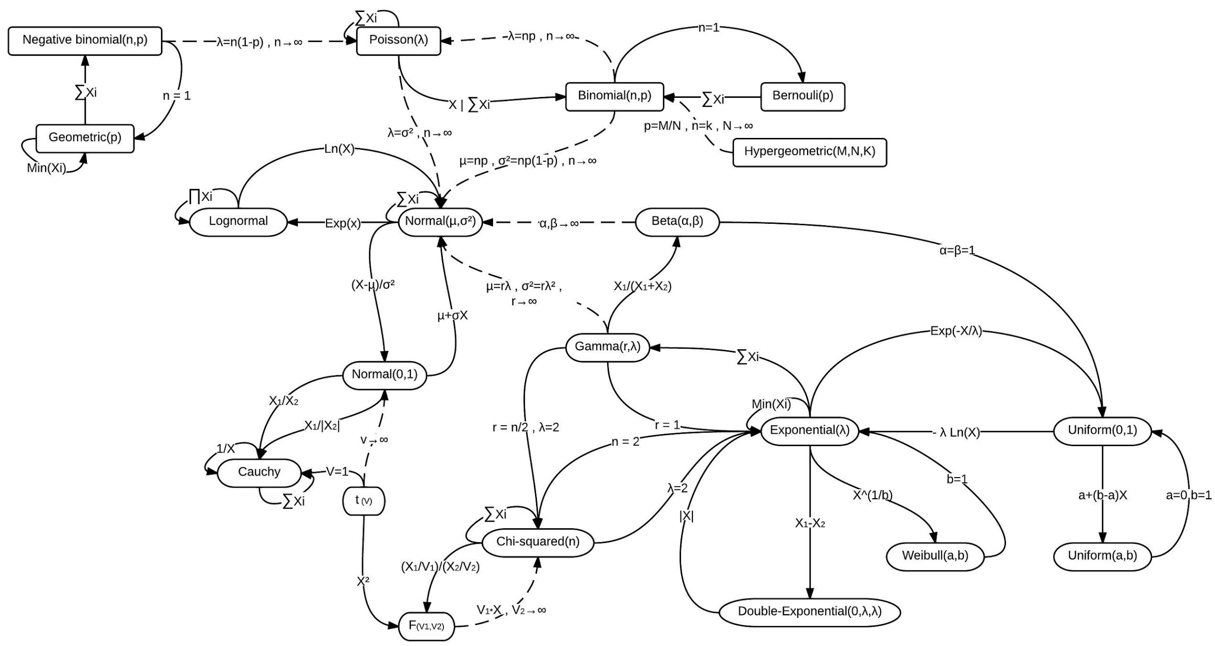


图 4.11: 各分布间的联系

可参考网站http://www.math.wm.edu/~leemis/chart/UDR/UDR.html

离散时间	连续时间
几何分布	指数分布
负二项分布	伽马分布
二项分布	泊松分布

4.4.1 伯努利过程与泊松过程

若每隔 Δt 进行一次试验，则伯努利试验也可以看作一个随时间而变化的过程。在伯努利试验中，到时刻 $n\Delta t$ 为止，共进行 n 次试验，这时成功次数服从二项分布. 而在泊松过程中，到时刻 t 的来到数则服从泊松分布为等待第一次成功，伯努利试验中的等待时间服从几何分布；而泊松过程中则服从指数分布. 它们都有无记忆性. 为等待第 r 次成功，伯努利试验中的等待时间服从帕斯卡分布；而泊松过程中则服从埃尔朗分布. 正如二项分布的泊松逼近，

错题记录

1. （茆 2.4.9）已知某商场一天来的顾客数 X 服从参数为 λ 的泊松分布，而每个来到商场的顾客购物的概率为 p，证明：此商场一天内购物的顾客数服从参数为 λp 的泊松分布.
2. （茆 2.5.8）统计调查表明，英格兰在 1875 年至 1951 年期间，在矿山发生 10 人或 10 人以上死亡的两次事故之间的时间 T（以日计）服从均值为 241 的指数分布，试求 $P(50 < T < 100)$.
3.

第 5 章 概率极限

考试重点

- 收敛性
- Borel-Cantelli 引理
- 各种概率不等式
- 特征函数在处理概率极限定理中的应用
- 各种大数定律及证明
- 中心极限定理

5.1 收敛

5.1.1 几乎必然收敛

定义 5.1 (几乎必然收敛)

若随机变量序列 $\{X_n : \Omega \to \mathbb{R}\}$ 与随机变量 $X : \Omega \to \mathbb{R}$ 间存在以下关系：

$$\mathbb{P}\{\omega \in \Omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon\} = 1, \quad \forall \varepsilon > 0$$

则称 X_n **几乎必然收敛** (converges almost surely) 于 X, 记为 $X_n \xrightarrow{\text{a.s.}} X$。

注 类似于微积分中逐点收敛的条件。考察满足条件 $\lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon$ 的样本点，这些样本点组成事件的概率为 1。

5.1.2 依概率收敛

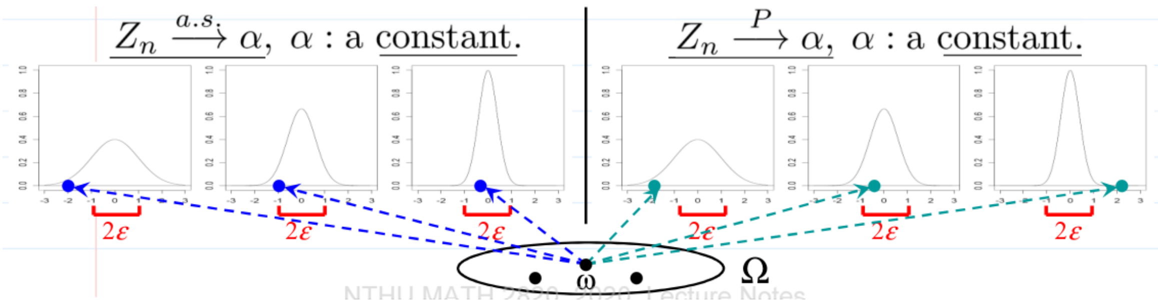
定义 5.2 (依概率收敛)

若随机变量序列 $\{X_n : \Omega \to \mathbb{R}\}$ 与随机变量 $X : \Omega \to \mathbb{R}$ 间存在以下关系：

$$\lim_{n\to\infty} \mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\} = 1, \quad \forall \varepsilon > 0.$$

则称 X_n **依概率收敛** (converges in probability) 于 X, 记为 $X_n \xrightarrow{\mathbb{P}} X$。

注 即对于事件 $A_i = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}$，其概率将逐渐变为 1。



