《概率与数理统计》讲义

作者：张昕
组织：经济学院
时间：March 5, 2023
版本：0.2

没必要的事不做，必要的事尽快去做。——折木奉太郎

目录

第 0 章 高中知识复习

0.1 组合理论

0.1.1 基本计数法则与列表

定理 0.1 (基本计数法则)

- 假定进行过程 I 有 n_1 种方法，而对于过程 I 的每一种方法，进行过程 II 都有 n_2 种方法。那么，依次进行过程 I 与 II 共有 $n_1 \times n_2$ 种方法。
- 假定进行过程 I 有 n_1 种方法，进行过程 II 都有 n_2 种方法。那么，进行过程 I 或 II 共有 $n_1 + n_2$ 种方法。

♡

例题 0.1 可重复的列表 (List with repetition) 7 位数的车牌，前三个必须是字母，后四个必须是数字，一共有多少编排车牌号的方法。

例题 0.2 不可重复的列表 (List without repetition) 7 位数的车牌，前三个必须是字母，后四个必须是数字，如果数字和字母均不能重复，一共有多少编排车牌号的方法。

我们将不可重复的列表称为排列（permutations）。

0.1.2 排列

定义 0.1 (排列与阶乘)

- 排列（**Permutations**）：与顺序有关的不可重复列表。
- 阶乘（**factorial**）：$n! = n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$。

♣

例题 0.3 一个球队有 9 名队员，先后发球，发球顺序共有多少种？

例题 0.4 某班级共有 6 名男生和 4 名女生，现根据班上某次测验成绩排序，假设所有同学成绩均不相同。

1. 一共有多少种排序方法？
2. 若男女分开排序，一共有多少种排序方法？

例题 0.5 十本书放在书架上，其中有 4 本数学书、3 本化学书、2 本历史书和一本语文书。现在要求所有相同类型书籍均紧挨着放，问一共有多少种放法？

0.1.3 组合

从 A，B，C，D 和 E 五个字母中选择三个形成组合，一共有几种选法？第一个字母有 5 种选法，第二个字母有 4 种选法，第三个字母有 3 种选法，但该结果是与顺序相关的。但在这个问题中，字母的顺序是无关的。所有选法应该有

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!r!} = \frac{5\cdot 4\cdot 3}{3\cdot 2\cdot 1} = 10$$

定义 0.2 (组合)

组合 (**Combinations**) 是元素不可重复的集合，因此与顺序无关。

对于 $r \leqslant n$，我们将 $\mathbf{C_n^r}$ 或者 $\binom{n}{r}$ 定义为

$$\mathbf{C_n^r} = \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

我们将 $\mathbf{C}_\mathbf{n}^\mathbf{r}$ 或者 $\binom{n}{r}$ 读作“n 取 r（n choose r）”，用来代表从 n 个元素中一次性选择 r 个元素的可能组合数。

例题 0.6 从 20 个人总选择 3 个组成委员会，一共有几种选法？

例题 0.7 12 个人中 5 女 7 男，现从中选取 2 女 3 男组成委员会，共有多少种选法？现在两位男士发生矛盾，并决定绝不一起工作，现在有多少种选法？

例题 0.8 现有一排 n 个信号塔，其中 m 个失效，n-m 个有效，有效信号塔之间没有差异，无效信号塔之间也没有差别。现在不能让连续两个信号塔均失效，共有多少种排序方法？

推论 0.1

一个非常有用的组合公式

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

这个公式的理解方法为，我将选 r 个元素分成两类：选某个特定元素和不选某个特定。如果选就是 $\binom{n-1}{r-1}$，如果不选就是 $\binom{n-1}{r}$。

定理 0.2 (二项式定理（The bionomial theorem））

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

例题 0.9 展开 $(x+y)^3$。

例题 0.10 证明 $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n$。

例题 0.11 一个有 n 个元素的集合，一共有几个子集，其中非空子集有几个。

例题 0.12 用“PEPPER”的六个字母排序，一共有多少种不同排序方式？

例题 0.13 9 面小旗，其中 4 红 2 白 3 蓝，不同排列方式代表不同信号，那么一共有多少种可能的信号？

0.1.4 划分

考虑下面的情境，将 n 个不同的元素分配到 r 个不同的组别，每个组别包含 $n_1, n_2, \cdots, n_r$ 个元素，并且 $\sum_{i=1}^{r} n_i = n$。第一组有 $\binom{n}{n_1}$ 种选法；对于第一组的每个选法，第二组又有 $\binom{n-n_1}{n_2}$ 种选法，以此类推。最终选法有

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

我们将这种关系称为**划分（partition）**。

定理 0.3 (划分与多项式定理（Partition and multinomial theorem））

如果 $n_1+n_2+\cdots+n_r=n$，我们定义

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

则

$$(x_1+x_2+\cdots+x_r)^n = \sum \binom{n}{n_1, n_2, \cdots, n_r} {x_1}^{n_1} {x_2}^{n_2} \cdots {x_r}^{n_r}$$

例题 0.14 警局有 10 名警察，现在需要 5 名巡逻，2 名值班，3 名待命。问将警察按照这样的工作分成三组共有多少种分法？

例题 0.15 10 名小朋友分成两组，A 组 5 名去踢足球，B 组 5 名去打篮球，一共有多少种分法？

例题 0.16 10 名小朋友分成两组，每组 5 名，两组进行篮球比赛，一共有多少种分法？

例题 0.17 现有 $n = 2^m$ 个小朋友进行淘汰赛，即将 n 个参赛者随机分成 $n/2$ 对，两两进行淘汰赛，胜者晋级，败者淘汰。直到最后一个胜者成为冠军。(为了简单，我们就考虑仅有 8 名参赛者。)

1. 第一轮比赛中，有多少不同可能的结果？
2. 整个淘汰赛，一共有多少种不同的结果，每一个结果展示了各论全部的淘汰信息。

例题 0.18 展开 $(x_1 + x_2 + x_3)^2$

0.1.5 插缝法与整数解的个数

再考虑下列情境，将 n 个不同的球放进 r 个不同的篮子，则一共有 r^n 种放法。若将 n 个相同的球放进 r 个不同的篮子，会有几种放法？我们可以把 n 个相同球排成一列，将其分成 r 组，实际就是在 $n-1$ 个缝隙中选 $r-1$ 个，即 $\binom{n-1}{r-1}$。若现在不允许篮子空置，则有 $\binom{n+r-1}{r-1}$ 种选法。

例题 0.19 $x_1 + x_2 = 3$ 的非负整数解和正整数解的个数分别是多少。

例题 0.20 再来考虑例 0.8 中信号塔的故事，是否有新的解法呢？那现在要求两个失效信号塔之间至少有两个有效信号塔，有几种排法？

0.2 集合论

0.2.1 集合基础知识与韦恩图

定义 0.3 (包含与相等)

包含 (Contain) $A \supset B$，称为 A 包含 B，或 B 包含于 A。

相等 (Indentical) $A = B \iff A \subset B$ and $A \supset B$.

