



概率论笔记

作者：肖程哲

时间：August 19, 2022



苟日新，日日新，又日新

目录

第 1 章 概率基础	1	3.3.1 delta 法	33
1.1 概率空间	1	3.3.2 预测	33
1.1.1 随机事件	1	3.4 熵与信息	33
1.1.2 概率空间	2	3.5 其他特征	33
1.1.3 概率的性质	3		
1.2 古典概型与几何概型	6	第 4 章 常见分布	34
1.2.1 古典概型	6	4.1 离散分布	34
1.2.2 几何概型	6	4.1.1 均匀分布	34
1.2.3 Bertrand 奇论	8	4.1.2 伯努利分布	34
1.3 条件概率与独立	9	4.1.3 二项分布	35
1.3.1 条件概率	9	4.1.4 几何分布	36
1.3.2 独立性	12	4.1.5 负二项分布	37
第 1 章 练习	14	4.1.6 多项分布	38
		4.1.7 泊松分布	39
第 2 章 随机变量	16	4.1.8 超几何分布	41
2.1 随机变量的分布	16	4.2 连续分布	42
2.2 多元随机变量	18	4.2.1 均匀分布	42
2.2.1 边际分布	19	4.2.2 指数分布	42
2.2.2 条件分布	19	4.2.3 伽马分布	44
2.2.3 独立	20	4.2.4 贝塔分布	45
2.3 随机变量的函数	21	4.3 正态分布及其导出分布	47
2.3.1 分布函数法	21	4.3.1 正态分布	47
2.3.2 Copula	22	4.3.2 卡方分布	48
2.3.3 概率密度函数法	23	4.3.3 F 分布	49
2.3.4 矩母函数法	24	4.3.4 t 分布	50
2.3.5 次序统计量	24	4.3.5 柯西分布	51
第 3 章 随机变量的数值特征	26	4.4 各分布间关系	53
3.1 期望	26		
3.1.1 均值	26	第 5 章 概率极限	55
3.1.2 方差	27	5.1 收敛	55
3.1.3 协方差	28	5.2 大数定理	57
3.1.4 条件期望	29	5.3 中央极限定理	58
3.2 矩母函数与特征函数	30		
3.2.1 矩	30	第 A 章 测度论基础	60
3.2.2 矩母函数	31	A.1 可测空间和可测映射	60
3.2.3 联合特征函数	32	A.1.1 集合及其运算	60
3.2.4 特征函数	32	A.1.2 集合系	61
3.3 估计与预测	33		
		附录 B 组合计数	62

第1章 概率基础

考试重点

□ 概率空间的定义

□ 古典概型与几何概型

□ 条件概率与独立

□ 伯努利试验

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 $\Omega = \{\omega\}$ 。若样本点有至多可数个, 则称为离散样本空间, 若不可数, 则称为连续样本空间。



定义 1.2 (事件的古典定义)

样本点 ω 的集合称为 (随机) 事件 (event)。



随机现象被抽象为集合。事件间的包含、相等、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: $AB\bar{C}$.
- 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: $AB\bar{C} \cup A\bar{B}C \cup \bar{A}BC$.
- 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: $\bar{A} \cap \bar{B} \cap \bar{C}$.
- 事件 “ A 、 B 、 C 不全发生” 可表示为: $\bar{A} \cup \bar{B} \cup \bar{C}$.

性质 集合的运算性质:

- 交换律

$$A \cup B = B \cup A, \quad AB = BA \quad (1.1)$$

- 结合律

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

- 分配律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (1.3)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \quad (1.4)$$

- 对偶律 (De Morgan 公式)

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad (1.5)$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}. \quad (1.6)$$

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 ($A - B = A\bar{B}$) .

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 \mathcal{F} ，是 Ω 的子集族 (collection of subsets)，应满足 σ 代数 的要求：

- $\emptyset \in \mathcal{F}$ ，代表无事发生；
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ，即 \mathcal{F} 对补集运算 (逻辑上的非) 封闭；
- $A_1, \dots, A_n, \dots \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ ，即 \mathcal{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 \mathcal{G} ，必存在 Ω 上唯一的 σ 域 $m(\mathcal{G})$ ，满足下列性质：

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

这个 $m(\mathcal{G})$ 称为包含 \mathcal{G} 的最小 σ 域，或由 \mathcal{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 \mathcal{G} ，所以 m 存在。再取 Σ 上满足此条件的 σ 域之交作为 $m(\mathcal{G})$ 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 $\mathcal{B}_{\mathbb{R}}$ 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 Borel σ 域，记为 \mathcal{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 \mathcal{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， \mathcal{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 ：
 $\mathcal{F} \rightarrow \mathbb{R}$ 称为概率的条件是：

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

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 $A_1, \dots, A_n, \dots \in \mathcal{F}$ 两两不交, 即 $A_i \cap A_j = \emptyset, \forall i \neq j$, 则

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

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



1.1.3 概率的性质

性质 概率的性质:

- $P(\emptyset) = 0$;
- $P(A^c) = 1 - P(A)$;
- 若 $A \subseteq B$ 则 $P(A) \leq P(B)$;

例题 1.2 口袋中有编号为 $1, 2, \dots, n$ 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 $1, 2, \dots, i$ 号球中取球即可。由题可知

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

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

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

例题 1.3

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

解 设 $P(A) \geq P(B)$, 则:

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

另一方面有:

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

命题 1.2 (加法公式)

基础形式:

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

一般形式:

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

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, \dots, A_n 有

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



例题 1.4 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

$$\begin{aligned} P(A_1) &= P(A_2) = \cdots = P(A_n) = \frac{1}{n} \\ P(A_1 A_2) &= P(A_1 A_3) = \cdots = P(A_{n-1} A_n) = \frac{1}{n(n-1)} \\ P(A_1 A_2 A_3) &= P(A_1 A_2 A_4) = \cdots = P(A_{n-2} A_{n-1} A_n) = \frac{1}{n(n-1)(n-2)} \\ &\vdots \\ P(A_1 A_2 \cdots A_n) &= \frac{1}{n!} \end{aligned}$$

再由概率的加法公式 1.2 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 $A_i \in \mathcal{F}, i = 1, 2, \dots$ 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

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

即若希望由有限可加性推出可列可加性, 则需要下式成立:

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

可写为:

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

若记

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

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

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

定义 1.6 (下连续)

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

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

则称它是下连续的.



定理 1.1

若 P 是 \mathcal{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 \mathcal{F} 中一个单调不减的事件序列, 即

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

若定义 $S_0 = \emptyset$, 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

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

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

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

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

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

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

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

由于 $\{B_n\}$ 单调不增，所以

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

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)

若事件 A 含有 k 个样本点，则事件 A 的概率为

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



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。以后我们经常研究摸球模型，意义即在于此。

例题 1.5 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 ($n \leq N$) 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

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



例题 1.6 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

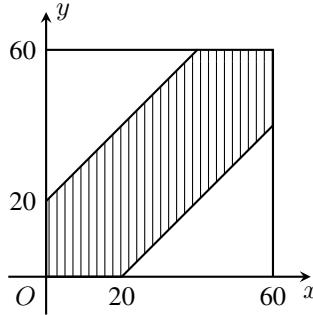


图 1.1：会面问题中的 Ω 与 A 。

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 1.7 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

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

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

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

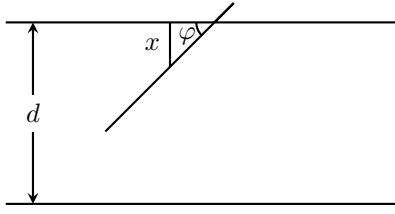
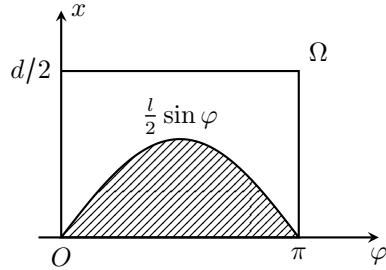


图 1.2: 蒲丰投针问题

图 1.3: 蒲丰投针问题中的 Ω 和 A

例题 1.8 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

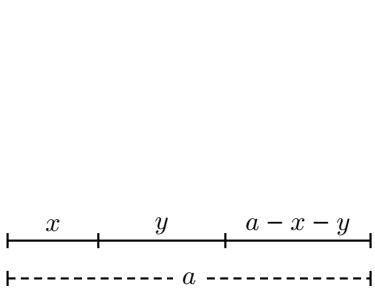
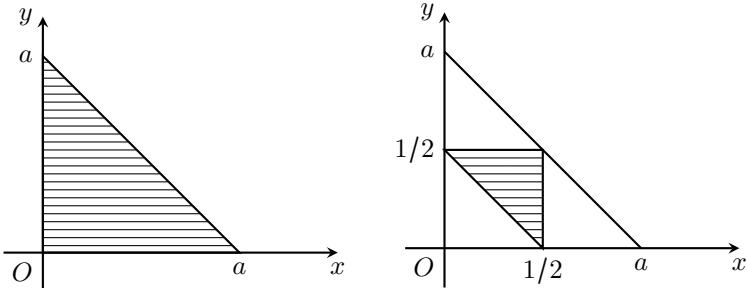
图 1.4: 长度为 a 的线段分成三段.

图 1.6: 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 $A, B \in \mathcal{F}$, 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 1.3

条件概率仍是一个概率测度, 即若设 $P(B) > 0$, 则

- $P(A|B) \geq 0$, $A \in \mathcal{F}$,
- $P(\Omega|B) = 1$,
- 若 \mathcal{F} 中的 $A_1, A_2, \dots, A_n, \dots$ 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 $A_1, A_2, \dots, A_n, \dots$ 互不相容, 所以 $A_1B, A_2B, \dots, A_nB, \dots$ 也

互不相容，故

$$\begin{aligned} P\left(\bigcup_{n=1}^{+\infty} A_n | B\right) &= \frac{P\left(\left(\bigcup_{n=1}^{+\infty} A_n\right)B\right)}{P(B)} = \frac{P\left(\bigcup_{n=1}^{+\infty}(A_n B)\right)}{P(B)} \\ &= \sum_{n=1}^{+\infty} \frac{P(A_n B)}{P(B)} = \sum_{n=1}^{+\infty} P(A_n | B). \end{aligned}$$

定理 1.3 (乘法法则)

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

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

泛化后有：

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



例题 1.9 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

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

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：

- 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

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

- 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

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

- 当 $c > 0, d = 0$ 时，称为**传染病模型**。此时，每次取出球后会增加下一次取到同色球的概率，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

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

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。

- 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 (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

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

定理 1.4 (全概率公式)

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

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



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 $AB_1, AB_2, \dots, AB_n, \dots$ 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 $P(AB_i) = P(B_i)P(A|B_i)$, 代入上式即得全概率公式。

推论 1.2

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



例题 1.10 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 .