对于几乎必然收敛的情况，每个样本点上的序列随机变量都在逐渐趋近收敛变量；而对于依概率收敛的情况，由于对某个样本点的概率为 0（例如连续随机变量），可能某一次样本点的序列随机变量接近收敛变量，下一次又远离收敛变量，只要保证总体概率趋近 1 即可。

命题 5.1

依概率收敛是几乎必然收敛的必要条件，但不是充分条件。

例题 5.1 依概率收敛而不几乎必然收敛 设样本空间为 $\Omega=(0,1]$，概率均匀分布在样本空间上，即 $P([a,b])=b-a, 0<a<b<1$。接下来对于 $k\in\mathbb{N}_+$，将区间 $(0,1]$ 分为 2^k 个等长的子区间，并分别记为 $I_{k,j}=(\frac{j-1}{2^k},\frac{j}{2^k}]$。

按以下方式定义随机变量序列 $\{X_n:\Omega\to\mathbb{R}\}$：

$$X_n(\omega)=\begin{cases}1,\omega\in I_{k,j}\\0,\omega\notin I_{k,j}\end{cases},\quad n=2^k+j-2$$

同时定义随机变量 $X:\Omega\to\mathbb{R}$ 为 $X(\omega)=0,\forall\omega\in\Omega$

由于

$$\mathbb{P}\{\omega\in\Omega:|X_n(\omega)-X(\omega)|<\varepsilon\}=1-\frac{1}{2^k}\to 1,\quad \forall 0<\epsilon<1$$

所以 $X_n\xrightarrow{\mathbb{P}}X$。然而对于任意 $\omega\in\Omega$，无论 k 为何数，此次分割中，总有一个区间包含此样本点。即有无数个 $X_i(\omega)=1$。所以

$$\mathbb{P}\{\omega\in\Omega:\lim_{n\to\infty}|X_n(\omega)-X(\omega)|<\epsilon\}=\mathbb{P}\{\varnothing\}=0$$

所以 X_n 不几乎必然收敛到 X。

引理 5.1

若 $X_n\xrightarrow{\mathbb{P}}X, Y_n\xrightarrow{\mathbb{P}}Y$，则：

1. $X_n\pm Y_n\xrightarrow{\mathbb{P}}X\pm Y$
2. $X_nY_n\xrightarrow{\mathbb{P}}XY$

证明

1. 因为

$$\{|X_n+Y_n-X-Y|\geq\varepsilon\}\subset\{|X_n-X|\geq\frac{\varepsilon}{2}\}\cup\{|Y_n-Y|\geq\frac{\varepsilon}{2}\}$$

所以

$$0\leq P\{|X_n+Y_n-X-Y|\geq\varepsilon\}\leq P\{|X_n-X|\geq\frac{\varepsilon}{2}\}+P\{|Y_n-Y|\geq\frac{\varepsilon}{2}\}$$

对上式取极限，并由夹逼定理得：

$$P\{|X_n+Y_n-X-Y|\geq\varepsilon\}=0$$

由 $Y_n\xrightarrow{\mathbb{P}}Y$ 易得 $-Y_n\xrightarrow{\mathbb{P}}-Y$，所以又有 $X_n-Y_n\xrightarrow{\mathbb{P}}X-Y$。

2. 由于 $2ab=(a+b)^2-a^2+b^2$，结合上一结论可知：$X_nY_n\xrightarrow{\mathbb{P}}XY$ 的充分条件是 $X_n\xrightarrow{\mathbb{P}}X\implies X_n^2\xrightarrow{\mathbb{P}}X^2$。任取 $\varepsilon>0$，有

$$\begin{aligned}P\{|X_n^2-X^2|\geq\varepsilon\}=&P\{|X_n-X||X_n+X|\geq\varepsilon\}\\=&P(\{|X_n-X||X_n+X|\geq\varepsilon\}\cap\{|X_n+X|\leq M\})\\&+P(\{|X_n-X||X_n+X|\geq\varepsilon\}\cap\{|X_n+X|>M\})\\\leq&P\{|X_n-X|\geq\varepsilon/M\}+P\{|X_n+X|>M\}\end{aligned}$$

而

$$\begin{aligned}P\{|X_n+X|>M\}\leq&P\{|X_n-X|+|2X|>M\}\\=&P(\{|X_n-X|+|2X|>M\}\cap\{|X_n-X|<1\})\\&+P(\{|X_n-X|+|2X|>M\}\cap\{|X_n-X|\geq1\})\\\leq&P\{|2X|>M-1\}+P\{|X_n-X|\geq1\}\end{aligned}$$

将 M 取足够大，使得对于任取的 $\delta>0$，有 $P\{|X|>(M-1)/2\}<\delta$ 成立。并且由于 $X_n\xrightarrow{\mathbb{P}}X$，所以

$$\exists N,\forall n>N,\quad \text{s.t. } P\{|X_n-X|\geq\varepsilon/M\},P\{|X_n-X|\geq1\}<\delta$$

综上

$$\begin{aligned}P\{|X_n^2 - X^2| \geq \varepsilon\} &\leq P\{|X_n - X| \geq \varepsilon/M\} + P\{|X_n + X| > M\} \\ &\leq P\{|X_n - X| \geq \varepsilon/M\} + P\{|2X| > M-1\} + P\{|X_n - X| \geq 1\} \\ &< 3\delta\end{aligned}$$

即 $X_n^2 \xrightarrow{\mathbb{P}} X^2$，所以 $X_nY_n \xrightarrow{\mathbb{P}} XY$

定理 5.1

若 $X_n \xrightarrow{\mathbb{P}} X$，$g(x)$ 是 $\mathbb{R}$ 上的连续函数，则 $g(X_n) \xrightarrow{\mathbb{P}} g(X)$

证明 若 $g_m(x)$ 是 m 次多项式函数，即 $g_m(x) = \sum_{i=0}^m a_i x^i$，则由引理5.1知有

$$g_m(X_n) \xrightarrow{\mathbb{P}} g_m(X)$$

因为 $g(x)$ 是连续函数，所以可以用多项式函数去逼近 $g(x)$，并且在任意有限区间上还是一致的，即对于任意 $M, \varepsilon > 0$，当 m 足够大的时候，存在 $g_m(x)$，满足

$$|g(x) - g_m(x)| < \varepsilon,\ |x| \leq M$$

接下来以 $|X|, |X_n| \leq M$ 为界，将目标拆为两部分：

$$\begin{aligned}&P\{|g(X_n) - g(X)| \geq \varepsilon\} \\ =&P(\{|g(X_n) - g(X)| \geq \varepsilon\} \cap \{|X|, |X_n| \leq M\}) \\ &+ P(\{|g(X_n) - g(X)| \geq \varepsilon\} \cap \{\{|X| \leq M\} \cup \{|X_n| \leq M\}\}) \\ \leq&P(\{|g(X_n) - g(X)| \geq \varepsilon\} \cap \{|X|, |X_n| \leq M\}) + P\{|X| \leq M\} + P\{|X_n| \leq M\}\end{aligned}$$