♣

定义 0.4 (交、并、补)

补 (Complement) $A^c\ or\ \bar{A} = \{\omega |\ \omega \notin A\}$

并 (Union) $A \cup B = \{\omega |\ \omega \in A\ or\ \omega \in B\}$

交 (Intersection) $A \cap B = \{\omega |\ \omega \in A\ and\ \omega \in B\}$.

♣

定理 0.4 (集合运算律 (the laws of set operations))

交换律 (Commutative laws) $A \cup B = B \cup A$，$A \cap B = B \cap A$。

结合率 (Associative laws) $(A \cup B) \cup C = A \cup (B \cup C)$，$(A \cap B) \cap C = A \cap (B \cap C)$。

分配率 (Distributive lwas) $(A \cup B) \cap C = AC \cup BC$，$(AB) \cup C = (A \cup C) \cap (B \cup C)$

自反率 $(A^c)^c = A, A = A \cap A = A \cup A$。

摩根律 (DeMorgan's laws) $(A \cup B)^c = A^c \cap B^c$，$(A \cap B)^c = A^c \cup B^c$。

♡

定义 0.5 (差、互斥)

差 (Difference) $A \backslash B = A \cap B^c = \{\omega |\ \omega \in A\ and\ \omega \notin B\}$，$A \backslash B = A - B = A - (A \cap B)$

互斥 (Disjoint) $A \cap B = \varnothing$

♣

0.2.2 集合的大小

定义 0.6 (等价)

等价 (equivalent): 两个集合 A 和 B,A 和 B 中的元素可以建立一一对应关系，即元素一样多，记作 $A \sim B$。♣

例题 0.21 $A = \{1, 2, 3, \cdots, n\}$ 与 $\mathbb{N} = \{1, 2, 3, \cdots, n, \cdots\}$ 哪个元素比较多？

解 显然，自然数集合 $\mathbb{N}$ 比较大。

例题 0.22

1. 奇数和偶数哪个比较多？
2. 自然数和偶数那个比较多？
3. 自然数和非负整数哪个比较多？

解

n,n>0	1	2	3	···	n	···
2n	2	4	6	···	2n	···
2n-1	1	3	5	···	2n-1	···
n-1	0	1	2	···	n-1	···

定义 0.7 (可数)

可数 **(countable)**: 可与自然数 $1, 2, 3, \cdots, n, \cdots$ 建立一一对应关系。整数 ($\mathbb{Z}$)、有理数 ($\mathbb{Q}$) 都可数。 ♣

例题 0.23 $[0, 1]$ 和 $\mathbb{N}$ 哪个元素比较多？

定义 0.8 (不可数 (uncountable))

不可数 **(uncountable)**：既不有限又不可数。无理数和实数 ($\mathbb{R}$) 均不可数。 ♣

定义 0.9 (集合的大小)

有限 (finite) 可与自然数 $1, 2, 3, \cdots, n$ 建立一一对应关系；

无限 (infinite) 不是有限；

可数 (countable) 可与自然数 $1, 2, 3, \cdots, n, \cdots$ 建立一一对应关系；

不可数 (uncountable) 既不有限又不可数；

至多可数 (at most countable) 有限或可数。 ♣

在后续的章节中，我们要接触**离散随机变量**和**连续随机变量**。

离散随机变量是指结果至多可数，比如投掷一次硬币，结果有 2 个，是有限多个；如果投掷硬币无限多次，就是可数多次。常言道，在赌场上你可以侥幸赢一两次，但一直赌下去吃亏的一定是你。这就涉及极限和级数的知识。

连续随机变量是指结果有不可数种，最典型的是时间，比如每个时点上股价的涨跌走势。这就涉及微积分的知识。所以请大家及时复习。

第 1 章 随机事件及其概率

1.1 样本空间和随机事件

定义 1.1 (基本概念)

随机现象 无法准确预知其结果的现象。

随机试验 (Experiment) 虽然全体可能结果可以预知，但是结果不能完全预知的试验。

样本空间 (Sample space) 随机试验每个可能的结果称为一个样本点，样本点组成的集合称为样本空间。

随机事件 (Event) 样本空间的子集 E 称为随机事件，即随机事件是包含着随机试验结果的集合，如果随机事件的结果包含于 E，那么我们就称事件 E 发生了。

♣

我们以投掷硬币为例：

- 我们无法准确预知硬币落下时哪一面朝上，因此投掷硬币哪一面朝上是一个<u>随机现象</u>。
- 但是我们可以预知的是，投掷硬币哪一面朝上只有两个可能的结果，正面朝上（Head）和背面朝上（Tail），因此这是一个<u>随机试验</u>。
- 正面朝上（Head）和背面朝上（Tail）是该随机试验的两个样本点，其构成的集合 $S = \{H, T\}$ 称为<u>样本空间</u>。
- $E = \{H\}$，因为 $H \in E \subset S$，所以 E 代表<u>事件</u>正面朝上发生了。

例题 1.1

1. 一个试验要先后投掷两个硬币，用集合写出其样本空间，以及事件第一枚硬币正面朝上。
2. 一个试验要同时投掷两个骰子，用集合写出其样本空间，以及事件骰子点数和为 7。
3. 一个试验要测试灯泡的寿命（单位：小时），用集合写出其样本空间，以及事件灯泡寿命不足 7 小时。
4. 一个试验要先后投掷两个硬币，用集合分别写出事件 E 第一枚硬币正面朝上，事件 E 第二枚硬币正面朝上，以及事件 U 至少一枚硬币正面朝上（即两枚硬币均不背面朝上）。
5. 一个试验要先后投掷两个硬币，用集合分别写出事件 E 至少一枚硬币正面朝上，事件 E 至少一枚硬币背面朝上，以及事件 I 恰好一枚硬币正面朝上，另一枚背面朝上。

定义 1.2 (事件的集合运算)

- 包含（Contain）：$E \subset F$，事件 E 必然会导致事件 F 的发生，例如事件考试成绩 90 分必然导致事件考试及格发生。
- 相等 (Equivalent)：$E = F$，即 $E \subset F$ *and* $F \subset E$。例如例 1.1.1 的第 4 题。
- 和事件 (Union)：$U = E \cup F = E + F = \{E \text{ or } F\}$，