敏感性问题的调查方案，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下，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 $P(\text{是})$, 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 $B_1, B_2, \dots, B_n, \dots$ 互不相容, 且包含事件 A 。即 $A \subseteq \bigcup_{i=1}^{\infty} B_i$, $B_i B_j = \emptyset, \forall i \neq j$, 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 1.11 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

$$\begin{aligned} P(B) &= 0.0004, & P(\bar{B}) &= 0.996, \\ P(A|B) &= 0.99, & P(A|\bar{B}) &= 0.001. \end{aligned}$$

由贝叶斯公式得

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

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 $A, B \in \mathcal{F}$ 满足

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

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

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

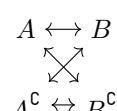
$$P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$$

则称此事件集相互独立 (mutually independent)



笔记 当 $P(A) > 0$ 时, $P(B|A) = P(B) \iff B \perp\!\!\!\perp A$, 由此可得到 B 独立于 A 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.



证明 由概率的性质知

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

又由 A 与 B 的独立性知

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

所以

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

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

例题 1.12 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

$$A_1 \cup \bar{A}_1 \bar{A}_2 A_3 \cup \bar{A}_1 \bar{A}_2 \bar{A}_3 \bar{A}_4 A_5 \cup \dots$$

又因为各次射击是独立的, 所以得

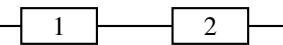
$$\begin{aligned} P(\text{甲获胜}) &= \alpha + (1 - \alpha)(1 - \beta)\alpha + (1 - \alpha)^2(1 - \beta)^2\alpha + \dots \\ &= \alpha \sum_{i=0}^{+\infty} (1 - \alpha)^i (1 - \beta)^i \\ &= \frac{\alpha}{1 - (1 - \alpha)(1 - \beta)}. \end{aligned}$$

同理可得

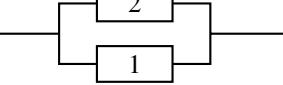
$$\begin{aligned} P(\text{乙获胜}) &= P(\bar{A}_1 A_2 \cup \bar{A}_1 \bar{A}_2 \bar{A}_3 A_4 \cup \dots) \\ &= (1 - \alpha)\beta + (1 - \alpha)(1 - \beta)(1 - \alpha)\beta + \dots \\ &= \beta(1 - \alpha) \sum_{i=0}^{+\infty} (1 - \alpha)^i (1 - \beta)^i \\ &= \frac{\beta(1 - \alpha)}{1 - (1 - \alpha)(1 - \beta)}. \end{aligned}$$

例题 1.13 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.

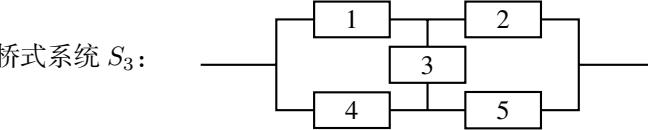
1. 串联系统 S_1 :



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 $S_1 = A_1 A_2$, 所以

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

(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

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

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

$$\begin{aligned} P(S_3|A_3) &= P((A_1 \cup A_4)(A_2 \cup A_5)) = P(A_1 \cup A_4)P(A_2 \cup A_5) \\ &= [1 - (1-p)^2]^2 = 0.9801. \end{aligned}$$

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）。

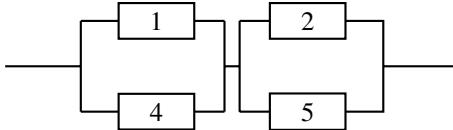


图 1.7: 先并后串系统

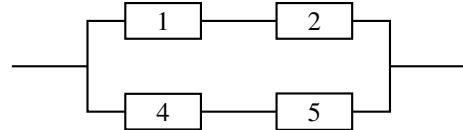


图 1.8: 先串后并系统

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

$$\begin{aligned} P(S_3) &= p[1 - (1-p)^2]^2 + (1-p)[1 - (1-p^2)^2] \\ &= 0.9 \times 0.9801 + 0.1 \times 0.9639 = 0.9785. \end{aligned}$$

定义 1.12 (事件域的独立性)

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

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

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



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

类似地可以定义 n 个试验 E_1, E_2, \dots, 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, \dots, E_n 相互独立。如果这 n 个独立试验还是相同的，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} ，则称这种试验为 n 重伯努利试验。

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 $x^2 + Bx + C = 0$ ，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 $1, 2, \dots, n$ 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 ($m \leq n+1$) 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 $0, 1, 2, \dots, 9$ 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 $1, 2, \dots, 9$ 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: $|P(AB) - P(A)P(B)| \leq \frac{1}{4}$
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.

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 $\omega \in \Omega$, 将其记为 $\{X \in B\} = X^{-1}(B)$. 由于 \mathbb{P} 定义在 \mathcal{F} 上, 故需 $X^{-1}(B) \in \mathcal{F}$.

定义 2.1 (可测性)

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



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

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

为随机元 (random element), 也称随机变量 (random variable). 其中 $\mathcal{F}_{\mathcal{X}}$



由于只考虑 $\mathcal{F}_{\mathcal{X}}$ 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. 新样本空间由 Borel 点集 构成, 对应的概率测度等于原像的.

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

定义 2.3 (离散与连续随机变量)

当 \mathcal{X} 是(至多可数的)离散点集, $\mathcal{F}_{\mathcal{X}}$ 由 \mathcal{X} 的所有子集组成, 则称其为离散随机变量 (discrete random variable).

当随机变量 $\mathcal{X} = \mathbb{R}^n$, 考虑 $\mathcal{F}_{\mathcal{X}}$ 为 $\{\prod_{i=1}^n (-\infty, x_i] : x_1, \dots, x_n \in \mathbb{R}\}$ 生成的 Borel 代数 (最小的 σ 代数), 则称其为连续随机变量 (continuous random variable).



2.1 随机变量的分布

定义 2.4

称随机元 X 诱导的概率测度

$$\mathbb{P}\{X \in \bullet\}, \bullet \in \mathcal{F}_{\mathcal{X}}$$

为 X 的概率分布 (distribution/law)



注 对于随机变量, 他的取值是随机的, 但他的分布是固定的

定义 2.5 (单变量分布函数)

一个函数 $F : \mathbb{R} \rightarrow [0, 1]$ 称为一个单变量分布函数, 当其满足以下性质时:

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

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

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



性质 $F(x)$ 最多只有可数个间断点

命题 2.1

对每个分布 $Q : \mathcal{B}_1 \rightarrow [0, 1]$ 都存在唯一一个分布函数 $F_Q : \mathbb{R} \rightarrow [0, 1]$ 使得 $F_Q(x) = Q[(-\infty, x)], \quad \forall x \in \mathbb{R}$ 成立。

**命题 2.2**

对每个分布函数 $F : \mathbb{R} \rightarrow [0, 1]$ 都存在唯一一个分布 $Q_F : \mathcal{B}_1 \rightarrow [0, 1]$ 使得 $Q_F[(-\infty, x)] = F(x), \quad \forall x \in \mathbb{R}$ 成立。

**定理 2.1**

分布函数可以唯一决定概率分布, 即:

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

把随机变量 X 服从分布函数 $F(x)$ 简记作 $X \sim F(x)$



	离散	连续
一元随机变量	概率质量函数 (pmf)	概率密度函数 (pdf)
	累积分布函数 (cdf)	
	矩母函数/特征函数 (mgf/chf)	
多元随机变量	联合概率质量函数 (joint pmf)	联合概率密度函数 (joint pdf)
	联合累积分布函数 (joint cdf)	
	联合矩母函数/特征函数 (joint mgf/chf)	

定义 2.6 (累积分布函数)

此时 $X = (X_1, \dots, X_n)^\top$ 的分布由(累积) 分布函数 (cumulative distribution function, c.d.f.)

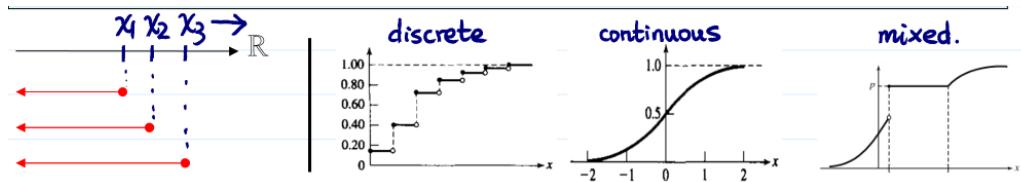
$$F_X(x_1, \dots, x_n) = \mathbb{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

唯一刻画. 把随机变量 X 服从分布函数 $F(x)$ 简记作 $X \sim F(x)$

**定义 2.7 (概率质量函数)**

当且仅当函数 $p(x)$ 满足下述条件时, 被称为概率质量函数 (probability mass function, p.m.f.):

- $p(x) \geq 0$



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



当 X 是离散型随机变量, 设 \mathcal{F}_X 由 \mathcal{X} 的所有子集组成, 此时 X 的分布由

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

唯一刻画. 其与分布函数间的关系为:

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

定义 2.8 (概率密度函数)

当且仅当函数 $f(x)$ 满足下述条件时, 被称为概率密度函数 (probability density function, p.d.f.):

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$



当 X 是连续型随机变量, 且 $F_X : \mathbb{R}^n \rightarrow [0, 1]$ 可微 (或者更一般地, 绝对连续), 此时 X 的分布由

$$f_X := \frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$$

唯一刻画. 其与分布函数间的关系为:

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

注 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. 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.

2.2 多元随机变量

定义 2.9 (随机向量)

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

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

构成一个 n 维随机向量, 亦称 n 维随机变量.



命题 2.3

若 B_n 为 \mathbb{R}^n 上任一博雷尔点集, 有

$$\{X(\omega) \in B_n\} \in \mathcal{F}$$



定义 2.10

称 n 元函数

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

为随机向量 $X(\omega)$ 的联合分布函数 (joint cdf).



当 $n = 2$ 时, 有

$$\mathbb{P}((a_1, b_1) \leq X < (a_2, b_2)) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \quad (2.1)$$

性质 多元分布函数的一些性质:

1. 单调性: 关于每个变元是单调不减函数;
- 2.

$$F(x_1, x_2, \dots, -\infty, \dots, X_n) = 0$$

$$F(+\infty, +\infty, \dots, +\infty) = 1$$

3. 关于每个变元右连续.

4. 在二元场合, 还应该有: 对任意 $a_1 < b_1, a_2 < b_2$, 都有

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geq 0$$

为保证2.1式中的概率的非负性, 性质 4 是必须的, 而且由性质 4 可以推出单调性, 但存在着反例说明, 由单调性并不能保证性质 4 的成立(见习题 12). 这是多元场合与一元场合的不同之处.