对于后半部分有：

$$\forall \delta > 0, \exists M > 0, \quad \text{s.t. } P\{|X| \leq M\}, P\{|X_n| \leq M\} < \delta$$

对于第一项，将 $\{|g(X_n) - g(X)| \geq \varepsilon\}$ 拆分为三部分：

$$\{|g(X_n) - g(X)| \geq \varepsilon\} \subset \{|g(X_n) - g_m(X_n)| \geq \varepsilon/3\} \cup \{|g_m(X_n) - g_m(X)| \geq \varepsilon/3\} \cup \{|g_m(X) - g(X)| \geq \varepsilon/3\}$$

又因为多项式函数逼近的一致性，稍加替换可得：

$$\begin{aligned}\{|g(X_n) - g_m(X_n)| \geq \varepsilon/3\} \cap \{|X| \leq M\} \cap \{|X_n| \leq M\} &= \varnothing \\ \{|g_m(X) - g(X)| \geq \varepsilon/3\} \cap \{|X| \leq M\} \cap \{|X_n| \leq M\} &= \varnothing\end{aligned}$$

所以

$$\begin{aligned}&P(\{|g(X_n) - g(X)| \geq \varepsilon\} \cap \{|X|, |X_n| \leq M\}) \\ \leq&P(\{|g_m(X_n) - g_m(X)| \geq \varepsilon/3\} \cap \{|X| \leq M\} \cap \{|X_n| \leq M+1\}) \\ \leq&P\{|g_m(X_n) - g_m(X)| \geq \varepsilon/3\} < \delta\end{aligned}$$

综上

$$P\{|g(X_n) - g(X)| \geq \varepsilon\} < 4\delta$$

即 $g(X_n) \xrightarrow{\mathbb{P}} g(X)$

笔记 相关证明技巧总结：对于事件 $\{|X+Y| > a+b\}$，有 $\{|X+Y| > a+b\} \subset \{|X| > a\} \cup \{|Y| > b\}$，从而有

$$P\{|X+Y| > a+b\} \leq P\{|X| > a\} + P\{|Y| > b\}$$

5.1.3 依分布收敛

定义 5.3 (依分布收敛)

设随机变量 $X, X_1, X_2, \ldots$ 的分布函数分别为 $F(x), F_1(x), F_2(x), \ldots$。若对 $F(x)$ 的任一连续点x，都有

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

则称 $\{F_n(x)\}$ **弱收敛** (converge weakly) 于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$；也称 $\{X_n\}$ **按分布收敛** (converge in distribution) 于 X，记作 $X_n \xrightarrow{d} X$

注 对于非连续点，其收敛值可能使得收敛函数不满足分布函数特征（见下例），需要定义这些非连续点上的值，使其满足分布函数特征。

例题 5.2 设 $\{X_n\}$ 服从退化分布：

$$P\left(X_n = \frac{1}{n}\right) = 1, \quad n \in N_+$$

其分布函数分别为：

$$F_n(x) = \begin{cases} 0, & x < 1/n, \\ 1, & x \geq 1/n. \end{cases}$$

因为 $F_n(x)$ 是在点 $x = 1$ 处有跳跃，所以当 $n \to +\infty$ 时，跳跃点位置趋于 0。于是我们很自然地认为 $\{F_n(x)\}$ 应该收敛于点 $x = 0$ 处的退化分布，即

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

但是，对任意的 n，有 $F_n(0) = 0$, 而 $F(0) = 1$, 所以

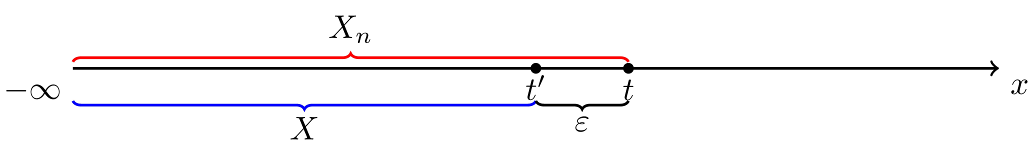
$$\lim_{n\to+\infty} F_n(0) = 0 \neq 1 = F(0)$$

命题 5.2

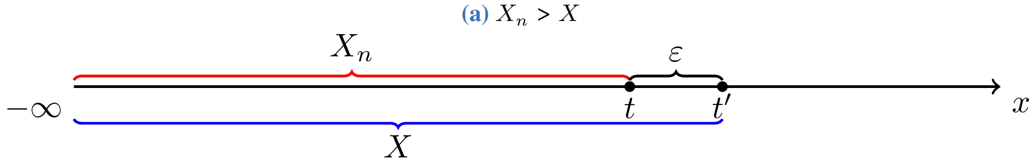
依分布收敛是依概率收敛的必要条件，但不是充分条件。但若依概率收敛到一个常数随机变量，则也可推出依分布收敛。

注 直观来看，X_n 依概率收敛到 X，意味着 n 很大时，二者的距离大概率很近。那么随着 n 越来越大，X_n 的行为自然会越来越像。对于某一区间 B，样本点集 $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ 将与点集 $X_n^{-1}(B)$ 越来越接近，其两随机变量出现在此区间的概率也将逐渐相等，这就是依分布收敛描述的东西。

证明 **必要性**：设随机变量 $X, X_1, X_2, \ldots$ 的分布函数分别为 $F(x), F_1(x), F_2(x), \ldots$。



(a) $X_n > X$



(b) $X_n < X$

如上图5.1a，直观来看 $P\{X_n \le t\} \approx F(t-\varepsilon)$。而严格来看有：

$$\begin{aligned}F_n(t) &= P(X_n \le t)\\ &\ge P(X_n \le t, X \le t-\varepsilon)\\ &= P(X \le t-\varepsilon) - P(X_n > t, X \le t-\varepsilon)\\ &\ge F(t-\varepsilon) - P(|X_n - X| > \varepsilon)\end{aligned}$$

数列 $\{F_n(t)\}$ 的极限未必存在，但可对两边取下极限得

$$F(t-\varepsilon) \le \varliminf_{n\to+\infty} F_n(t)$$

同理，如上图5.1b，根据“对称性”，交换 X_n 与 X 的位置有：

$$F(t+\varepsilon) \ge F_n(t) - P(|X_n - X| > \varepsilon)$$

对两边取下极限得

$$\varlimsup_{n\to+\infty} F_n(t) \le F(t+\varepsilon)$$

综上有：

$$F(t-\varepsilon) \le \varliminf_{n\to+\infty} F_n(t) \le \varlimsup_{n\to+\infty} F_n(t) \le F(t+\varepsilon)$$