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$$

- 积事件 (Intersection)：$I = EF = E \cap F = \{E \text{ and } F\}$，

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots$$

- 互斥 (Exclusive)：$E \cap F = \varnothing$，事件 E 和事件 F 不能同时发生。
- 逆事件或对立事件 (Complement)：$E \cup F = S$ *and* $E \cap F = \varnothing$，记为 $\bar{E} = E^c$，样本空间 S 中事件 E 之外所有的可能结果。
- 差事件：$E - F = E \cap \bar{F} = E \cap F^c$

♣

例题 1.2 用韦恩图（Venn diagram）描述上述关系。

定理 1.1 (集合运算律 (the laws of set operations))

交换律 (Commutative laws) $E \cup F = F \cup E$，$E \cap F = F \cap E$。
结合率 (Associative laws) $(E \cup F) \cup G = E \cup (F \cup G)$，$(E \cap F) \cap G = E \cap (F \cap G)$。
分配率 (Distributive lwas) $(E \cup F) \cap G = EF \cup FG$，$(EF) \cup G = (E \cup F) \cap (F \cup G)$
自反率 $(E^c)^c = E$。
摩根律 (DeMorgan's laws) $(E \cup F)^c = F^c \cap E^c$，$(E \cap F)^c = F^c \cup E^c$。

1.2 随机事件的概率

定义 1.3

定理 1.2 (概率公理)

非负性 $0 \le P(E) \le 1$
完备性 $P(S) = 1$
可列可加性 对于互斥的事件 $E_1,E_2,\cdots$（即 $E_iE_j = \varnothing$ 当 $i \neq j$），则

$$P\left(\bigcup_1^n E_i\right) = \sum_{i=1}^n P(E_i)$$

对满足上述条件的 $P(E)$ 称为事件 E 的概率。

命题 1.1

- $P(\varnothing) = 0$
- $P(E) = 1 - P(E^c)$
- if $E \subset F$, then $P(E) \le P(F)$
- $P(E \cup F) = P(E) + P(F) - P(EF)$. Especially,$P(E \cup F) = P(E) + P(F), for\ EF = \varnothing$.
- $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(EG) + P(EFG)$
- $P(E - F) = P(E) - P(EF)$.Especially,if $F \subset E$,then $P(E - F) = P(E) - P(F)$.

1.3 古典概型（等可能概型）

定义 1.4 (古典概型与古典概率)

试验的样本空间 S 为有限集，即 $S = \{1, 2, \cdots, N\}$, 且每个事件的概率相等，即

$$P(\{1\}) = P(\{2\}) = \cdots = P(\{N\})$$

根据公理 2 与公理 3，则

$$P(\{i\}) = \frac{1}{N} \qquad i = 1, 2, \cdots, N$$

根据公理 3，则

$$P(E) = \frac{E\text{的结果数}}{S\text{的事件数}}$$

例题 1.3 投掷两个骰子，朝上的面数字之和为 7 的概率。
例题 1.4 一个碗里有 6 个白球和 5 个黑球，从中随机取出 3 个球，问恰好 1 白 2 黑的概率。
例题 1.5 在 10 对夫妇中随机选出 5 人，问 5 人中一对夫妇都没有的概率。

例题 1.6 从 6 男 9 女中选出 5 人组成委员会，问该委员会由 3 男 2 女组成的概率。
例题 1.7 箱子里有 n 个球，其中一个做了标记，从中取出 k 个球，问取出被标记的球的概率。
例题 1.8 某人手里有五张扑克牌，这 5 张牌数字是连续的，但又不全是同一花色，我们就称之为“顺子 (Straight)”。问这个人从一副 52 张的扑克牌中恰好摸到一手顺子的概率。
例题 1.9 某人手里有五张扑克牌，其中 3 张点数一样，另 2 张点数也一样，我们就称之为“福尔豪斯 (full house)”，即“三张加一对”。问这个人恰好摸到一手福尔豪斯的概率。
例题 1.10 我们继续打扑克。现在四个人打比赛，将 52 张牌分给 4 位选手，问其中一个人拿到 13 张黑桃 (spades) 的概率，再问每人各都拿到一张 A（ace）的概率。
例题 1.11 房间里面有 n 个人，问所有人生日均不是同一天的概率。n 多大时，此概率小于 $\frac{1}{2}$。
例题 1.12 将扣在桌子上的 52 张牌依次翻开，直到出现一张 A 位置，比较接下来翻开的牌是黑桃 A 和梅花 2 的概率。
例题 1.13 一只橄榄球队有 20 名进攻球员和 20 名防守球员，现在给他们安排住宿，每两人一间。问所有房间都是同类型球员的概率。
例题 1.14 （配对问题）房间里 N 个人参加舞会，所有人把帽子都抛向空中，并混在一起落地，然后所以人再随机拿一顶帽子，问所有人都没有拿到自己原有帽子的概率。
例题 1.15 十对夫妇坐一圈，没有一对夫妇坐一起的概率。

1.4 条件概率

投掷两个骰子，问两个数字之和为 8 的概率。现在我先偷看了一个骰子数字是 3，那么现在两个骰子数字之和为 8 的概率是多少。

E 表示“两个骰子之和为 8”，F 表示“第一个骰子数字为 3”。我们将上述概率称为事件 F 已经发生的情况下 E 发生的条件概率 (conditional Probability)，记作 $P(E|F)$。

定义 1.5 (条件概率 (conditional Probability))

If $P(F) > 0$,then

$$P(E|F) = \frac{P(EF)}{P(F)}$$

♣

例题 1.16 我的钥匙找不到了，但是我有八成的把握我的钥匙在我外衣的衣兜里，我又确信有一半的概率在左衣兜，有另一半的概率在右衣兜。现在我去摸摸了我外衣的左衣兜却没发现钥匙，问钥匙在右衣兜的条件概率。
例题 1.17 投掷一枚硬币两次，问以下条件下两枚硬币都朝上的概率。

1. 第一次正面朝上
2. 至少一次正面朝上

例题 1.18 四个人玩扑克，我们就叫他们东南西北吧。现在南北一共已经拿了 8 张黑桃了，问东有剩下 5 张黑桃里的三张的概率。
例题 1.19 盒子里有 r 个红球和 b 个蓝球。随机从中无放回地取出 n 个球（$n \leq r + b$）。已知在取出来的 n 个球中有 k 个蓝球了，问取出的第一个球是蓝球的概率。
例题 1.20 你要选课了，选我的课得 A 的概率是 1/2，选别的老师的课则得 A 的概率是 2/3。你犹豫不决，决定投个硬币来决定选谁的课。问你选了我的课并且得了 A 的概率。
例题 1.21 箱子里有 8 红 4 白共 12 个球。现在有序无放回地取出两个球。问两个球都是红球的概率。