2.2.1 边际分布

定义 2.11

对于多维随机变量 X , 只考虑其中一个分量的分布时, 称其为 X 的边缘分布或边缘分布. 对于分量 X_i , 其边缘分布函数 (marginal cdf) 为:

$$F_{X_i}(x_i) = \mathbb{P}\{X_i \leq x_i\} = F(\infty, \dots, x_i, \dots, \infty)$$



2.2.2 条件分布

定义 2.12

对一切使 $P(Y = y_j) = p_{\cdot j} = \sum_{i=1}^{+\infty} p_{ij} > 0$ 的 y_j , 称

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

为给定 $Y = y_j$ 条件下 X 的条件分布列. 若 $p_X(x) = 0$, 则定义其为 0.



设二维连续随机变量 (X, Y) 的联合密度函数为 $p(x, y)$, 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 $P(X \leq x | Y = y)$. 但是, 因为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 $P(X \leq x | Y = y)$, 一个很自然的想法是: 将 $P(X \leq x | Y = y)$ 看

成是 $h \rightarrow 0$ 时 $P(X \leq x | y \leq Y \leq y + h)$ 的极限, 即

$$\begin{aligned} P(X \leq x | Y = y) &= \lim_{h \rightarrow 0} P(X \leq x | y \leq Y \leq y + h) \\ &= \lim_{h \rightarrow 0} \frac{P(X \leq x, y \leq Y \leq y + h)}{P(y \leq Y \leq y + h)} \\ &= \lim_{h \rightarrow 0} \frac{\int_{-\infty}^x \int_y^{y+h} p(u, v) dv du}{\int_y^{y+h} p_Y(v) dv} \\ &= \lim_{h \rightarrow 0} \frac{\int_x^{\infty} \left\{ \frac{1}{h} \int_y^{y+h} p(u, v) dv \right\} du}{\frac{1}{h} \int_y^{y+h} p_Y(v) dv}. \end{aligned}$$

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

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1}{h} \int_y^{y+h} p_Y(v) dv &= p_Y(y), \\ \lim_{h \rightarrow 0} \frac{1}{h} \int_y^{y+h} p(u, v) dv &= p(u, y). \end{aligned}$$

所以

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

至此, 我们可以定义连续随机变量的条件分布如下.

定义 2.13

对一切使 $p_Y(y) > 0$ 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

$$\begin{aligned} F(x|y) &= \int_{-\infty}^x \frac{p(u, y)}{p_Y(y)} du, \\ p(x|y) &= \frac{p(x, y)}{p_Y(y)}. \end{aligned}$$

同理对一切使 $p_Y(y) > 0$ 的 x , 给定 $X = x$ 条件下 Y 的条件分布函数和条件密度函数分别为

$$\begin{aligned} F(y|x) &= \int_{-\infty}^y \frac{p(x, v)}{p_X(x)} dv \\ p(y|x) &= \frac{p(x, y)}{p_X(x)}. \end{aligned}$$



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 的概率密度函数与概率三定理的对应:

$$\text{乘法法则} \quad p_{XY}(x, y) = p_{Y|X}(y|x)p_X(x), \quad f_{XY}(x, y) = f_{Y|X}(y|x)f_X(x)$$

$$\text{全概率公式} \quad p_Y(y) = \sum_x p_{Y|X}(y|x)p_X(x), \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx$$

$$\text{Bayes 原理} \quad p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)p_X(x)}{\sum_x p_{Y|X}(y|x)p_X(x)}, \quad f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x)f_X(x)dx}$$

2.2.3 独立

定义 2.14 (独立随机变量)

若随机变量 $X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ 联合分布函数可分解成各分量边缘分布函数的乘积, 即:

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

则称随机变量 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:

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

对于离散情况:

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



2.3 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.

定义 2.15 (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

**定理 2.3 (事件法)**

设 $Y = g(X)$ 是随机向量 $X = (X_1, X_2, \dots, X_n)$ 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 2.1 已知随机变量 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 0

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 $\{X\}_n$ 各分类独立, 则其变换 $\{Y\}_n = g(\{X\}_n)$ 各分类也独立.



证明

2.3.1 分布函数法

通过下式获取随机变量的函数的分布:

$$F_Y(y) = \begin{cases} \int_{A_y} f_X(x)dx \\ \sum_{x \in A_y} p_X(x) \end{cases}, \quad A_y = \{x | g(x) \leq y\}$$

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 2.2 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

$$F_Y(y) = P(Y \leq y) = P(-\sqrt{y} < X \leq \sqrt{y}) + P(X = \sqrt{y} = F_X(\sqrt{y}) - F_X(-\sqrt{y})) + 0 \quad y \geq 0$$

$$\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(\sqrt{y}) \\
&= f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \frac{-1}{2\sqrt{y}} \\
&= \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y}))
\end{aligned}$$

例题 2.3 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

$$F_Y(y) = P(Y \leq y) = P(-\sqrt{y} < X \leq \sqrt{y}) + P(X = \sqrt{y} = F_X(\sqrt{y}) - F_X(-\sqrt{y})) + 0 \quad y \geq 0$$

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

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, 与 2.1 类似

例题 2.4 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

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

$$\begin{aligned}
F_Z(z) &= \iint_{Q_z} f(x, y) dx dy = \int_{-\infty}^0 \int_{xz}^{\infty} + \int_0^{\infty} \int_{-\infty}^{xz} f(x, y) dy dx \\
&= \int_{-\infty}^0 \int_z^{-\infty} + \int_0^{\infty} \int_{-\infty}^z xf(x, xv) dv dx \quad (\text{set } y = xv) \\
&= \int_{-\infty}^0 \int_{-\infty}^z (-x)f(x, xv) dv dx + \int_0^{\infty} \int_{-\infty}^z xf(x, xv) dv dx \\
&= \int_{-\infty}^z \int_{-\infty}^{\infty} |x| f(x, xv) dx dv
\end{aligned}$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz) dx$$

2.3.2 Copula

定义 2.16

设连续型实值随机变量 X 有分布函数 F , 易见 F 在 $\bar{\mathbb{R}} = [-\infty, +\infty]$ 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 有分布函数 F , 则 $F(X) \sim \text{Uniform}([0, 1])$

**证明**

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

定理 2.6

设连续型实值随机变量 X 有分布函数 F , 且设 $U \sim \text{Uniform}([0, 1])$, 则 $F^{-1}(U) \stackrel{d}{=} X$, 其中 $\stackrel{d}{=}$ 表示分布相同 (equal in distribution).

**定理 2.7 (Sklar 定理)**

考虑多个连续型实值随机变量 X_1, \dots, X_k , 记 X_i 的分布函数为 F_i . 我们称 $(F_1(X_1), \dots, F_k(X_k))$ 的分布函数 $C : [0, 1]^k \rightarrow [0, 1]$ 为相应的 **Copula**, 适合

$$C(F_1(x_1), \dots, F_k(x_k)) = \mathbb{P}\{X_1 \leq x_1, \dots, X_k \leq x_k\}, \quad \forall x_1, \dots, x_k$$

这个结果称为 **Sklar 定理**



这个结果, 在金融统计中有颇多应用. 稍作诠释的话, Copula 提取了变量间的相关性, 通过粘合边际能够恰好地表示总体. 人们可以构造各种各样的 Copula, 对真实世界进行建模.

2.3.3 概率密度函数法

定理 2.8

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

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

否则为 0

若 g 为分段单调函数, 则分段计算上是结果, 再进行相加

**定理 2.9**

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:

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

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, 再通过积分获取 \mathbf{Y} 的概率密度函数.

例题 2.5 已知随机变量 X_1, X_2 的联合概率密度函数 $f(x_1, x_2)$, 求 $Y_1 = \frac{X_2}{X_1}$ 的分布

解 令 $Y_2 = X_1$, 则:

$$x_1 = y_2 \equiv w_1(y_1, y_2)$$

$$x_2 = y_1 y_2 \equiv w_2(y_1, y_2)$$

$$\frac{\partial w_1}{\partial y_1} = 0, \frac{\partial w_1}{\partial y_2} = 1, \frac{\partial w_2}{\partial y_1} = y_2, \frac{\partial w_2}{\partial y_2} = y_1$$

$$J = \begin{vmatrix} 0 & 1 \\ y_2 & y_1 \end{vmatrix} = -y_2$$

所以:

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(y_2, y_1 y_2) |y_2|$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1 Y_2}(y_1, y_2) dy_2 = \int_{-\infty}^{\infty} f_{X_1 X_2}(y_2, y_1 y_2) |y_2| dy_2 \quad (2.2)$$

命题 2.4

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



证明

2.3.4 矩母函数法

2.3.5 次序统计量

定义 2.17

设 X_1, X_2, \dots, X_n 为随机变量, 将其按大小排序后记为 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, 则将 $x_{(i)}$ 称为该样本的第 i 个次序统计量. 其中

- 最小次序统计量定义为: $X_{(1)} = \min\{x_1, \dots, x_n\}$
- 最大次序统计量定义为: $X_{(n)} = \max\{x_1, \dots, x_n\}$
- 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, \dots, X_n 独立, 但其次序统计量一般不独立

例题 2.6 若 X_1, X_2, \dots, X_n 独立同分布, 求 $X_{(1)}$ 与 $X_{(n)}$ 的分布
解

$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) \\ &= [F(x)]^n \\ f_{X_{(n)}}(x) &= \frac{d}{dx} F_{X_{(n)}}(x) \\ &= n f(x) [F(x)]^{n-1} \end{aligned}$$

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

定理 2.10

第 i 个次序统计量的概率密度函数为:

$$f_{X_{(k)}} = C(n; 1, k-1, n-k) f(x) [F(x)]^{k-1} [1 - F(x)]^{n-k}$$

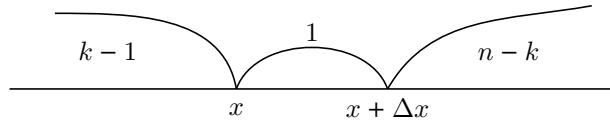


图 2.1: $X_{(k)}$ 取值的示意图

定理 2.11

次序统计量 $(x_{(i)}, x_{(j)})$ ($i < j$) 的联合分布密度函数为

$$\begin{aligned}
p_{ij}(y, z) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(y)]^{i-1} [F(z) - F(y)]^{j-i-1} \\
&\quad \cdot [1 - F(z)]^{n-j} p(y)p(z), \quad y \leq z,
\end{aligned}$$



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 2.2).

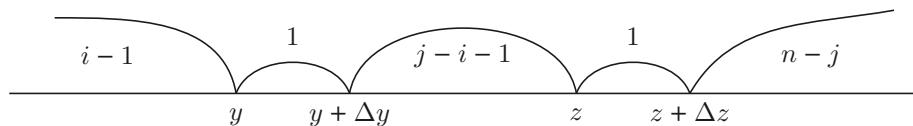


图 2.2: $x_{(i)}$ 与 $x_{(j)}$ 取值的示意图

例题 2.7 若 X_1, X_2, \dots, X_n 独立同分布, 求 $R = X_{(n)} - X_{(1)}$ 的分布

解

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

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

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

3.1 期望

定义 3.1

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

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x) dF_X(x)$$

为 $g(X)$ 的期望 (expectation).