再令 $\varepsilon \to 0+$ 得

$$F(t-0) \le \varliminf_{n\to+\infty} F_n(t) \le \varlimsup_{n\to+\infty} F_n(t) \le F(t+0)$$

所以当 t 为 $F(x)$ 上的连续点时，$F(t-0) = F(t+0)$。根据夹逼定理，此时 $\{F_n(t)\}$ 的极限存在且

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

充分性（常数随机变量）：设 $X = c$,c 为常数，其分布函数退化为：

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

所以对任意的 $\varepsilon > 0$, 有

$$\begin{aligned}P\{|X_n - X| \ge \varepsilon\} &= P\{X_n \ge c+\varepsilon\} + P\{X_n \le c-\varepsilon\}\\ &\le P\{X_n > c + \frac{\varepsilon}{2}\} + P\{X_n \le c-\varepsilon\}\\ &= 1 - F_n(c+\varepsilon/2) + F_n(c-\varepsilon).\end{aligned}$$

由于 $x = c+\varepsilon/2$ 和 $x = c-\varepsilon$ 均为 $F(x)$ 的连续点，且 $F_n(x) \xrightarrow{W} F(x)$，所以当 $n \to +\infty$ 时，有

$$F_n(c+\varepsilon/2) \to F(c+\varepsilon/2) = 1, \quad F_n(c-\varepsilon) \to F(c-\varepsilon) = 0$$

由此得

$$P\{|X_n - X| \ge \varepsilon\} \to 0$$

例题 5.3 依分布收敛而不依概率收敛 设 X 的分布列为

$$P(X=-1) = \frac{1}{2}, \quad P(X=1) = \frac{1}{2}$$

令 $X_n = -X$, 则 X_n 与 X 同分布。即 X_n 与 X 有相同的分布函数。故 $X_n \xrightarrow{L} X$。

但对任意 $0 < \varepsilon < 2$，有

$$P\{|X_n - X| \ge \varepsilon\} = P(2|X| \ge \varepsilon) = 1 \nrightarrow 0$$

即 X_n 不是依概率收敛于 X。

$$
\begin{tikzcd}
\text{几乎必然收敛} \arrow[r] & \text{依概率收敛} \arrow[r] & \text{按分布收敛} \arrow[l, bend right, dotted, "\text{常数随机变量}"']
\end{tikzcd}
$$

图 5.2: 三种收敛的关系

定理 5.2 (连续映射定理)

设 $g:\mathbb{R}^k\mapsto\mathbb{R}$ 为区间 B 上的连续 函数，并且 $P(X\in B)=1$，则随机向量序列通过次映射后收敛情况与原先相同，即：

$$X_n^{(i)}\xrightarrow{\text{a.s.}}X^{(i)}\Rightarrow g(X_n^{(1)},\cdots X_n^{(k)})\xrightarrow{\text{a.s.}}g(X^{(1)},\cdots X^{(k)})$$

$$X_n^{(i)}\xrightarrow{\mathbb{P}}X^{(i)}\Rightarrow g(X_n^{(1)},\cdots X_n^{(k)})\xrightarrow{\mathbb{P}}g(X^{(1)},\cdots X^{(k)})$$

$$(X_n^{(i)})\xrightarrow{d}(X^{(i)})\Rightarrow g(X_n^{(1)},\cdots X_n^{(k)})\xrightarrow{d}g(X^{(1)},\cdots X^{(k)})$$

证明 证明见书 **[vaart_1998]** 中定理 2.3。

推论 5.1 (Slutsky 定理)

若 $X_n\xrightarrow{d}X, Y_n\xrightarrow{P}a$，其中 a 是常数，则：

- $Y_nX_n\xrightarrow{d}aX$
- $Y_n+X_n\xrightarrow{d}a+X$
- $\frac{X_n}{Y_n}\xrightarrow{d}\frac{X}{a}$，若对所有的 n 都有 $P(Y_n\neq 0)=1$，并且 $a\neq 0$

证明 由条件可知，$(X_n,Y_n)\xrightarrow{d}(X,a)$，且 $g_1(X_n,Y_n)=X_nY_n, g_2(X_n,Y_n)=X_n+Y_n, g_3(X_n,Y_n)=\frac{X_n}{Y_n}$ 都是连续函数（g_3 在定理中的限制下），所以仍然保持依分布收敛。

注 当 $Y_n\xrightarrow{P}a$ 才能得出 $(X_n,Y_n)\xrightarrow{d}(X,a)$；若 Y_n 依分布收敛到其他随机变量，则未必成立。

定理 5.3 (theta 方法的极限定理)

设

$$\frac{\sqrt{n}(X_n-\theta)}{\sigma}\xrightarrow{d}N(0,1)$$

那么对于连续函数 g，若 $g'(\theta)\neq 0$ 存在，则

$$\frac{\sqrt{n}(g(X_n)-g(\theta))}{\sigma|g'(\theta)|}\xrightarrow{d}N(0,1)$$

证明

定理 5.4 (连续性定理)

设 $F_n(x)$ 是一个累计函数序列，并且分别对应矩母函数 $M_n(t)$；$F(x)$ 是一个累计函数，并且对应矩母函数 $M(t)$。那么，$F(x)$ 的所有连续点上 $F_n(x)\xrightarrow{M}F(x)$ 的充要条件是：

$$\forall t\in U(0,\delta), s.t. \lim_{n\to\infty}M_n(t)=M(t)$$

其中 $U(0,\delta)$ 代表某一包含零点的开区间。

若矩母函数不存在，可替换成特征函数

$$\forall t\in\mathbb{R},,\quad \text{s.t.}\ \lim_{n\to\infty}\varphi_n(t)=\varphi(t)$$

注 不能替换成密度函数或质量函数

例题 5.4 离散情况：设均匀随机变量 $X_n \sim U(-\frac{1}{2}, \frac{1}{2})$，那么 $X_n \xrightarrow{d} 0$。然而 $P_n(x) \to 0, \forall x \in \mathbb{R}$（不满足质量函数定义），并且 $\lim_{n\to\infty} F_n(0) = \frac{1}{2}$（不满足累积函数定义）

连续情况：设累积函数为 $F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, 0 < x < 1$，那么 $F_n \xrightarrow{d} U(0,1)$。然而 $f_n(x)$ 无极限

例题 5.5 泊松分布收敛于正态分布 令 $X_n \sim P(\lambda_n)$，并且 $\lim_{n\to\infty} \lambda_n = \infty$。将变量标准化：

$$Z_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

那么

$$M_{Z_n}(t) = \mathrm{e}^{-t\sqrt{\lambda_n}} M_{X_n}(t)(\frac{t}{\sqrt{\lambda_n}}) = \exp(-t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}}-1}))$$

由于

$$\lim_{n\to\infty} \ln(M_{Z_n}(t)) = \lim_{\lambda_n\to\infty} -t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}}-1}) = \frac{t^2}{2}$$

所以 $Z_n \xrightarrow{d} N(0,1)$，即若 λ 很大，可通过 $N(\lambda,\lambda)$ 估计 $P(\lambda)$

5.2 大数定理

5.2.0.1 大数定理

大数定律讨论的是在什么条件下，随机变量序列的算术平均依概率收敛到其均值的算术平均。

定义 5.4 (大数定律的一般形式)

设有一随机变量序列 $\{X_n\}$，若满足：

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

则称该随机变量序列 $\{X_n\}$ 服从**大数定律**。

笔记 柯西分布无均值、方差，不存在此定理。

不同的大数定律的差别只是对不同的随机变量序列 $\{X_n\}$ 而言, 有的是相互独立的随机变量序列, 有的是相依的随机变量序列, 有的是同分布的随机变量序列, 有的是不同分布的随机变量序列等等.