定理 1.3 (乘法公式 (The multiplication rule))

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

$$P(A_1A_2A_3\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1})$$

♡

例题 1.22 还记得舞会上扔帽子的那 N 个人吗？问他们中有 k 个人恰好拿到了自己原有帽子的概率。

例题 1.23 把 52 张牌随机分成 4 堆，每堆 13 张，问每堆都恰好有一张 A 的概率。

1.5 全概率公式与贝叶斯公式

我们将 E 展开。$E = E \cap S = E \cap (F \cup F^c) = EF \cup EF^c$。

定理 1.4 (乘法公式 (The multiplication rule))

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \end{aligned}$$

例题 1.24

a) 保险公司把客户分成两类，一类为容易出事故者，另一类为安全者。第一类一年内出事故的概率为 0.4，而另一类为 0.2。假如全部人口中容易出事故者占 30%。现在有一新客户来购买保险，问这个人在购买保险后一年内出事故的概率为多少。

b) 假如一个投保人在购买保险后的一年内出了事故，问他是容易出事故者的概率。

例题 1.25 你做选择题，你的策略是，如果会就直接答，如果不会就猜一个。现在你知道正确答案的概率是 p，猜中的概率是 1/m。我判卷的时候发现你把这道题做对了，我就想知道你真正会做这道题的概率。

例题 1.26 一种血液化验的新方法，有 95% 的把握能够把患有某癌症的病人筛查出来，但是该化验同样对健康人有 1% 的假阳性概率。假如目前人群中患有这种癌症的病人仅占 0.5%。现在某病人化验结果为阳性，问该病人确实罹患该癌症的概率。

例题 1.27 本医生认为，如果我认为某病人罹患某病的可能性达到 80%，我会建议他手术，如果我不确定，我会建议他做进一步检查，但是该检查痛苦且昂贵。

现在老李来找我看病，我对他患该病的把握仅有 60%，于是我让老李去做检查，结果检查结果是阳性。正当我要做手术时，我发现老李隐瞒了病史，老李患有糖尿病。

这让我很不开心，虽然这并不影响我 60% 的判断，但却影响了检查的结果，因为这个又贵又痛苦的检查虽然不会给健康人给出阳性报告，但是对患有糖尿病且又患有该病的患者有 30% 的概率给出阳性报告。

请问，现在我该怎么办，是直接做手术，还是让老李重新做检查呢？

例题 1.28 柯南本来有六成把握某人是犯罪嫌疑人。现在得知犯案的人是左撇子，而这个人恰好是个左撇子。已知，人群中左撇子的占比是 20%。问新的证据下，柯南认为这个人犯罪的把握有几成？

例题 1.29 箱子 1 里有 n 个红球，箱子 2 里有 n 个篮球。现在，从箱子 1 里随机拿走一个球，再从箱子 2 里拿一个球放到箱子 1 里。直到两个箱子的球都被取走，即箱子 1 取了 2n 次，箱子 2 拿取了 n 次。问箱子 1 里取走的最后一个球是红球的概率。

定义 1.6 (Odd ration)

$$Odd\ ratio = \frac{P(A)}{P(A^c)}$$

当我们找到新的证据 B，在 B 成立的条件下 A 发生与不发生的概率为：

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

则 A 新的 odd ration 为

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

例题 1.30 硬币 A 有 1/4 的概率正面朝上，硬币 B 有 3/4 的概率正面朝上。现在箱子里有 A 和 B 硬币各一枚，我随机摸了一枚，抛了一次，结果正面朝上，问这枚硬币是 A 硬币的概率。

定理 1.5 (全概率公式 (the law of total Probability))

假设 $F_1, F_2, \cdots, F_n$ 是一组可以构成样本空间的彼此互斥的事件，也就是必定恰有一个事件会发生，这样的事件也被称为完备事件组。那么可以把事件 E 写成一组互斥事件 EF_i 的和事件

$$E = \bigcup_{i=1}^{n} EF_i$$

所以我们有全概率公式

$$P(E) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

定理 1.6 (贝叶斯公式 (Bayes's formula))

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

该公式的含义是我们可以根据试验的结果（E）来修订我们对试验的猜测（$P(F_j)$）。

例题 1.31 一架飞机失踪了。现在推测该飞机会坠落在三个区域中的一个。如果对每个区域进行搜索，在该区域搜索到飞机的概率为 $1 - \beta_i$。

搜索队已经完成对区域 1 的搜索，并未发现飞机。我们就可以根据这个信息去修订我们原来对三个区域搜索到飞机的概率。试问如何修订呢？

直觉告诉我们，在第一区域没搜索到飞机，那飞机坠毁在第一区域的概率会下降，坠毁在其他区域的概率会增加。

例题 1.32 下面我们来介绍一个电影里经常看到的游戏。我有三张形状相同的牌，第一张两面全是红色，第二张两面全是黑色，第三张一面红一面黑。现在随机拿出一张，向上的一面为红色，问你应该赌另一面是黑色呢还是红色呢？

例题 1.33 我有三种电池，第一种连续使用超过 100 小时的概率达到了 0.7，另两种分别是 0.4 和 0.3。已知箱子里三种电池的占比是 2:3:5。

1. 我随机拿了一个电池，问使用时长超过 100 小时的概率。
2. 现在我随便拿的电池的使用时长已经超过 100 小时了，那么他是第 i 种电池的概率。

1.6 独立事件

定义 1.7 (独立 (Independent))

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

例题 1.34 投掷两个硬币，第一个正面为上的概率是 1/2，第二个正面的概率是 1/2，两者是独立的，则两个都是正面的概率是 1/4。

例题 1.35 投掷两个骰子，E_1 表示“骰子之和是 6”，F 表示“第一个骰子数字是 4”，问两者独立吗？现在 E_2 为“两个骰子数字和是 7”，问它和 F 独立吗？。

定理 1.7

If $A \perp B$, then $A \perp B^c$，$A^c \perp B$，$A^c \perp B^c$

例题 1.36 如果 A 和 B 独立，A 也和 C 独立，那么 A 和 BC 独立吗？我们再来投掷两个骰子，A 表示“骰子之和

是 7”，B 表示“第一个骰子为 4”，C 表示“第二个骰子为 3”。已知 A 和 B 独立，同理 A 和 C 也独立。但是很明显 A 和 BC 不独立，因为 $P(A|BC) = 1$。