注 当 $F_X(x)$ 在 x_0 处连续可导时, $dF_X(x_0) = f_X(x_0)dx$; 当 x_0 为区间断点时时, $dF_X(x_0) = p_X(x_0)\delta(x_0)dx$. 期望算子 \mathbb{E} 是一个线性泛函, 仅适用于可积的随机变量.

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 $\text{Cov}(X, Y)$ 或 σ_{XY} .
4. 定义 X 与 Y 的相关系数 (correlation coefficient) 为 $\sigma_{XY}/(\sigma_X\sigma_Y)$, 记为 $\text{Cor}(X, Y)$ 或 ρ_{XY} . 若 $\rho_{XY} = 0$, 则称 X 与 Y 不相关



3.1.1 均值

注

- 随机变量的均值可看作其加权平均, 权重为其 pdf 或 pmf, 也即其质心. 从大数定律 (5) 的角度看, 也可解释为其长期均值.
- 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相同, 方差的为其平方

定理 3.1

均值为随机变量的线性映射, 即:

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

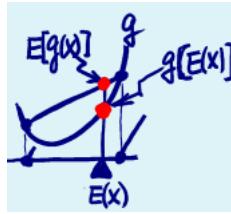


定理 3.2 (0)

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 $\mathbb{E}(X/Y) \neq \mathbb{E}(X)/\mathbb{E}(Y)$

定理 3.3

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



一个重要结果是, 若 $g(X) \geq 0$, 则 $\mathbb{E}[g(X)] = 0 \implies g(X) \stackrel{\text{a.s.}}{=} 0$, 即 $\mathbb{P}\{g(X) = 0\} = 1$. 其证明可通过 **Markov 不等式**

$$\mathbb{P}\{g(X) \geq \varepsilon\} \leq \mathbb{E}[g(X)]/\varepsilon, \quad \forall \varepsilon > 0$$

完成, 其中需要用到概率的连续性, 即 $\lim_{n \rightarrow \infty} A_n = A \implies \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

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) $X \mapsto \frac{X - \mathbb{E}X}{\sqrt{\text{Var}(X)}}$.

3.1.2 方差

定理 3.4

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



定理 3.5

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



定理 3.6

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

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



定理 3.7

若 X_1, \dots, X_n 相互独立, 则:

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



考虑均方误差 (mean squared error)

$$\text{MSE}(X; \theta) = \mathbb{E}[|X - \theta|^2], \quad \theta \in \mathbb{R},$$

通过方差偏差分解 (variance-bias decomposition)

$$\text{MSE}(X; \theta) = \text{Var}(X) + |\mathbb{E}X - \theta|^2$$

可以说明 $\theta \mapsto \text{MSE}(X; \theta)$ 在 $\mathbb{E}X$ 处取到最小值 $\text{Var}(X)$.

投影 (projection) 和正交分解的思想在各种内积空间中应用广泛, 这里是 $\mathbb{E} = \text{proj}_{\mathbb{R}}$, 概率论中关于子事件域 \mathcal{G} (随机元 X , resp.) 的条件期望几何直观是 $\text{proj}_{\mathcal{G}}(\text{proj}_{\sigma(X)}, \text{resp.})$, 线性模型 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 中拟合值为 $\hat{\mathbf{y}} = \text{proj}_{\text{Col}(\mathbf{X})}\mathbf{y}$.

定理 3.8 (Chebyshev 不等式)

设随机变量 X 的均值与方差分别为: μ, σ^2 , 则:

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



证明 设 $f(x)$ 为 X 的概率密度函数, 令 $R = \{x : |x - \mu| > t\}$

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

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

推论 3.1

$$\text{Var}(X) = 0 \implies P(X = \mu) = 1$$



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.

定理 3.9

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



定理 3.10

$$\text{Cov}\left(a + \sum_{i=1}^n b_i X_i, c + \sum_{j=1}^m d_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j) = \mathbf{b}^T \Sigma \mathbf{d}$$



定理 3.11

独立是不相关的充分条件, 但不是必要条件



定理 3.12

$-1 \leq \rho_{XY} \leq 1$, 当且仅当 X 与 Y 间为线性关系时取等号



证明

定理 3.13

平移与缩放随机变量都不影响其协方差, 即:

$$|\text{Cov}(a + bX, c + dY)| = |\text{Cor}(X, Y)|$$



3.1.4 条件期望

定义 3.3

若 $h(Y)$ 在给定 $X = x$ 下的条件分布 (定义2.12) 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。

定理 3.14

若随机变量 X, Y 独立, 则:

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



证明 由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。

笔记 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 3.15 (重期望公式)

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



证明

定理 3.16

联合期望公式:

$$E_{X,Y} = E_X E_{Y|X} = E_Y E_{X|Y}$$

泛化情况:

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



定理 3.17 (重方差公式)

随机变量 Y 的方差可作如下分解:

$$\text{Var}_Y(Y) = \text{Var}_X[\mathbb{E}_{Y|X}(Y|X)] + E_X[\text{Var}_{Y|X}(Y|X)]$$

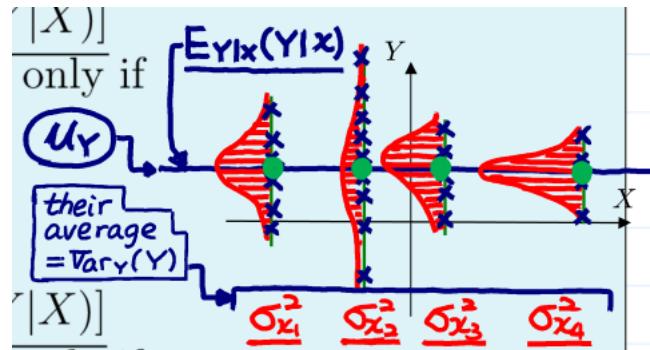
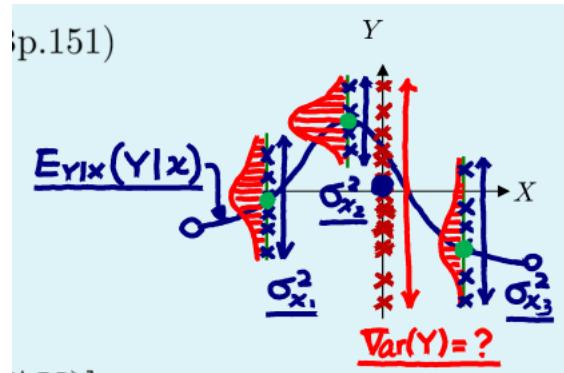


推论 3.2

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

当且仅当 $E_{Y|X}(Y|X) = E_Y(Y)$ 时取等号





推论 3.3

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

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

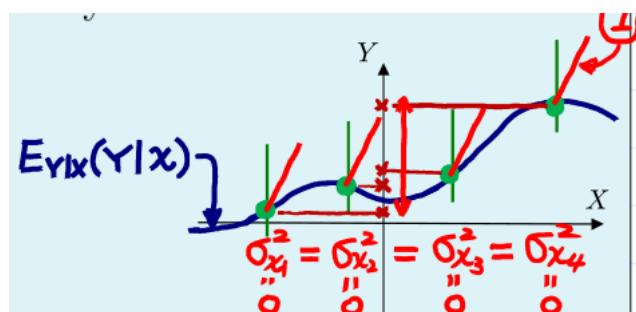


3.2 矩母函数与特征函数

3.2.1 矩

定义 3.4

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 v_k ;



易知矩与中心矩间存在以下关系:

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

特别的有:

$$\begin{aligned} E(X) &= \mu_1 \\ \text{Var}(X) &= v_2 = \mu_2 - \mu_1^2 = \mu_2 - \mu_1^2 \end{aligned}$$

3.2.2 矩母函数

定义 3.5

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 3.18

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 3.19

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 3.1

若 a, b 为常数, 则

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



定理 3.20

若 X, Y 独立, 则

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

泛化情况: 若 X_1, \dots, X_n 相互独立, 则

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



3.2.3 联合特征函数

定义 3.6

对于随机变量 X_1, \dots, X_n , 若下式期望存在:

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

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 3.2

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



定理 3.21

当且仅当:

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

时, X_1, \dots, 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 3.22

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



3.2.4 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 3.7

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

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

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

$$\phi_{X_1 \dots X_n}(t_1, \dots, t_n) = \mathbb{E}(e^{i(t_1 X_1 + \dots + t_n X_n)})$$



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 3.3

随机变量的特征函数总是存在



命题 3.4

若矩母函数存在, 则其与特征函数之间满足关系:

$$\phi_X(t) = M_X(it)$$



定理 3.23

特征函数可通过以下逆变换得到分布：

离散 $p_X(x) = \lim_{T \rightarrow \infty} \int_{-T}^T e^{-itx} \phi_X(t) dt$

连续 $f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt$



3.3 估计与预测

3.3.1 delta 法

3.3.2 预测

3.4 熵与信息

3.5 其他特征

第4章 常见分布

考试重点

- 各分布的特征
- 多维正态分布
- 各分布联系与转换

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, \dots, 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 $X \sim U(a_1, \dots, a_n)$ 。



离散均匀分布的特征：

参数 $a_i \in \mathbb{R}$

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

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

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

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

实例 丢一个均匀的骰子

4.1.2 伯努利分布

定义 4.2 (Bernoulli 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**Bernoulli 分布** (Bernoulli distribution) (也称 0-1 分布)，记为 $X \sim B(p)$ 。



Bernoulli 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 $\sigma^2 = p(1-p)$

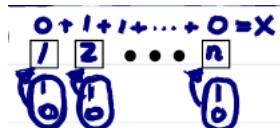
实例 丢一次硬币, p 代表某一面出现的概率

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 Bernouli 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 Bernouli 实验 $X_1 + \dots + X_n, X_i \sim B(p)$, 则这些随机变量之和 $Y = X_1 + \dots + X_n$ 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 (参见定理3.20)
- 定义

均值 $\mu = np$, 求解方法有:

- 定义 (下述凑一法)
- 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 (下述凑一法)
- 矩母函数在 $t = 0$ 的二阶导, $\text{Var}(X) = E(X^2) - (E(X))^2$
- 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

笔记 求解二项分布的均值时, 可通过将其凑成概率质量函数之和 (为 1) 的形式:

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

此称为凑一法 (sum to one, STO)。同理，求方差时也可使用：

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

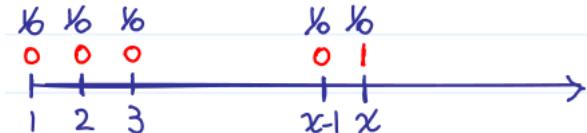
命题 4.1

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

4.1.4 几何分布

定义 4.4

若进行无限次独立的 Bernoulli 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3.20)
- 定义