定理 5.5 (切比雪夫大数定律)

设 $\{X_n\}$ 为一列两两不相关的随机变量序列，若每个 X_i 的方差存在，且有共同的上界，即 $\mathrm{Var}(X_i) \leq c$, $i = 1, 2, \ldots$，则 $\{X_n\}$ 服从大数定律5.4。

注 切比雪夫大数定律只要求 $\{X_n\}$ 互不相关, 并不要求它们是同分布的。假如 $\{X_n\}$ 是独立同分布的随机变量序列, 且方差有限, 则 $\{X_n\}$ 必定服从大数定律.

证明 因为 $\{X_n\}$ 两两不相关，故

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \leq \frac{c}{n}$$

再由切比雪夫不等式得：

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \geq \varepsilon\right) < \frac{\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)}{\varepsilon^2} < \frac{c}{n\varepsilon^2}, \quad \forall \varepsilon > 0$$

于是当 $n \to +\infty$ 时，有

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

定理 5.6 (马尔可夫大数定律)

对随机变量序列 $\{X_n\}$，若

$$\frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) \to 0$$

成立，则其服从大数定律。此条件被称为**马尔可夫条件**。

注 马尔可夫条件对 $\{X_n\}$ 已经没有任何同分布、独立性、不相关的假定。证明过程与切比雪夫大数定律类似；切比雪夫大数定律显然可由马尔可夫大数定律推出。

例题 5.6 设 $\{X_n\}$ 为一同分布、方差存在的随机变量序列，且 X_n 仅与 X_{n-1} 和 X_{n+1} 相关，而与其他的 X_i 不相关。试问该随机变量序列 $\{X_n\}$ 是否服从大数定律？

解 考虑其马尔可夫条件

$$\frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n}\operatorname{Var}(X_i) + 2\sum_{i=1}^{n-1}\operatorname{Cov}(X_i, X_{i-1})\right]$$

记 $\operatorname{Var}(X_i) = \sigma^2$，则 $\left|\operatorname{Cov}(X_i, X_j)\right| \le \sigma^2$，于是有

$$\frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) \le \frac{1}{n^2}\left[n\sigma^2 + 2(n-1)\sigma^2\right] \to 0$$

即马尔可夫条件成立，故 $\{X_n\}$ 服从大数定律.

定理 5.7 ((辛钦) 弱大数定理)

设 $\{X_n\}$ 为一独立同分布 的随机变量序列，若 X_i 的数学期望存在 且为 μ，则 $\{X_n\}$ 服从大数定律，即：

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu$$

证明 因为 $\{X_n\}$ 同分布，故其特征函数相同，记为 $\varphi(t)$。又因为 $\varphi'(0) = i\mathbb{E}(X_i)$，则其位于 0 点附近的展开式为：

$$\varphi(t) = \varphi(0) + \varphi'(0)t + o(t) = 1 + i\mu t + o(t)$$

由 $\{X_n\}$ 的独立性知 $\overline{X_n}$ 的特征函数为：

$$\varphi_{\overline{X_n}}(t) = \left[\varphi(\frac{t}{n})\right]^n = \left[1 + i\mu\frac{t}{n} + o(\frac{t}{n})\right]^n$$

对于任意 t 有：

$$\lim_{n\to\infty}\varphi_{\overline{X_n}}(t) = \mathrm{e}^{i\mu t}$$

再根据定理5.4与命题5.2，上述定理成立。

例题 5.7 蒙特卡洛积分（平均值法） 为计算积分 $I(f) = \int_0^1 f(x)\mathrm{d}x$，生成随机数 $X_1,\cdots,X_n$ i.i.d. $\sim U(0,1)$，并且计算 $\hat{I}(f) = \overline{Y_n}, Y_i = f(X_n)$。由于 $X_1,\cdots,X_n$ i.i.d.，所以 $Y_1,\cdots,Y_n$ i.i.d.；并且 $E(Y_i) = E(f(X_i)) = \int_0^1 f(x)\cdot 1\mathrm{d}x = I(f)$，所以 $\hat{I}(f) \xrightarrow{\mathbb{P}} E(Y_i) = I(f)$

例题 5.8 样本方差 设随机变量 $X_1,\cdots,X_n$ i.i.d.，并且均值和方差分别为 μ,σ。令：

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n}) = (\frac{1}{n}\sum_{i=1}^{n}X_i^2) - \overline{X_n}^2$$

由于 $g(x) = x^2$ 是连续函数，所以 $\overline{X_n}^2 \xrightarrow{\mathbb{P}} \mu^2$。由于 $X_1^2,\cdots X_n^2$ i.i.d.，且 $E(X_i^2) = \sigma^2 + \mu^2$，所以 $\overline{X_n^2} \xrightarrow{\mathbb{P}} \mu^2 + \sigma^2$。因此：

$$S_n^2 \xrightarrow{\mathbb{P}} \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

例题 5.9 若 $X_n \sim t_n$，则 $X_n \xrightarrow{d} N(0,1)$。令 $Z \sim N(0,1), U_1,\cdots,U_n \sim \chi_1^2$ 相互独立，则 $\frac{Z}{\sqrt{(U_1+\cdots+U_n)/n}} \sim t_n$。由

于 $E(U_i)=1$，所以 $\overline{U_n}\xrightarrow{\mathbb{P}}1$，所以 $\sqrt{\overline{U_n}}\xrightarrow{\mathbb{P}}1$，所以 $t_n=\frac{Z}{\sqrt{\overline{U_n}}}\xrightarrow{d}N(0,1)$（Slutsky 定理5.1）。

例题 5.10 若 $X_n\sim F_{m,n}$，则 $mX_n\xrightarrow{d}\chi_m^2$。令 $U\sim\chi_m^2, V_1,\cdots,V_n\sim\chi_1^2$ 相互独立，则 $\frac{U/m}{(V_1+\cdots+V_n)/n}\sim F_{m,n}$。由于 $E(V_i)=1$，所以 $\overline{V_n}\xrightarrow{\mathbb{P}}1$，所以 $mF_{m,n}=\frac{U}{\overline{V_n}}\xrightarrow{d}\chi_m^2$（Slutsky 定理5.1）。