这表明三个乃至多个事件独立的条件要更苛刻。

定义 1.8

$$A, B\ and\ C\ are\ independent. \Leftrightarrow$$
$$P(AB) = P(A)P(B)$$
$$P(AC) = P(A)P(C)$$
$$P(BC) = P(B)P(C)$$
$$P(ABC) = P(A)P(B)P(C)$$

♣

例题 1.37 如果 A，B 和 C 彼此独立，A 和 $B \cup C$ 独立吗？

定义 1.9 (重复试验和伯努利试验)

一个大的试验由多个相互独立的子试验构成，且每个子试验都有相同的子样本空间及相同的概率结果。那么这种试验就被称为重复试验（Trials)。

假如重复试验的每个子试验的只有发生或不发生两种互斥的结果，那么这样的重复试验称为伯努利试验（Bernoulli Trials)。

♣

例题 1.38 一个重复 n 次的伯努利试验，成功的概率为 p，不成功的概率为 p，问：

a) 至少成功一次的概率
b) 任意成功 k 次的概率
c) 全部成功的概率
d) 如果将伯努利试验重复无限次，那么全部成功的概率。

例题 1.39 进行独立重复试验，每次试验投掷两个骰子，并记录数字和。问“和为 5”出现在“和为 7”之前的概率。

例题 1.40 氪金游戏 我在开手办盲盒，每次开出手办 i 概率为 p_i，且 $\sum_{i=1}^{n} p_i = 1$。假定每次开盲盒都是独立。现在我已经开出 k 个手办了。A_i 表示“至少有一个手办 i 了”，若 $i \neq j$。问：

a) $P(A_i)$
b) $P(A_i \cup A_j)$
c) $P(A_i|A_j)$

例题 1.41 点数问题 (Problem of points) 两个赌徒按照某种规则赌了起来，并约定赢者通吃，但是赌局在未分胜负时中途停止，问应该如何分配赌资呢？

1654 年，法国的赌徒 $deMéré$ 爵士向数学家 Pascal 提出了这个问题。Pascal 认为赌徒分配赌资的依据应该是赌局进行下去时他们各自获胜的概率。Pascal 解决了一些特例，而后通过与 Fermat 的通讯，两者不仅完全解决了点数问题，还搭建了类似问题的数学框架。因此，有些人把 Pascal 和 Fermat 建立通信的那一天作为概率论的诞生日。

多次独立重复试验，每次成功的概率为 p，失败概率是 1-p。在第 m 次失败前已经成功 n 次的概率。假如每次成功 A 得一分，失败则 B 的一分。如果 A 先得到 n 分 A 就获胜，如果 B 先得到 m 分 B 就获胜。问 A 获胜的概率有多大？

例题 1.42 游戏有 r 个玩家，每个玩家 i 手里有 n_i 单位的赌资。赌局的每个阶段随机选两个人对局，胜者从败者处拿走 1 单位赌资。如果某玩家输光所有赌资就出局，直到仅剩一人时游戏结束，他就是最终获胜者。假设每场对局都是独立的，且玩家实力势均力敌，即每场对局两个玩家均有等概率击败对方。问玩家 i 获胜的概率。

1.7 $P(\cdot|F)$ 同样是概率

定理 1.8

条件概率 $P(\cdot|F)$ 也满足概率三公理。

例题 1.43

a) 保险公司把客户分成两类，一类为容易出事故者，另一类为安全者。第一类一年内出事故的概率为 0.4，而另一类为 0.2。假如全部人口中容易出事故者占 30%。现在有一新客户来购买保险，问这个人在购买保险后一年内出事故的概率为多少。
b) 现在某客户在第一年已经出了一次事故，问他在第二年再出事故的概率。

定义 1.10 (条件独立)

如果在 B 已经发生的条件下，A_1 发生的概率不因 A_2 是否发生而改变，即在给定 B 发生的条件下 A_1 和 A_2 独立，我们就称给定 B 发生条件下 A_1 和 A_2 条件独立（conditionally independent）。

$$P(A_1|A_2B) = P(A_1|B)\ or,\ equivalently\ P(A_1A_2|B) = P(A_1|B)P(A_2)|B$$

例题 1.44 先验 (prior) 概率和后验 (posterior) 概率

第 2 章 随机变量

2.1 随机变量

定义 2.1 (随机变量)

定义在随机试验样本空间 S 的实值函数 $X = X(\omega)$。

例题 2.1 一次性投掷三枚硬币，记录正面向上的硬币个数 X。写出该随机试验的样本空间，写出一个定义在该样本空间的实值函数，也就是是一个随机变量，并写出该实值函数取值对应的概率。

例题 2.2 坛子里放着标号 1 到 20 的 20 个球，无放回地随机取 3 个。随机变量 X 是选中球的最大数字。

- 写出随机变量 X 的取值
- 每个对应取值的概率
- 求 $P\{X > 10\}$

例题 2.3 投掷一枚不均匀的硬币，正面朝上概率 p，背面朝上概率 1-p。现独立重复该试验，直到出现正面，或投掷 n 次为止。X 表示投掷的次数。

- 写出随机变量 X 的取值
- 每个对应取值的概率

2.2 离散随机变量

定义 2.2 (离散随机变量)

- 离散（**discrete**）：随机变量 X 取值至多可数多个，则称为现象为离散，X 称为离散随机变量。
- 至多可数（**at most countable**）：有限（finite）或可数（countable）。
- 可数（**countable**）：能够和自然数集 N 建立一一对应的关系。

定义 2.3 (概率分布、分布律、概率函数，mass function)

$$P\{X = x_i\} = p(x_i) = p_i$$

$$s.t.\ \ 1)\ p(x_i) = p_i \geqslant 0\ for\ i = 1, 2, \cdots$$

$$2)\ \sum_{n=1}^{\infty} p(x_i) = \sum_{n=1}^{\infty} p_i = 1$$

例题 2.4 用表格和图形（直方图）展示的随机变量 X 的概率分布。$p_1 = 0.2$，$p_2 = 0.5$，$p_3 = 0.3$。

定义 2.4 ((累积) 分布函数 (cdf,cumulative distribution function))

$$F(b) = P\{X \leqslant b\}$$

或记为 $X \sim F(X)$、$F_X(x)$。

命题 2.1 (cdf 的性质)