均值 $\mu = \frac{1}{p}$, 求解方法有：

- $E(X) = \sum_{k=1}^{\infty} P(X \geq k)$
- 矩母函数在 $t = 0$ 的一阶导
- 定义 (下述微分法)

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

- 定义 (下述微分法)
- 矩母函数在 $t = 0$ 的二阶导, $\text{Var}(X) = E(X^2) - (E(X))^2$

实例 买彩票时，中奖所需购买张数。

 **笔记** 微分法求解几何分布均值:

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

微分法求解几何分布方差:

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

$$\begin{aligned} \text{Var}(X) &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= \frac{2}{p^2} + \frac{1}{p} - \frac{1}{p^2} p^2 \\ &= \frac{1-p}{p^2} \end{aligned}$$

命题 4.2

离散分布中有且只有几何分布具备无记忆性:

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

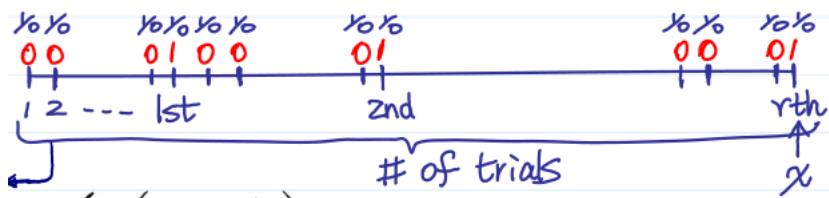


证明

4.1.5 负二项分布

定义 4.5

若进行无限次独立的 Bernoulli 实验，则第 r 次出现“1”的实验次数遵循**负二项分布** (negative binomial distribution)，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 $p(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} p & x \in \mathbb{N}_r \\ 0 & \text{其他} \end{cases}$

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3.20)

• 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4.3)

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, $\text{Var}(X) = E(X^2) - (E(X))^2$

• 独立几何分布之和 (定理4.3)

实例 买彩票时, 中奖 r 次所需购买张数。

命题 4.3

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

证明

命题 4.4

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

证明

4.1.6 多项分布

定义 4.6 (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, \dots, p_r 。令 X_i 代表得出结果 i 的次数, 则随机向量 $\mathbf{X} = (X_1, \dots, X_r)$ 遵循多项分布 (multinomial distribution), 记为 $\mathbf{X} \sim \text{Multinomial}(n, p_1, \dots, p_r)$ 。

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 $p(x) = \begin{cases} \binom{n}{x_1 \dots x_r} p_1^{x_1} \dots p_r^{x_r} & x \in \mathbb{N}^r \& \sum_{i=1}^r x_i = n \\ 0 & \text{其他} \end{cases}$

矩母函数 $M(t_1, \dots, t_r) = (p_1 e^{t_1} + \dots + p_r e^{t_r})^n$, 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 $\text{Var}(X_i) = np_i(1 - p_i)$

协方差 $\text{Cov}(X_i, X_j) = -p_i p_j$, 求解方法有:

- 矩母函数求解 $E(X_i X_j)$
- 混一法求解 $E(X_i X_j)$

实例

注 多项分布是二项分布的泛化。

4.1.7 泊松分布

定理 4.1 (Poisson 逼近)

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



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

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

从而

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

对任意的 k ($k = 0, 1, 2, \dots$) 成立

定义 4.7 (Poisson 分布)

若随机变量 X 的概率分布列满足以下形式:

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

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
- 混一法

均值 $\mu = \lambda$

方差 $\sigma^2 = \lambda$

实例

公共汽车站来到的乘客数

命题 4.5

Poisson 分布之和仍是 Poisson 项分布。若 $X_1 + \dots + X_k$ 独立且 $X_i \sim P(\lambda_i)$, 则 $Y = X_1 + \dots + X_k \sim P(\lambda_1 + \dots + \lambda_k)$



证明

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

$$\begin{aligned} & \because f(1) = [f\left(\frac{1}{n}\right)]^n \\ & \text{且 } a = f(1) = [f\left(\frac{1}{2}\right)]^2 \geq 0 \\ & \therefore f\left(\frac{1}{n}\right) = a^{\frac{1}{n}} \\ & \therefore f\left(\frac{m}{n}\right) = [f\left(\frac{1}{n}\right)]^m = a^{\frac{m}{n}} \end{aligned}$$

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 4.8 (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 $[t_0, t_0 + t]$ 中的发生次数数只与时间间隔长度 t 有关而与时间起点 t_0 无关。以 $P_k(t)$ 记在长度为 t 的时间区间中发生 K 次事件 A 的概率。

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 $\lim_{t \rightarrow 0} \psi(t)/t = 0$



对 $\Delta t > 0$, 考虑 $[0, t + \Delta t)$ 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:

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

(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 $a = 0$, 则 $P_0(t) = 0$, 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 $\Delta t \rightarrow 0$ 时, 有

$$\begin{aligned} P_0(\Delta t) &= e^{-\lambda\Delta t} = 1 - \lambda\Delta t + o(\Delta t) \\ P_1(\Delta t) &= 1 - P_0(\Delta t) - \psi(\Delta t) = \lambda\Delta t + o(\Delta t) \\ \sum_{l=2}^{\infty} P_{k-l}(t)P_l(\Delta t) &\leq \sum_{l=2}^{\infty} P_l(\Delta t) = \psi(\Delta t) = o(\Delta t) \end{aligned}$$

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 $P_0(t) = e^{-\lambda t}$, 故有 $P'_1(t) = \lambda[e^{-\lambda t} - P_1(t)]$, 可解得 $P_0(t) = \lambda t e^{-\lambda t}$, 依次可递推可解得:

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

正是参数为 λt 的泊松分布。

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, 记为 $X \sim h(n, N, M)$.

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, r, \quad (4.1)$$

其中

超几何分布的特征:

参数 $r = \min\{M, n\}$, 且 $M \leq N, n \leq N, n, N, M$ 均为正整数.

$$\text{概率质量函数 } p(k) = \begin{cases} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} & k = 0, 1, \dots, r \\ 0 & \text{其他} \end{cases}$$

矩母函数 存在, 但没有简单的表达

$$\text{均值 } \mu = \frac{rn}{n+m}$$

$$\text{方差 } \sigma^2 = \frac{rnm(n+m-r)}{(n+m)^2(n+m-1)}$$

实例 从一个有限总体中进行不放回抽样

注 若 $m, n \rightarrow \infty$ 时有 $p_{m,n} = \frac{n}{m+n} \rightarrow p$, 则超几何分布可近似为二项分布。 $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \cong \binom{n}{k} p^k (1-p)^{n-k}$

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 $X \sim U(a, b)$ 。

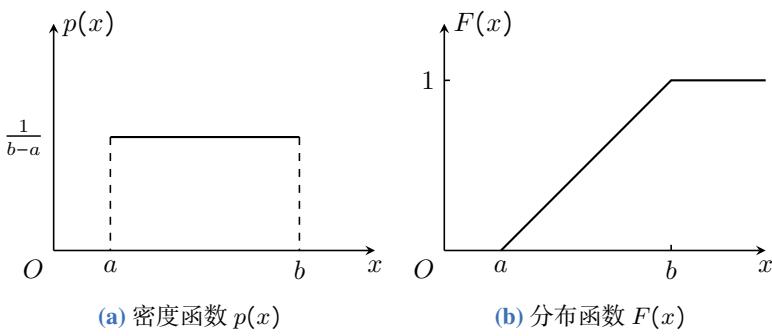


图 4.1: (a, b) 上的均匀分布

均匀分布的特征：

$$\text{参数} \quad a < b, \quad a, b \in \mathbb{R}$$

$$\text{概率质量函数} \quad p(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{其他} \end{cases}$$

$$\text{分布函数} \quad F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x < b; \\ 1, & x \geq b. \end{cases}$$

$$\text{矩母函数} \quad M(t) = \frac{e^{bt} - e^{at}}{t(b-a)}, \quad t \in \mathbb{R}$$

$$\text{均值} \quad \mu = \frac{a+b}{2}$$

$$\text{方差} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

注 常利用 $U(0, 1)$ 生成特定分布的伪随机数（参见定理2.6）

4.2.2 指数分布

定义 4.11

若随机变量 X 的密度函数（见图 4.2）为

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

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

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

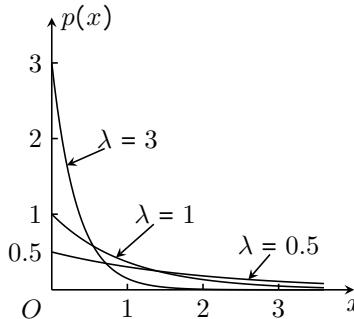


图 4.2: 参数为 λ 的指数分布密度函数

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 有其均值可看出, 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

命题 4.6

如果某设备在任何长为 t 的时间 $[0, t]$ 内发生故障的次数 $N(t)$ 服从参数为 λt 的泊松分布, 则相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布.



证明 设 $N(t) \sim P(\lambda t)$, 即

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

注意到两次故障之间的时间间隔 T 是非负随机变量, 且事件 $\{T \geq t\}$ 说明此设备在 $[0, t]$ 内没有发生故障, 即 $\{T \geq t\} = \{N(t) = 0\}$, 由此我们得

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

当 $t \geq 0$ 时, 有

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

所以 $T \sim Exp(\lambda)$, 相继两次故障之间的时间间隔 T 服从参数为 λ 的指数分布.

4.2.3 伽马分布

定义 4.12

若随机变量 X 的密度函数为

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

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



图 4.3 给出若干条 λ 固定、 α 不同的伽玛密度函数曲线, 从图中可以看出:

- $0 < \alpha < 1$ 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.