5.2.0.2 强大数定理

大数定律讨论的是在什么条件下，随机变量序列的算术平均几乎必然收敛到其均值的算术平均。

定义 5.5 (强大数定律的一般形式)

设有一随机变量序列 $\{X_n\}$，若满足：

$$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{\text{a.s.}}\frac{1}{n}\sum_{i=1}^{n}E(X_i)$$

则称该随机变量序列 $\{X_n\}$ 服从**强大数定律**。♣

5.3 中央极限定理

中央极限定理讨论在什么条件下，独立随机变量和的分布函数会收敛到正态分布。

5.3.1 独立同分布下的中心极限定理

定理 5.8 (林德伯格－莱维 (Lindeberg-Levy) 中央极限定理)

设随机变量 $X_1,\cdots,X_n,\text{i.i.d.}$，且 $E(X_i)=\mu,\text{Var}(X_i)=\sigma^2$，令

$$\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n}X_i, T_n=\sum_{i=1}^{n}X_i$$

则：

$$\lim_{n\to\infty}P(\frac{T_n-n\mu}{\sigma\sqrt{n}}\leq x)=\lim_{n\to\infty}P(\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\leq x)=\Phi(x)$$

其中 $\Phi(x)$ 是 $N(0,1)$ 的累积函数，即：

$$\frac{T_n-n\mu}{\sigma\sqrt{n}}=\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\xrightarrow{d}N(0,1)$$

♡

证明 设 $X_n-\mu$ 的特征函数为 $\varphi(t)$。因为 $E(X_n-\mu)=0,\ \text{Var}(X_n-\mu)=\sigma^2$, 所以有

$$\varphi'(0)=0,\ \varphi''(0)=-\sigma^2$$

于是特征函数 $\varphi(t)$ 有展开式

$$\varphi(t)=\varphi(0)+\varphi'(0)t+\varphi''(0)\frac{t^2}{2}+o(t^2)=1-\frac{1}{2}\sigma^2t^2+o(t^2)$$

则 $\{Y_n^*\}$ 的特征函数为

$$\varphi_{Y_n^*}(t)=\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n=\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n$$

从而有

$$\lim_{n\to+\infty}\varphi_{Y_n^*}(t)=\mathrm{e}^{-t^2/2}$$

而 $\mathrm{e}^{-t^2/2}$ 正是 $N(0,1)$ 分布的特征函数，根据定理5.4，上述定理得证。

例题 5.11 测量误差估计 假设每次测量结果为独立同分布的随机变量 $X_1,\cdots,X_n$, 其均值与方差分别为 μ,σ^2. 由

中央极限定理可知 $\frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma} \xrightarrow{d} N(0,1)$. 由例5.8可知 $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$, 即 $\frac{\sigma}{S_n} \xrightarrow{\mathbb{P}} 1$(定理5.1). 所以:

$$\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n} = \frac{\sqrt{n}(\overline{X_n}-\mu)}{\sigma}\frac{\sigma}{S_n} \xrightarrow{d} N(0,1)\cdot 1 = N(0,1)$$

即可以通过分布 $N(0,\frac{S_n^2}{n})$ 估计测量误差 $\overline{X_n}-\mu$

注 由定理4.3可知 $\frac{\sqrt{n}(\overline{X_n}-\mu)}{S_n} \sim t_{n-1}$. 但由例5.9可知,$n\to\infty$ 时,t 分布收敛于正态分布, 所以不冲突.

推论 5.2 (棣莫弗 – 拉普拉斯 (de Moivre-Laplace) 中心极限定理)

若随机变量 $X_1,\cdots,X_n$ i.i.d. $\sim B(1,p)$, 则 $T_n \sim B(n,p)$. 其中 $E(T_n)=nE_(X_i)=np, \mathrm{Var}(T_n)=n\,\mathrm{Var}(X_i)=np(1-p)$。根据中央极限定理有:

$$\frac{T_n-np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

即当 n 很大时，$B(n,p)$ 可用 $N(np,np(1-p))$ 近似。

注 其他与二项分布一样, 可以通过独立同分布的随机变量相加得到的分布, 也可在一定条件下近似为正态分布. 例如 Gamma 分布 (可通过指数分布相加得到), 泊松分布 (泊松分布相加), 负二项分布 (几何分布相加), 参见图4.11

注 由定理4.1可知，当 $B(n,p)$ 中 n 很大 p 很小时，二项分布可近似为泊松分布；而泊松分布 λ 很大时，又可近似为正态分布。两者相比，一般在 p 较小时，用泊松分布近似较好；而在 $np>5$ 和 $n(1-p)>5$ 时，用正态分布近似较好。

因为二项分布是离散分布，而正态分布是连续分布，所以用正态分布作为二项分布的近似计算中，作些修正可以提高精度。若 $k_1<k_2$ 均为整数，一般先作如下修正后再用正态近似 $P(k_1\le T_n\le k_2)=P(k_1-0.5<T_n<k_2+0.5)$。

例题 5.12 若 $T_n\sim b(25,0.4)$，通过以下修正的正态近似得到 $P(5\le\mu_n\le15)$：

$$\begin{aligned}P(5\le T_n\le 15) &= P(5-0.5<T_n<15+0.5)\\ &\approx \Phi\left(\frac{15+0.5-10}{\sqrt{6}}\right)-\Phi\left(\frac{5-0.5-10}{\sqrt{6}}\right)\\ &= 2\Phi(2.245)-1=0.9754.\end{aligned}$$

5.3.2 独立不同分布下的中心极限定理

设 $\{X_n\}$ 是一个相互独立的随机变量序列，它们具有有限的数学期望和方差：

$$E(X_i)=\mu_i,\ \mathrm{Var}(X_i)=\sigma_i^2$$

为讨论随机变量的和 $Y_n=\sum_{i=1}^n X_i$，先将其标准化。由于

$$\begin{aligned}E(Y_n) &= \mu_1+\mu_2+\cdots+\mu_n,\\ \sigma(T_n) &= \sqrt{\mathrm{Var}(X_i)}=\sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}\end{aligned}$$

且记 $\sigma(Y_n)=S_n$，则 Y_n 的标准化为

$$Y_n^*=\frac{Y_n-(\mu_1+\mu_2+\cdots+\mu_n)}{S_n}=\sum_{i=1}^n\frac{X_i-\mu_i}{S_n}$$