1.(有界性) $\lim_{b\to\infty} F(b) = 1$

2.(有界性) $\lim_{b\to-\infty} F(b) = 0$

3.(单调非减) $If\ x_1 < x_2,\ then\ F(x_1) \leqslant F(X_2)$

4.(右连续性) $F(x_0+) = F(x_0)$

5.$P\{a < X \leqslant b\} = F(b) - F(a)$

例题 2.5 针对例 2.2.1，求 X 的累积分布函数（cdf）和 $P\{X \leqslant 1.5\}, P\{1.5 < X \leqslant 2.5\}, P\{2 \leqslant X \leqslant 3\}, P\{X > 1.5\}$。

例题 2.6 已知 X 的 cdf 为

$$F(X) = \begin{cases} 0 & if\ x < 0, \\ \frac{1}{2} & if\ 0 \leqslant x < 1, \\ \frac{2}{3} & if\ 1 \leqslant x < 2, \\ \frac{11}{12} & if\ 2 \leqslant x < 3, \\ 1 & if\ 3 \leqslant x. \end{cases}$$

求 X 的概率分布和 $P\{X < 3\}, P\{X = 1.5\}, P\{X > \frac{1}{2}\}, P\{2 < X \leqslant 4\}$。

2.3 期望

定义 2.5 (期望 (Expectation) 或期望值 (Expected value))

$$E(X) = \sum_{n=1}^{\infty} x_i p_i = \sum_{n=1}^{\infty} x_i p(x_i)$$

为了简便，$E(X) = EX$。期望是随机变量函数取值的加权平均，权重就是该取值的概率。

♣

例题 2.7 X 表示“投掷一枚骰子的点数”，求 $E(X)$.

例题 2.8

$$I = \begin{cases} 1 & if\ A, \\ 0 & if\ A^c. \end{cases}$$

求 $E(I)$

例题 2.9 120 名同学乘坐 3 辆校车去郊游。第一辆车 36 名同学，第二辆 40 名，第三辆 44 名。到站下车后，我随机找了一名同学，问他“你坐的哪辆车呀”，即问他车上有几名同学。X 表示“这名同学乘坐车辆的人数”，求 $E(X)$，与每辆车上同学数量的平均数比，哪个大呢？

2.4 随机变量函数的期望

定义 2.6 (随机变量函数)

X 是一个随机变量，g(X) 是该随机变量的一个函数。g(X) 也是一个随机变量，也有自己概率分布，也有自己的 $E[g(X)]$。

♣

例题 2.10 随机变量 X 的取值 {-1,0,1}，且 p(-1)=0.2，p(0)=0.5，p(1)=0.3。通过做表求 $E[X^2]$。

定义 2.7 (随机变量函数期望)

$$E[g(X)] = \sum_{n=1}^{\infty} g(x_i)p(x_i)$$

♣

推论 2.1

$$E[aX + b] = aE[X] + b$$

♡

定义 2.8 (n 阶矩 (nth moment of X))

$$E(X^2) = \sum_{n=1}^{\infty} x_i^2 p(x_i)$$

♣

2.5 方差

定义 2.9 (方差 (Variance) 和标准差 (standard deviation))

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

$$SD(X) = \sqrt{Var(X)}$$

♣

推论 2.2

$$D(X) = Var(X) = E(X^2) - (E(X))^2 = EX^2 - (EX)^2$$

♡

例题 2.11 随机变量 X 代表投掷均匀骰子的点数，求 $E[X]$ 和 $Var[X]$。

推论 2.3

$$Var(aX + b) = a^2 Var(X)$$

♡

2.6 常用离散随机变量

定义 2.10 (两点分布、0-1 分布、伯努利 (Bernoulli) 分布)

$$p(x_i) = \begin{cases} p & if\ x_1, \\ 1 - p & if\ x_2. \end{cases}$$

♣

定义 2.11 (二项 (Binomial) 分布)

n 重伯努利试验，每次试验发生的概率为 p，X 表示试验发生的次数。则

$$X \sim B(n, p) \Leftrightarrow P\{X = k\} = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

♣

例题 2.12 投掷 5 枚硬币，硬币相互独立，五枚硬币正面朝上的概率分布。

例题 2.13 某厂生产的螺丝次品率为 0.1，并假设每枚螺丝的质量相互独立。现工厂将 10 个螺丝装一盒，并承诺每盒最多出现一个次品，否则可以退货，问退货率。

例题 2.14 赌徒押注 1 到 6 之间的一个数字，庄家投掷三个骰子，若出现赌徒押的数字一次，则给 1 单位奖励，两次则 2 单位，三次则 3 单位，否则庄家要拿走赌徒 1 单位。问该游戏公平吗？

例题 2.15 某性状由显性基因 d(dominant gene) 和隐形基因 r(recessive gene) 决定。问一对混合型 (hybrid) 基因的父母的四个孩子中，有三个孩子具有显性性状的概率。

例题 2.16 一个通信系统有 n 个元件，超过一半元件工作正常，系统就能正常运行。假设各原件都是独立的，且各元件正常工作的概率为 p。问

1. p 为何值时，5 元件系统比 3 元件系统可靠？
2. 何时 2k+1 个元件的系统比 2k-1 个元件的系统更可靠。

定理 2.1

if $X \sim B(n, p)$, then

$$E(X) = np \quad Var(X) = np(1-p)$$

推论 2.4

如果 $X \sim B(n, p)$, 那么 p(k) 先单调递增，再单调递减，在 $k = \frac{n+1}{p}$ 时取最大值。

定义 2.12 (泊松分布 (Poisson))

$$X \sim P(\lambda) \Leftrightarrow p(k) = P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}$$

例题 2.17 某本书某一页的印刷错误数服从参数为 0.5 的泊松分布。计算该页至少有一处错误的概率。

推论 2.5

当二项分布的 n 足够大，p 足够小，二项分布会趋向 $\lambda = np$ 的泊松分布。

例题 2.18 一台机器生产出的零件的次品率是 0.1，十个这样的零件中至多一个次品的概率。

推论 2.6

if $X \sim P(\lambda)$, then $E(X) = \lambda$, $Var(X) = \lambda$.