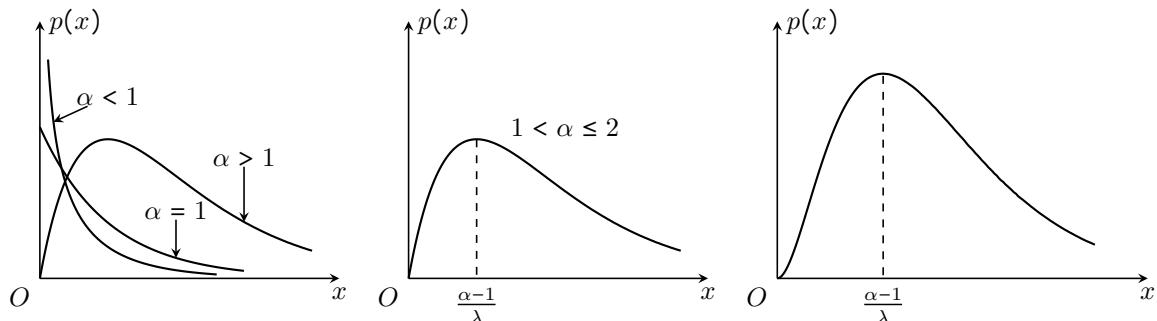


图 4.3: λ 固定、不同 α 的伽玛密度曲线

Gamma 分布的特征:

矩母函数 $M(t) = (\frac{\lambda}{\lambda-t})^\alpha, t < \lambda$

均值 $\mu = \frac{\alpha}{\lambda}$

方差 $\sigma^2 = \frac{\alpha}{\lambda^2}$

Gamma 函数 $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ 的特征:

- $\Gamma(1) = 1$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\Gamma(n) = (n - 1)!, n \in \mathbb{Z}$
- $\Gamma(\frac{\alpha}{2}) = \frac{\sqrt{\pi}(\alpha-1)!}{2^{\alpha-1}(\frac{\alpha-1}{2})!}, \alpha = 2k + 1, k \in \mathbb{N}$

命题 4.7

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, \dots, X_k , i.i.d. $\sim E(\lambda)$, 则 $Y = X_1 + \dots + X_k \sim \Gamma(k, \lambda)$ 。更进一步有: 若随机变量 X'_1, \dots, X'_k , i.i.d. $\sim \Gamma(\alpha_i, \lambda)$, 则 $Y' = X'_1 + \dots + X'_k \sim \Gamma(\alpha_1 + \dots + \alpha_k, \lambda)$ 。



证明

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

$$M_{Y'}(t) = \prod_{i=1}^k M_{X'_i}(t_i) = \prod_{i=1}^k \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_i} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1 + \dots + \alpha_k}$$

命题 4.8

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



证明

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

命题 4.9

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



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 4.10



4.2.4 贝塔分布

定义 4.13

称以下函数

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

为贝塔函数, 其中参数 $a > 0, b > 0$.



命题 4.11

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

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

由伽玛函数的定义知

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

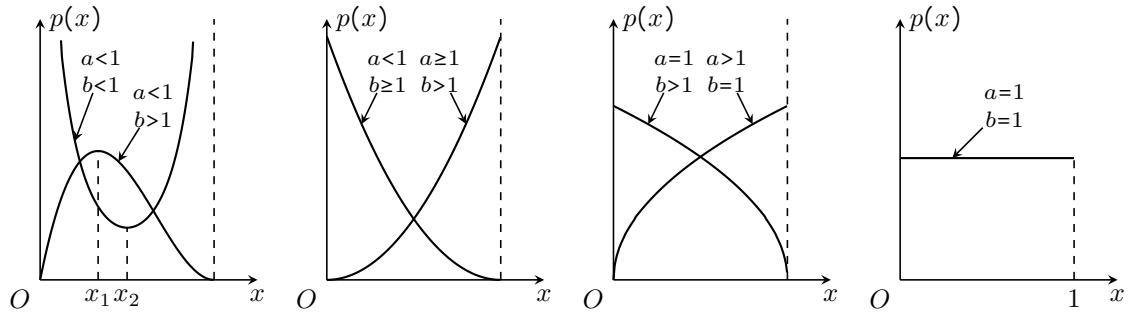


图 4.4: 贝塔密度函数曲线

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

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

$\text{beta}(a, b)$,

定义 4.14

若随机变量 X 的密度函数为

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

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 4.4 可以看出:

- $a < 1, b < 1$ 时, $p(x)$ 是下凸的单峰函数.
- $a > 1, b > 1$ 时, $p(x)$ 是上凸的单峰函数.
- $a < 1, b \geq 1$ 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 $Be(1, 1) = U(0, 1)$.

Beta 分布的特征:

$$\text{矩母函数} \quad M(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{a+r}{a+b+r} \right) \frac{t^k}{k!}, \quad t < \lambda$$

$$\text{均值} \quad \mu = \frac{a}{a+b}$$

$$\text{方差} \quad \sigma^2 = \frac{ab}{(a+b+1)(a+b)^2}$$

注

$$\beta(1, 1) = U(0, 1)$$

命题 4.12

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

4.3 正态分布及其导出分布

4.3.1 正态分布

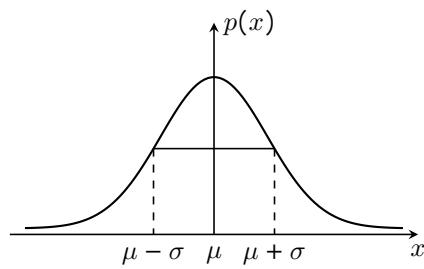
定义 4.15

若随机变量 X 的密度函数为

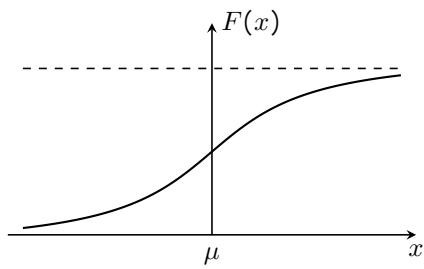
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R} \quad (4.2)$$

则称 X 服从正态分布 (Normal distribution), 称 X 为正态变量, 记作 $X \sim N(\mu, \sigma^2)$ 。其中参数 $\mu \in \mathbb{R}, \sigma > 0$ 。

正态分布的密度函数 $p(x)$ 的图形如图 4.5(a) 所示, 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。



(a) 密度函数 $p(x)$



(b) 分布函数 $F(x)$

图 4.5: 正态分布

正态分布的特征:

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

均值 $\mu = \mu$

方差 $\sigma^2 = \sigma^2$

命题 4.13

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 4.14

若独立随机变量 X_1, \dots, X_k 满足 $X_i \sim N(\mu_i, \sigma_i^2)$, 则其和 $Y = X_1 + \dots + X_k \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2)$

证明

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

推论 4.1

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

定理 4.3

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

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



证明

4.3.2 卡方分布**定义 4.16**

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

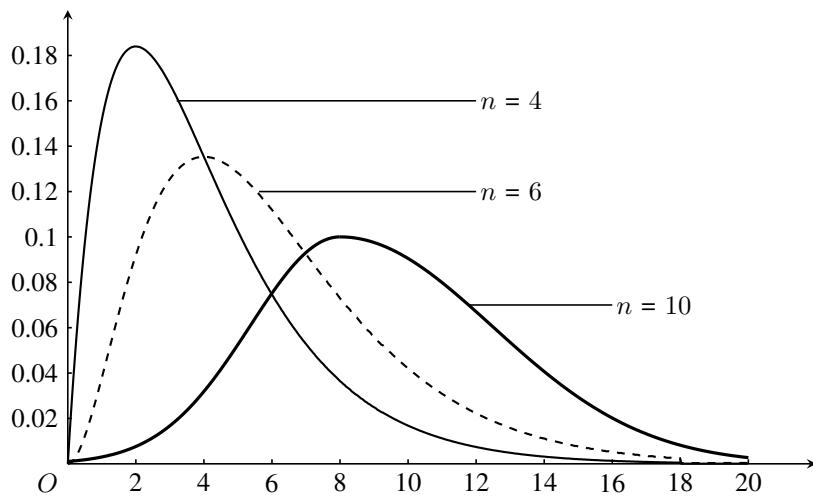


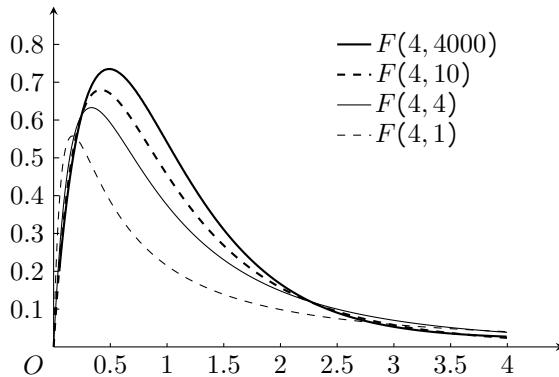
卡方分布的特征:

参数 $n \in \mathbb{N}_+$ 概率密度函数 $p(x) = \begin{cases} \frac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y > 0 \\ 0 & \text{其他} \end{cases}$ 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 4.6。

图 4.6: $\chi^2(n)$ 分布的密度函数

图 4.7: F 分布的密度函数, 是一个只取非负值的偏态分布**命题 4.15**

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



证明

推论 4.2

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

**4.3.3 F 分布****定义 4.17**

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

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

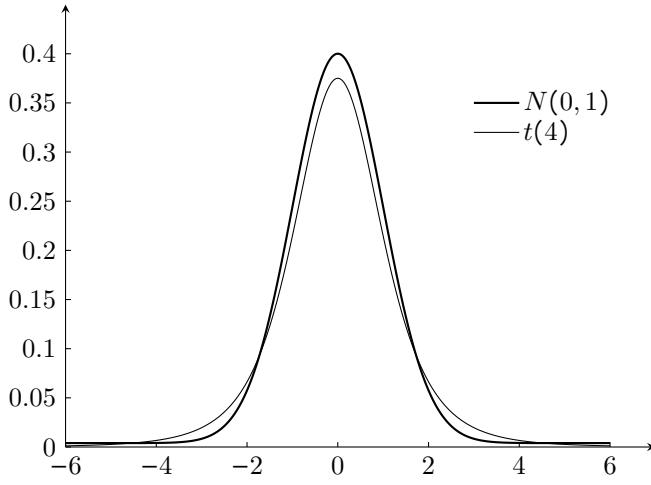
矩母函数 不存在

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 4.16 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

图 4.8: t 分布与 $N(0, 1)$ 的密度函数, t 分布尾更重

变量商的分布的密度函数公式2.2。 Z 的密度函数为

$$\begin{aligned} p_Z(z) &= \int_0^{+\infty} x_2 p_1(zx_2) p_2(x_2) dx_2 \\ &= \frac{z^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} 2^{\frac{m+n}{2}} \int_0^{+\infty} x_2^{\frac{m+n}{2}-1} e^{-\frac{x_2}{2}(1+z)} dx_2. \end{aligned}$$

运用变换 $u = \frac{x_2}{2}(1+z)$, 可得

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

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 4.18

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 4.8), 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 $p(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}\pi\Gamma(\frac{n}{2})} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$

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

命题 4.17

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 求导可得 t 分布的密度函数为

$$\begin{aligned} p_t(y) &= y p_F(y^2) = \frac{\Gamma(\frac{1+n}{2})(\frac{1}{n})^{\frac{1}{n}}}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} (y^2)^{\frac{1}{2}-1} \left(1 + \frac{1}{n}y^2\right)^{-\frac{1+n}{2}} y \\ &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}\pi\Gamma(\frac{n}{2})} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < y < +\infty. \end{aligned}$$