如果要求 Y_n^* 中各项 $(X_i-\mu_i)/S_n$ “均匀地小”，即对任意的 $\tau>0$，要求事件

$$A_k=\left\{\frac{|X_i-\mu_i|}{S_n}>\tau\right\}=\{|X_i-\mu_i|>\tau S_n\}$$

发生的可能性小，或直接要求其概率趋于 0。为达到这个目的，要求

$$\lim_{n\to+\infty}P\left\{\max_{1\le i\le n}|X_i-\mu_i|>\tau S_n\right\}=0.$$

因为

$$P\left\{\max_{1\le i\le n}|X_i-\mu_i|>\tau S_n\right\}=P\left\{\cup_{i=1}^n(|X_i-\mu_i|>\tau S_n)\right\}\le\sum_{i=1}^n P(|X_i-\mu_i|>\tau S_n)$$

不妨考虑

$$P(|X_i-\mu_i|>\tau S_n)=\mathbb{E}[\mathbb{1}(|X_i-\mu_i|>\tau S_n)]\le\mathbb{E}[\frac{(x-\mu_i)^2}{\tau^2S_n^2}\cdot\mathbb{1}(|X_i-\mu_i|>\tau S_n)]$$

只要上式右侧之和趋于零，就可保证 Y_n^* 中各加项“均匀地小”，故有以下定义：

定义 5.6 (林德贝格条件)

若 $\{X_n\}$ 是一个相互独立的随机变量序列，其均值与方差分别为 μ_i,σ_i^2。记 $S_n=\sqrt{\sigma_1^2+\sigma_2^2+\cdots+\sigma_n^2}$，若对任意的 $\tau>0$，有

$$\lim_{n\to+\infty}\frac{1}{\tau^2S_n^2}\sum_{i=1}^n\mathbb{E}[(x-\mu_i)^2\cdot\mathbb{1}(|X_i-\mu_i|>\tau S_n)]=0$$

则称 $\{X_n\}$ 满足**林德贝格条件**。

定理 5.9 (林德贝格中心极限定理)

设独立随机变量序列 $\{X_n\}$ 满足林德贝格条件，则对任意的 x，有

$$\lim_{n\to+\infty}P\left\{\frac{1}{S_n}\sum_{i=1}^n(X_i-\mu_i)\le x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^x\mathrm{e}^{-t^2/2}\,\mathrm{d}t$$

命题 5.3

假如独立随机变量序列 $\{X_n\}$ 具有同分布和方差有限的条件，则必定满足林德贝格条件。也就是说定理5.8是定理 5.9 的特例。

证明 设连续随机变量序列 $\{X_n\}$ 独立同分布，其密度函数、均值、方差分别为为 $f(x),\mu_i=\mu,\sigma_i=\sigma$。这时 $S_n=\sigma\sqrt{n}$，由此得

$$\frac{1}{S_n^2}\sum_{i=1}^n\int_{|x_i-\mu_i|>\tau S_n}(x-\mu_i)^2f(x)\,\mathrm{d}x=\frac{n}{n\sigma^2}\int_{|x-\mu|>\tau\sigma\sqrt{n}}(x-\mu)^2f(x)\,\mathrm{d}x.$$

因为方差存在，即

$$\mathrm{Var}(X_i)=\int_{-\infty}^{+\infty}(x-\mu)^2f(x)\,\mathrm{d}x<+\infty.$$

所以其尾部积分一定有

$$\lim_{n\to+\infty}\int_{|x-\mu|>\tau\sigma\sqrt{n}}(x-\mu)^2f(x)\,\mathrm{d}x=0,$$

故林德贝格条件满足。

林德贝格条件虽然比较一般，但该条件较难验证，下面的李雅普诺夫条件则比较容易验证，因为它只对矩提出要求，因而便于应用。

定理 5.10 (李雅普诺夫中心极限定理)

设 $\{X_n\}$ 为独立随机变量序列，若存在 $\delta>0$，满足

$$\lim_{n\to+\infty}\frac{1}{S_n^{2+\delta}}\sum_{i=1}^n\mathbb{E}\left(|X_i-\mu_i|^{2+\delta}\right)=0,\tag{5.1}$$

则对任意的 x, 有

$$\lim_{n\to+\infty}P\left\{\frac{1}{S_n}\sum_{i=1}^n(X_i-\mu_i)\le x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^x\mathrm{e}^{-t^2/2}\,\mathrm{d}t$$

	大数定律	强大数定律	中心极限定理
收敛	依概率收敛	依概率 1 收敛	依分布收敛
伯努利实验	伯努利	博雷尔	棣莫弗 − 拉普拉斯
独立同分布	辛钦	科尔莫戈罗夫	林德伯格 − 莱维
	泊松		
一般场合	切比雪夫	科尔莫戈罗夫	林德伯格 − 费勒
	马尔可夫		

其中 μ_i 与 S_n 如前所述.

错题记录

1. (茆 4.1.2)
2. (茆 4.3.4) 在伯努利试验中, 事件 A 出现的概率为 p, 令
$$X+n=\begin{cases}1, & \text{若在第 } n \text{ 次及第 } n+1 \text{ 次试验中 } A \text{ 出现}\\ 0, & \text{其他.}\end{cases}$$
证明 $\{X_n\}$ 服从大数定律.
3. (茆 4.3.12) (伯恩斯组大数定律) 设 $\{X_n\}$ 是方差有界的随机变量序列, 且当 $|k-l| \to +\infty$ 时, 一致地有 $\mathrm{Cov}(X_k, X_l) \to 0$, 证明 $\{X_n\}$ 服从大数定律.
4. (茆 4.3.13) (格涅坚科大数定律) 设 $\{X_n\}$ 是随机变量序列, 若记
$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ a_n = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$$
则 $\{X_n\}$ 服从大数定律的充要条件是
$$\lim_{n\to+\infty} E\left(\frac{(Y_n-a_n)^2}{1+(Y_n-a_n)^2}\right)=0$$
5. (茆 4.3.14) 设 $\{X_n\}$ 为独立同分布的随机变量序列, 方差存在。又设 $\sum_{n=1}^{+\infty} a_n$ 为绝对收敛级数。令 $Y = \sum_{i=1}^{n} X_i$, 证明 $\{a_nY_n\}$ 服从大数定律。
6. (茆 4.3.16) 设 $\{X_n\}$ 为独立同分布的随机变量序列, 其方差有限, 且 X_n 不恒为常数。如果 $S_n = \sum_{i=1}^{n} X_i$, 试证: 随机变量序列 $\{S_n\}$ 不服从大数定律.
7. (李 5.25)
8. (李 5.45)
9. (李 5.46)
10. (李 5.60)

附录 A 测度论基础

简而言之，测度论可以理解为在抽象空间建立类似于实变函数中测度、积分和导数那样的分析系统。

A.1 可测空间和可测映射

A.1.1 集合及其运算

考虑一个任意非空集合 X，称之为**空间**。X 的子集以 $A.B, C, \cdots$ 等记之，称之为这个空间的**集合**。空集记为 $\varnothing$。X 的成员称为**元素**。元素 x 属于集合 A，记作 $x \in A$；反之，元素 x 不属于集合 A，则用记号 $x \notin A$ 来表示。