定义 2.13 (几何分布 (Geometric distribution))

独立重复试验，直到发生为止，X 表示需要试验的次数。

$$p(k) = P\{X = k\} = (1-p)^{n-1}p \quad with\ E(X) = \frac{1}{p}\ Var(X) = \frac{1-p}{p^2}$$

例题 2.19 坛子里有 N 白 M 黑共（N+M）个球。取出球并放回，直到取出黑球为止。问正好取 n 次和至少取 k 次的概率。

定义 2.14 (负二项分布 (Negative binomial distribution))

独立重复试验，直到成功 r 次为止，X 表示需要试验的次数。

$$p(r) = P\{X = r\} = \binom{n-1}{r-1}p^r(1-p)^{n-r} \quad with\ E(X) = \frac{r}{p}\ Var(X) = \frac{r(1-p)}{p^2}$$

例题 2.20Banach 火柴问题 某烟民有两盒火柴，一盒放左边口袋，一盒放右边口袋，需要火柴时就随机拿一盒并随机取出一根。假设两盒火柴开始时均有 N 根火柴。问当他第一次发现一盒火柴空了的时候，另一盒还有 k 根火柴的概率。

定义 2.15 (超几何分布 (Hypergeometric distribution))

坛子里有 N 个球，其中 M 白，N-M 黑。随机无放回取出 n 个样本。X 表示白球数

$$p(k)=P\{X=k\}=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{n}{k}}\quad with\ E(X)=n\frac{M}{N}\ Var(X)=n\frac{M}{N}(1-\frac{M}{N})(\frac{N-n}{N-1})$$

♣

例题 2.21 假设山里有 N 只大熊猫，但是科学家们并不知道具体数字。于是科学家们先在这座山里抓住 M 只大熊猫，标记后放回。过一段时间后再捉 n 只。X 表示 n 只大熊猫中被标记的只数。若这段实际内大熊猫总数没有变化，每只大熊猫被抓住的概率相等，那么 X 服从超几何分布。这种估计方法被称为极大似然估计 (maximum likelihood estimate)。

2.7 连续随机变量

定义 2.16 (连续随机变量)

$Given\ F(X),\ X\subseteq R,\ \exists f\geqslant 0,\ \forall B\subseteq X\ s.t.$

$$P\{X\in B\}=\int_B f(x)\,dx$$

我们称 X 为连续 (continuous) 随机变量，f 为 X 的概率密度函数 (pdf,probability dense function),F 是 X 的累积分布函数 (cdf,cumulative distribution function)。

♣

命题 2.2

1. $f(x)\geqslant 0$;
2. $\int_{-\infty}^{+\infty} f(x)\,dx=1$;
3. $P\{X=a\}=0$;
4. $\forall\, a\leqslant b,\ P\{a<X\leqslant b\}=P\{a\leqslant X\leqslant b\}=\int_a^b f(x)\,dx=F(b)-F(a)$;
5. $P\{X<a\}=P\{X\leqslant a\}=\int_{-\infty}^a f(x)\,dx=F(a)$;
6. if f is cts at a, then $F'(a)=f(a)$。

♠

例题 2.22

$$f(x)=\begin{cases} kx & if\ 0\leqslant x<3,\\ 2-\frac{x}{2} & if\ 3\leqslant x<4,\\ 0 & otherwise.\end{cases}$$

a) 求 k;
b) 求 F(X);
c) 求 $P\{1<X\leqslant 3.5\}$

例题 2.23

$$F(x)=\begin{cases} 0 & if\ x<0,\\ x^2 & if\ 0\leqslant x\leqslant 1,\\ 1 & if\ 1\leqslant x.\end{cases}$$

a) 求 $P\{0.3<X<0.7\}$;
b) 求 f(x)。

例题 2.24 某电脑死机前连续运行的时长是一个连续随机变量，其密度函数（pdf）为

$$f(x)=\begin{cases} \lambda e^{-x/100}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

a) 求电脑死机前运行 50 至 100 小时的概率；

b) 求电脑运行不到 100 小时就死机了的概率。

例题 2.25 某电子元件寿命为连续随机变量，其密度函数（pdf）为

$$f(x) = \begin{cases} 0, & x \leqslant 100 \\ \dfrac{100}{x^2}, & x > 100 \end{cases}$$

现在某电子设备包含 5 个这样的电子元件，且这些电子元件的寿命相互独立。问 150 小时内，这台电子设备 5 个这样的元件中恰好有 2 个需要更换的概率。

2.8 连续随机变量的期望和方差

定义 2.17 (期望)

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

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

♣

例题 2.26 求 $E(X)$ 和 $E(X^2)$ 和 $Var(X)$，当 X 的密度函数（pdf）为

$$f(x) = \begin{cases} 2x, & 0 \leqslant x \leqslant 1 \\ 0, & otherwise \end{cases}$$

例题 2.27 求 $E(e^X)$，当 X 的密度函数（pdf）为

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & otherwise \end{cases}$$

2.9 常用连续随机变量

定义 2.18 (均匀分布 (Uniformly distribution))

$$f(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & otherwise \end{cases}$$

♣

例题 2.28 求 (α, β) 上均匀分布的 $f(x), F(x), E(X), Var(X)$。

例题 2.29 $X \sim U(0, 10)$，求 $X < 3$，$X > 6$ 和 $3 < X < 8$ 的概率。

例题 2.30 某公交车每 15 分钟一班，例如 7:00、7:15 和 7:30 到站，现在我到车站的时间服从 7 点到 7 点半之间的均匀分布。求

a) 我等公交车的时间不超过 5 分钟；

b) 我等公交车的时间超过了 10 分钟。

定义 2.19 (指数分布 (Exponential distribution))

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

♣

例题 2.31 $X \sim E(\lambda)$，求 $F(x), E(X), Var(X)$。

例题 2.32 某电话通话时长服从参数为 0.1 的指数分布。现在有人在我之前来到电话亭打电话。问

a) 我等待时间超过 10 分钟的概率；

b) 我等待时间在 10 到 20 分钟的概率。

例题 2.33 无记忆性 (memoryless) 你去政务大厅办业务，有两个窗口办理你要办理的业务，但两个窗口分别正在接到张三和李四，他们俩个任一办完业务就轮到你了。现在，每位客户的服务时长都服从参数为 λ 的指数分布。问你是三个人中最后一个办完业务的概率。

线索：指数分布具有无记忆性，即 $P\{X > s+t|X > s\} = P\{X > t\}$

例题 2.34Hazard rate or failure rate 某零件已经使用了 t 小时，求它不能再继续使用 dt 小时的概率。为了方便，我们定义 $\bar{F}(x) = \{X > x\}$。

$$P\{X \in (t, t+dt)|X > t\} = \frac{[P\{X \in (t, t+dt), X > t\}}{[P\{X > t\}} = \frac{[P\{X \in (t, t+dt)\}}{[P\{X > t\}} = \frac{f(t)}{\bar{F}(x)}dt$$

我们定义 $\dfrac{f(t)}{\bar{F}(x)}$ 为 hazard rate $\lambda(x)$。

如果零件寿命服从指数分布，就具有无记忆性。那么对于年龄为 t 的零件，它剩下的寿命和新的完全一样。那么 hazard rate 就是常数 λ。

例题 2.35 瑞利分布 (Rayleigh distribution) $\lambda(t) = a + bt$，其 $F(t) = 1 - e^{-at-bt^2/2}$，$f(t) = (a+bt)e^{-(at+bt^2/2)}$，当 $a = 0$，则是瑞利分布。