4.3.5 柯西分布

定义 4.19

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 $X \sim C(\mu, \sigma)$ 。其中参数 $\mu \in \mathbb{R}, \sigma > 0$ 。



柯西分布的特征：

累积函数 $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(\mu + \sigma x)$

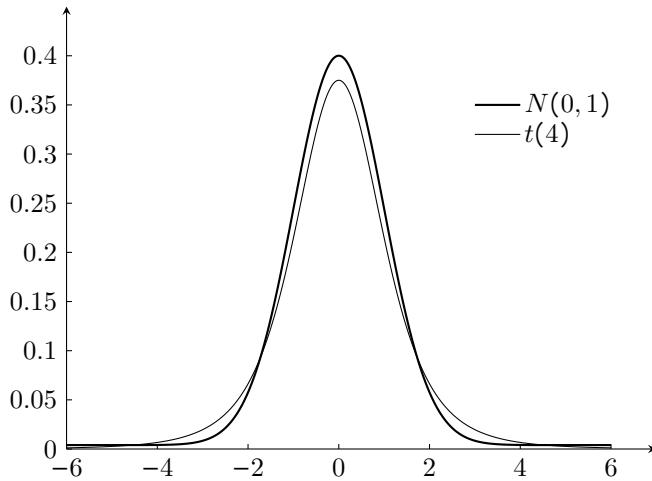
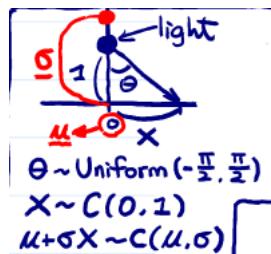
矩母函数 除 $t = 0$ 外不存在

特征函数 $\phi_X(t) = \exp(i\mu t - \sigma|t|)$

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。

图 4.9: t 分布与 $N(0, 1)$ 的密度函数**命题 4.18**

$$C(0, 1) = t_1$$

命题 4.19

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 4.20**

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 4.21

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

证明

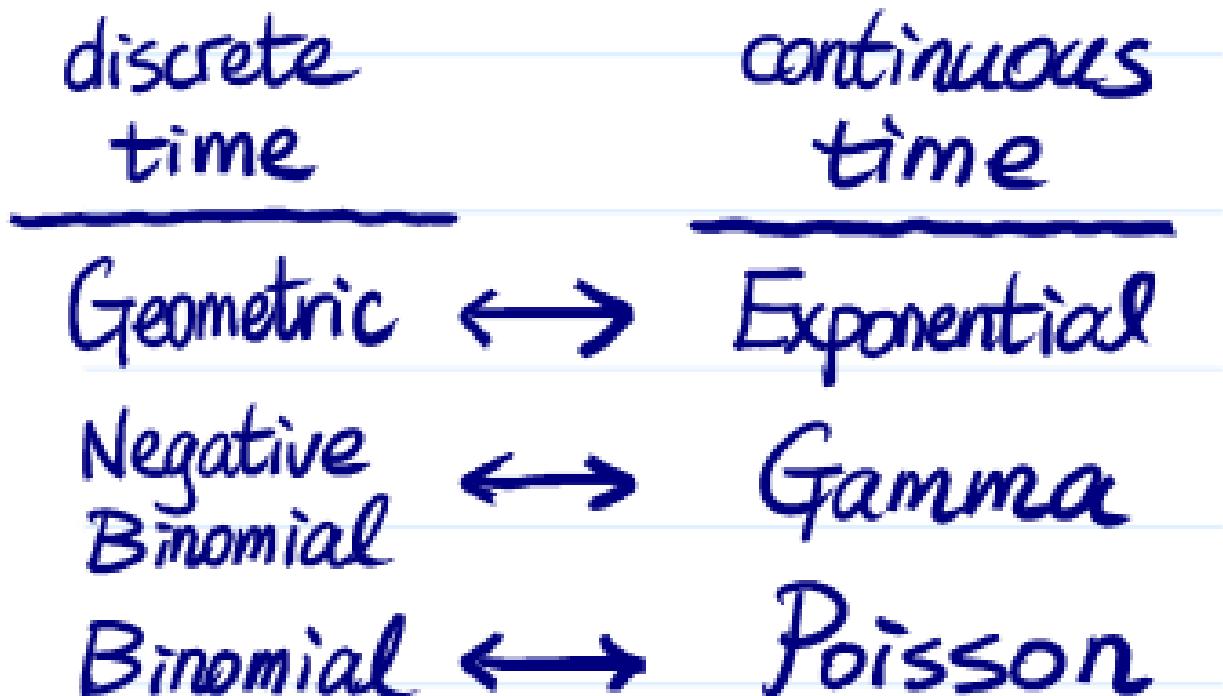
$$\phi_{X_1+\dots+X_n} = \prod_{i=1}^n \phi_{X_i}(t) = \exp(it \sum_{i=1}^k \mu_i - |t| \sum_{i=1}^k \sigma_i)$$

推论 4.3

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

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

4.4 各分布间关系



可参考网站 <http://www.math.wm.edu/~leemis/chart/UDR/UDR.html>

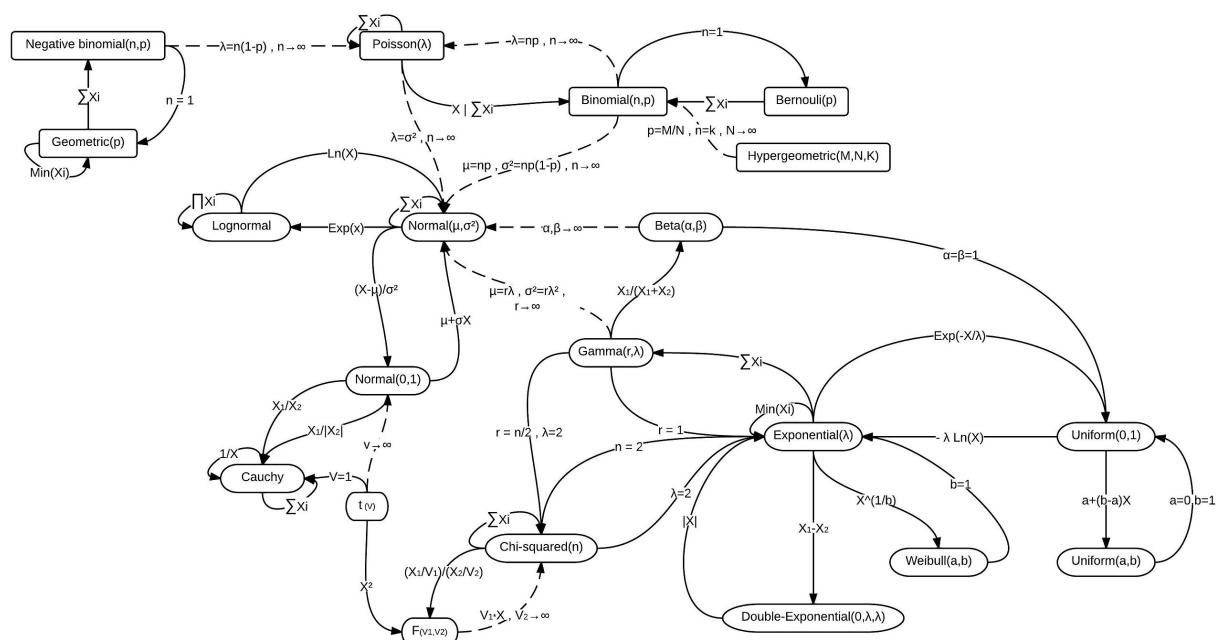


图 4.10: 各分布间的联系

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，如雷尔-康特立引理
- 中心极限定理

5.1 收敛

定义 5.1

若随机变量序列 $\{X_n : \Omega \rightarrow \mathbb{R}\}$ 与随机变量 $X : \Omega \rightarrow \mathbb{R}$ 间存在以下关系：

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon\right\} = 1, \forall \epsilon > 0$$

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 $\lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon$ 的样本点，这些样本点组成事件的概率为 1。

定义 5.2

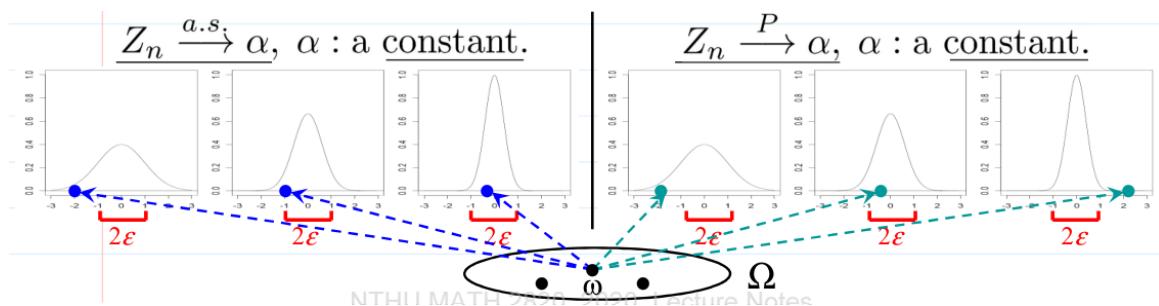
若随机变量序列 $\{X_n : \Omega \rightarrow \mathbb{R}\}$ 与随机变量 $X : \Omega \rightarrow \mathbb{R}$ 间存在以下关系：

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\} = 1, \forall \epsilon > 0.$$

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 Z_i , 事件 $A_i = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\}$, $P(A_i)$ 将逐渐变为 1。



注 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，下一次又离开收敛变量，只有保证总体概率趋近 1 即可。

易见 $X_n \xrightarrow{\text{a.s.}} X_\infty \implies X_n \xrightarrow{\mathbb{P}} X_\infty$, 但是反之不成立，例如概率空间 $((0, 1], \mathcal{B}_{(0,1]}, \text{Uniform}((0, 1]))$ 上的打字机序列

$$\xi_n = \mathbb{1}_{(n/2^k - 1, (n+1)/2^k - 1]}, \quad 2^k \leq n < 2^{k+1}$$

适合 $\mathbb{P}\{|\xi_n| > \epsilon\} \leq 1/2^k \rightarrow 0$, 而 $\overline{\lim} \xi_n \equiv 1$ 且 $\underline{\lim} \xi_n \equiv 0$, 所以 $\lim \xi_n(\omega)$ 对 $\forall \omega \in (0, 1]$ 都不存在。

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的实例: 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 $\{X_n : \Omega \rightarrow \mathbb{R}\}$:

$$X_n(\omega) = \begin{cases} 1, \omega \in I_{k,j} \\ 0, \omega \notin I_{k,j} \end{cases}, \quad n = 2^k + j - 2$$

同时定义随机变量 $X : \Omega \rightarrow \mathbb{R}$ 为 $X(\omega) = 0, \forall \omega \in \Omega$