定义 A.1

空间 X 上定义的实函数

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

称为 A 的**指示函数**。

定义 A.2 (集合的运算)

给定集合 A 和 B，集合

余	$A^{\mathsf{C}} := \{x \in X \vert x \notin A\}$
并	$A \cup B := \{x \vert (x \in A) \vee (x \in B)\}$
交	$A \cap B := \{x \vert (x \in A) \wedge (x \in B)\}$
差	$A \setminus B := \{x \vert (x \in A) \wedge (x \notin B)\}$
对称差	$A \triangle B := (A \setminus B) \cup (B \setminus A)$

定义 A.3 (集合的极限)

设 $\{A_n\}$ 是一个集合序列，

1. 若

$$\forall n \in N_+, \quad A_{n+1} \subseteq A_n$$

则称 $\{A_n\}$ 为**非降**的，记为 $A_n \uparrow$, 并称

$$\lim_{n\to\infty} A_n := \bigcup_{n=1}^{+\infty} A_n$$

为它的**极限**.

2. 若

$$\forall n \in N_+, \quad A_n \subseteq A_{n+1}$$

则称 $\{A_n\}$ 为**非增**的，记为 $A_n \downarrow$, 并称

$$\lim_{n\to\infty} A_n := \bigcap_{n=1}^{+\infty} A_n$$

为它的**极限**.

非降或非增的集合序列统称为**单调序列**. 因此, <u>单调集合序列总有极限</u>.

定义 A.4 (上极限与下极限)

对于任意给定的一个集合序列 $\{A_n\}$, 集合序列 $\{B_n = \bigcap_{k=n}^{+\infty} A_k\}$ 与 $\{B'_n = \bigcup_{k=n}^{+\infty} A_k\}$ 分别是非降和非增的, 因而分别有极限:

$$\underline{\lim}_{n\to\infty} A_n = \lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{+\infty} A_k$$

$$\overline{\lim}_{n\to\infty} A_n = \lim_{n\to\infty} B'_n = \bigcap_{n=1}^{\infty} B'_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{+\infty} A_k$$

分别称其为 $\{A_n\}$ 的**上极限**与**下极限**

显然, 记号 $x \in \overline{\lim}_{n\to\infty} A_n$ 意味元素 x 属于序列 $\{A_n\}$ 中的无穷多个集合, 而记号 $x \in \underline{\lim}_{n\to\infty} A_n$, 则表明除去 $\{A_n\}$ 中的有限个集合外, 元素 x 属于该序列的其余集合. 于是我们有

$$\underline{\lim}_{n\to\infty} A_n \subseteq \overline{\lim}_{n\to\infty} A_n$$

定义 A.5

如果 $\underline{\lim}_{n\to\infty} A_n = \overline{\lim}_{n\to\infty} A_n$, 则认为 $\{A_n\}$ 的极限存在, 并且将

$$\lim_{n\to\infty} A_n := \underline{\lim}_{n\to\infty} A_n = \overline{\lim}_{n\to\infty} A_n$$

称为它的**极限**.

A.1.2 集合系

定义 A.6 (可测函数)

设 $y = g(x)$ 是 $\mathbb{R}$ 到 $\mathbb{R}$ 上的一个映照, 若对于一切 $\mathbb{R}$ 中的 Borel 点集 B_1 均有

$$\{x : g(x) \in B_1\} \in \mathcal{B}_1$$

其中 $\mathcal{B}_1$ 为 $\mathbb{R}$ 上 Borel σ 域, 则称 $g(x)$ 是一元博雷尔函数, 也称为一元**可测函数**

附录 B 组合计数

全部组合分析公式的推导基于下列两条原理：

乘法原理 若进行 A_1 过程有 n_l 种方法，进行 A_2 过程有 n_2 种方法，则进行 A_1 过程后再接着进行 A_2 程共有 $n_1 \cdot n_2$ 种方法

加法原理 若进行 A_1 过程有 n_l 种方法，进行 A_2 过程有 n_2 种方法，假定 A_1 过程与 A_2 过程是并行的，则进行过程 A_1 或过程 A_2 的方法共有 $n_1 + n_2$ 种

排列与组合的定义及其计算公式如下：

排列 从 n 个不同元素中任取 $r(r \le n)$ 个元素排成一列 (考虑元素先后出现次序)，称此为一个排列，其总数记为 P_n^r。按乘法原理有：

$$P_n^r = n \times (n-1) \times \cdots \times (n-r+1) = \frac{n!}{(n-r!)}$$

若 $r = n$，则称为全排列，记为 ${}_n$。显然，全排列 $P_n = n!$。

重复排列 从 n 个不同元素中每次取出一个，<u>放回后</u> 再取下一个，如此连续取 r 次所得的排列称为重复排列。此种重复排列数共有 n^r 个。

组合 从 n 个不同元素中任取 $r(r \le n)$ 个元素并成一组 (不考虑元素间的先后次序)，称此为一个组合，其总数记为 $\binom{n}{r}$ 或 C_n^r。按乘法原理有：

$$\binom{n}{r} = \frac{P_n^r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

在此规定 $0! = 1$ 与 $\binom{n}{0} = 1$.

重复组合 从 n 个不同元素中每次取出一个，<u>放回后</u> 再取下一个，如此连续取 r 次所得的组合称为重复组合. 其总数为 $\binom{n+r-1}{r}$。注意这里的 r 也允许大于 n.

重复组合数的得出可如下考虑：将 n 个元素看作 n 个盒子，使用 $n+1$ 个插板区分（相邻两插板组成一个盒子），取 r 次视为往盒子中放入 r 个球。每种取法可视为将 $n-1$ 个插板（两头的不能动）与 r 个球进行放置。共有 $r+n-1$ 个位置，但只要其中的插板（或球）放置好后，球（或插板）的位置自然固定。所以其总数为 $\binom{n+r-1}{n-1}$（或 $\binom{n+r-1}{r}$，两者相同）。

	有序	**无序**
放回	重复排列 n^r	重复组合 $\binom{n+r-1}{r}$
不放回	排列 $\frac{n!}{(n-r!)}$	组合 $\binom{n}{r}$

定理 B.1 (牛顿二项式定理)

若对于任意实数 α 定义

$$\binom{\alpha}{r} = \frac{A_r^\alpha}{r!} = \frac{\alpha(\alpha-1)\cdots(\alpha-r+1)}{r!}$$

则有牛顿二项式：

$$(1+x)^\alpha = \sum_{r=0}^{\infty}\binom{\alpha}{r}x^r$$

证明