由于对任意 $0 < \epsilon < 1$

$$\mathbb{P}\{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon\} = 1 - \frac{1}{2^K} \rightarrow 1$$

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

$$\mathbb{P}\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| < \epsilon\right\} = \mathbb{P}\{\emptyset\} = 0$$

所以 X_n 不几乎必然收敛到 X 。

定义 5.3

设随机变量 X, X_1, X_2, \dots 的分布函数分别为 $F(x), F_1(x), F_2(x), \dots$ 。若对 $F(x)$ 的任一连续点 x , 都有

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

则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 记作 $F_n(x) \xrightarrow{W} F(x)$; 也称 $\{X_n\}$ 按分布收敛于 X , 记作 $X_n \xrightarrow{d} X$

**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。

**定理 5.1 (连续映射定理)**

设 $g : \mathbb{R}^k \mapsto \mathbb{R}$ 为连续函数, 则随机向量序列通过次映射后收敛情况与原先相同:

$$X_n^{(i)} \xrightarrow{\text{a.s.}} X^{(i)} \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{\text{a.s.}} g(X^{(1)}, \dots, X^{(k)})$$

$$X_n^{(i)} \xrightarrow{\mathbb{P}} X^{(i)} \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{\mathbb{P}} g(X^{(1)}, \dots, X^{(k)})$$

$$(X_n^{(i)}) \xrightarrow{d} (X^{(i)}) \Rightarrow g(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{d} g(X^{(1)}, \dots, X^{(k)})$$



注 随机向量依分布收敛是指联合分布函数收敛。

定理 5.2 (Slutsky 定理)

若 $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} a$, 其中 a 是常数, 则:

- $Y_n X_n \xrightarrow{d} aX$
- $Y_n + X_n \xrightarrow{d} a + X$
- $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a}$, 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 $g_1(X_n, Y_n) = X_n Y_n, g_2(X_n, Y_n) = X_n + Y_n, g_3(X_n, Y_n) = \frac{X_n}{Y_n}$ 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,

$$\forall t \in U(0, \delta), s.t. \lim_{n \rightarrow \infty} M_n(t) = M(t)$$

是在 $F(x)$ 的所有连续点上 $F_n(x) \xrightarrow{M} F(x)$ 的充要条件。



注 若矩母函数不存在, 替换成特征函数, 定理也成立。但是, 不能替换成密度函数或质量函数

例题 5.2 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 $P_n(x) \rightarrow 0, \forall x \in \mathbb{R}$ (不满足质量函数定义), 并且 $\lim_{n \rightarrow \infty} F_n(0) = \frac{1}{2}$ (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 5.3 泊松分布收敛于正态分布 令 $X_n \sim P(\lambda_n)$, 并且 $\lim_{n \rightarrow \infty} \lambda_n = \infty$ 。将变量标准化:

$$Z_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

那么

$$M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} M_{X_n}(t) \left(\frac{t}{\sqrt{\lambda_n}} \right) = \exp(-t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}} - 1}))$$

由于

$$\lim_{n \rightarrow \infty} \ln(M_{Z_n}(t)) = \lim_{\lambda_n \rightarrow \infty} -t\sqrt{\lambda_n} + \lambda_n(e^{\frac{t}{\sqrt{\lambda_n}} - 1}) = \frac{t^2}{2}$$

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

定理 5.5 (弱大数定理)

设随机变量 X_1, \dots, X_n 相互独立, 且 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$, 那么

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu$$



注 柯西分布无均值、方差, 不能使用此定理。

证明 易知

$$E(\bar{X}_n) = \mu, \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

再通过切比雪夫不等式有：

$$P(|\bar{X}_n - \mu| > \epsilon) < \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0$$

定理 5.6 (强大数定理)

设随机变量 X_1, \dots, X_n 相互独立，且 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$ ，那么

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{a.s.}} \mu$$



例题 5.4 蒙特卡洛积分 为计算积分 $I(f) = \int_0^1 f(x)dx$ ，生成随机数 X_1, \dots, X_n i.i.d. $\sim U(0, 1)$ ，并且计算 $\hat{I}(f) = \bar{Y}_n, Y_i = f(X_i)$ 。由于 X_1, \dots, X_n i.i.d.，所以 Y_1, \dots, Y_n i.i.d.；并且 $E(Y_i) = E(f(X_i)) = \int_0^1 f(x) \cdot 1 dx = I(f)$ ，所以 $\hat{I}(f) \xrightarrow{\mathbb{P}} E(Y_i) = I(f)$

例题 5.5 样本方差 设随机变量 X_1, \dots, X_n i.i.d.，并且均值和方差分别为 μ, σ^2 。令：

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - \bar{X}_n^2 +$$

由于 $g(x) = x^2$ 是连续函数，所以 $\bar{X}_n^2 \xrightarrow{\mathbb{P}} \mu^2$ 。由于 X_1^2, \dots, X_n^2 i.i.d.，且 $E(X_i^2) = \sigma^2 + \mu^2$ ，所以 $\bar{X}_n^2 \xrightarrow{\mathbb{P}} \mu^2 + \sigma^2$ 。因此：

$$S_n^2 \xrightarrow{\mathbb{P}} \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

例题 5.6 若 $X_n \sim t_n$ ，则 $X_n \xrightarrow{d} N(0, 1)$ 。令 $Z \sim N(0, 1), U_1, \dots, U_n \sim \chi_1^2$ 相互独立，则 $\frac{Z}{\sqrt{(U_1+\dots+U_n)/n}} \sim t_n$ 。由于 $E(U_i) = 1$ ，所以 $\bar{U}_n \xrightarrow{\mathbb{P}} 1$ ，所以 $\sqrt{\bar{U}_n} \xrightarrow{\mathbb{P}} 1$ ，所以 $t_n = \frac{Z}{\sqrt{\bar{U}_n}} \xrightarrow{d} N(0, 1)$ (定理5.2)。

例题 5.7 若 $X_n \sim F_{m,n}$ ，则 $mX_n \xrightarrow{d} \chi_m^2$ 。令 $U \sim \chi_m^2, V_1, \dots, V_n \sim \chi_1^2$ 相互独立，则 $\frac{U/m}{(V_1+\dots+V_n)/n} \sim F_{m,n}$ 。由于 $E(V_i) = 1$ ，所以 $\bar{V}_n \xrightarrow{\mathbb{P}} 1$ ，所以 $mF_{m,n} = \frac{U}{\bar{V}_n} \xrightarrow{d} \chi_m^2$ (定理5.2)。

5.3 中央极限定理

定理 5.7 (中央极限定理)

设随机变量 X_1, \dots, X_n i.i.d.，且 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2$ ，令

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, T_n = \sum_{i=1}^n X_i$$

则：

$$\lim_{n \rightarrow \infty} P\left(\frac{T_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \lim_{n \rightarrow \infty} P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x\right) = \Phi(x)$$

其中 $\Phi(x)$ 是 $N(0, 1)$ 的累积函数，即：

$$\frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$



证明 令 $W_i = \frac{X_i - \mu}{\sigma}$, 则 $E(W_i) = 0$, $\text{Var}(W_i) = 1$ 。再令 $Z_n = \frac{T_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i$ 。则

$$\begin{aligned} M_{Z_n}(t) &= \left[M_W\left(\frac{t}{\sqrt{n}}\right) \right]^n \\ &= \left[M_W(0) + M'_W(0) \frac{t}{\sqrt{n}} + \frac{M''_W(0)}{2} \left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\frac{1}{n}\right) \right]^n \\ &= \left[1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right]^n \\ &\rightarrow e^{\frac{t^2}{2}} \end{aligned}$$

根据定理5.4, $Z_n \xrightarrow{d} N(0, 1)$ 。

注 若矩母函数不存在, 可替换为特征函数。

例题 5.8 二项分布近似为正态分布 若随机变量 X_1, \dots, X_n i. i. d. $\sim B(1, p)$, 则 $T_n \sim B(n, p)$. 其中 $E(T_n) = nE(X_i) = np$, $\text{Var}(T_n) = n\text{Var}(X_i) = np(1-p)$. 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4.10
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.

例题 5.9 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, \dots, X_n , 其均值与方差分别为 μ, σ^2 . 由中央极限定理可知 $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$. 由例5.5可知 $S_n^2 \xrightarrow{\mathbb{P}} \sigma^2$, 即 $\frac{\sigma}{S_n} \xrightarrow{\mathbb{P}} 1$ (定理5.2). 所以:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \frac{\sigma}{S_n} \xrightarrow{d} N(0, 1) \cdot 1 = N(0, 1)$$

即可以通过分布 $N(0, \frac{\sigma^2}{n})$ 估计测量误差 $\bar{X}_n - \mu$

注 由定理4.3可知 $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$. 但由例5.6可知, $n \rightarrow \infty$ 时, t 分布收敛于正态分布, 所以不冲突.

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, \dots 等记之，称之为这个空间的**集合**。空集记为 \emptyset 。 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 $A^c := \{x \in X | x \notin A\}$

并 $A \cup B := \{x | (x \in A) \vee (x \in B)\}$

交 $A \cap B := \{x | (x \in A) \wedge (x \in B)\}$

差 $A \setminus B := \{x | (x \in A) \wedge (x \notin B)\}$

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

$$\lim_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{+\infty} A_n$$

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

$$\lim_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{+\infty} A_n$$

为它的极限.

非降或非增的集合序列统称为单调序列. 因此, 单调集合序列总有极限.



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:

$$\begin{aligned}\underline{\lim}_{n \rightarrow \infty} A_n &= \underline{\lim}_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{+\infty} A_k \\ \overline{\lim}_{n \rightarrow \infty} A_n &= \overline{\lim}_{n \rightarrow \infty} B'_n = \bigcap_{n=1}^{\infty} B'_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{+\infty} A_k\end{aligned}$$

分别称其为 $\{A_n\}$ 的上极限与下极限



显然, 记号 $x \in \overline{\lim}_{n \rightarrow \infty} A_n$ 意味元素 x 属于序列 $\{A_n\}$ 中的无穷多个集合, 而记号 $x \in \underline{\lim}_{n \rightarrow \infty} A_n$, 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

$$\underline{\lim}_{n \rightarrow \infty} A_n \subseteq \overline{\lim}_{n \rightarrow \infty} A_n$$

定义 A.5

如果 $\underline{\lim}_{n \rightarrow \infty} A_n = \overline{\lim}_{n \rightarrow \infty} A_n$, 则认为 $\{A_n\}$ 的极限存在, 并且将

$$\lim_{n \rightarrow \infty} A_n := \underline{\lim}_{n \rightarrow \infty} A_n = \overline{\lim}_{n \rightarrow \infty} A_n$$

称为它的极限.



A.1.2 集合系

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_1 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_1 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 $r(r \leq n)$ 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，放回后再取下一个，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 $r(r \leq n)$ 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，放回后再取下一个，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