例题 2.36 有种说法，各年龄段吸烟者的死亡率是非吸烟者死亡率的两倍。这是什么意思呢？$\lambda_s(t)$ 表示年龄 t 的吸烟者的 hazard rate，$\lambda_n(t)$ 表示年龄 t 的非吸烟者的 hazard rate。$\lambda_s(t) = 2\lambda_s(t)$。根据计算，我们就省略计算过程，结论是一个吸烟者能存活到一定年龄的概率是非吸烟者的相应概率的平方。

定义 2.20 (正态分布 (Normal distribution))

$$X \sim N(\mu, \sigma^2) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$$

♣

命题 2.3

1. 密度函数关于 $x = \mu$ 对称；
2. $x = \mu$，密度函数取得最大值 $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma}$；
3. 密度函数曲线在 $x = \mu \pm \sigma$ 处有拐点，并以 x 轴为渐近线；
4. 固定 σ，只改变 μ, 会使密度函数曲线左右移动，而不改变曲线形状；
5. 固定 μ，只改变 σ, 不会改变密度函数的对称轴，但会改变最大值。
6. if $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

例题 2.37 $X \sim N(\mu, \sigma^2)$，求 $E(X), Var(X)$。

♠

定义 2.21 (标准正态分布 (standard Normal distribution))

$$X \sim N(0,1) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

♣

命题 2.4

1. $\Phi(-x) = 1 - \Phi(x)$
2. if $X \sim N(\mu, \sigma^2)$,then $Z = \dfrac{(X-\mu)}{\sigma} \sim N(0,1)$

♠

例题 2.38 $X \sim N(3, 9)$，查表得到 $P\{2 < X < 5\}, P\{X > 0\}, P\{|X-3| > 6\}$。

例题 2.39 某负心汉被指控为某新生儿的父亲，但他申诉说在孩子出生前 290 天到 240 天之前均在国外，并未与孩子的母亲见面。某医生提供证言称，母亲的孕期近似服从参数为 $\mu = 270$ 和 $\sigma^2 = 100$ 的正态分布。如果你是法官，你怎么看？

推论 2.7 (3σ 准则)

对于正态分布，99.74% 的取值位于 $(\mu-3\sigma,\mu+3\sigma)$，95.44% 的取值位于 $(\mu-2\sigma,\mu+2\sigma)$，68.26% 的取值位于 $(\mu-\sigma,\mu+\sigma)$。

2.10 随机变量函数的分布

例题 2.40 $X\sim U(0,1)$，求 $Y=X^n$ 的分布。

例题 2.41 X 为连续型随机变量，其概率密度为 f_X，求 $Y=X^2$ 的分布。

例题 2.42 X 为连续型随机变量，其概率密度为 f_X，求 $Y=|X|$ 的分布。

定理 2.2

随机变量 X 具有概率密度函数 $f_X(x)$。假设 g(x) 严格单调且可微（g'(x)>0 或 g'(x)<0），因此 g(x) 必定连续，Y=g(X) 为连续型随机变量，其密度函数为

$$f_Y(y)=\begin{cases} f_X[g^{-1}(y)]|\frac{d}{dx}g^{-1}(y)|, & \forall\, x\ s.t.\ y=g(x) \\ 0, & otherwise \end{cases}$$

其中 $g^{-1}(y)$ 为 y=g(x) 的反函数，即满足 y=g(x) 的 x。

第 2 章 随机变量

2.1 随机变量

定义 2.1 (随机变量)

定义在随机试验样本空间 S 的实值函数 $X = X(\omega)$。

例题 2.1 一次性投掷三枚硬币，记录正面向上的硬币个数 X。写出该随机试验的样本空间，写出一个定义在该样本空间的实值函数，也就是一个随机变量，并写出该实值函数取值对应的概率。

例题 2.2 坛子里放着标号 1 到 20 的 20 个球，无放回地随机取 3 个。随机变量 X 是选中球的最大数字。

- 写出随机变量 X 的取值
- 每个对应取值的概率
- 求 $P\{X > 10\}$

例题 2.3 投掷一枚不均匀的硬币，正面朝上概率 p，背面朝上概率 1-p。现独立重复该试验，直到出现正面，或投掷 n 次为止。X 表示投掷的次数。

- 写出随机变量 X 的取值
- 每个对应取值的概率

2.2 离散随机变量

定义 2.2 (离散随机变量)

- 离散（**discrete**）：随机变量 X 取值至多可数多个，则称为现象为离散，X 称为离散随机变量。
- 至多可数（**at most countable**）：有限（finite）或可数（countable）。
- 可数（**countable**）：能够和自然数集 N 建立一一对应的关系。

定义 2.3 (概率分布、分布律、概率函数，mass function)

$$
\begin{aligned}
&P\{X = x_i\} = p(x_i) = p_i \\
s.t.\ \ &1)\ p(x_i) = p_i \geqslant 0\ for\ i = 1, 2, \cdots \\
&2)\ \sum_{n=1}^{\infty} p(x_i) = \sum_{n=1}^{\infty} p_i = 1
\end{aligned}
$$

例题 2.4 用表格和图形（直方图）展示的随机变量 X 的概率分布。$p_1 = 0.2$，$p_2 = 0.5$，$p_3 = 0.3$。

定义 2.4 ((累积) 分布函数 (cdf,cumulative distribution function))

$$F(b) = P\{X \leqslant b\}$$

或记为 $X \sim F(X)$、$F_X(x)$。

命题 2.1 (cdf 的性质)

1.(有界性) $\lim\limits_{b\to\infty} F(b) = 1$

2.(有界性) $\lim\limits_{b\to-\infty} F(b) = 0$

3.(单调非减) $If\ x_1 < x_2,\ then\ F(x_1) \leqslant F(X_2)$

4.(右连续性) $F(x_0+) = F(x_0)$

5.$P\{a < X \leqslant b\} = F(b) - F(a)$

例题 2.5 针对例 2.2.1，求 X 的累积分布函数（cdf）和 $P\{X \leqslant 1.5\}, P\{1.5 < X \leqslant 2.5\}, P\{2 \leqslant X \leqslant 3\}, P\{X > 1.5\}$。

例题 2.6 已知 X 的 cdf 为

$$F(X) = \begin{cases} 0 & if\ x < 0, \\ \frac{1}{2} & if\ 0 \leqslant x < 1, \\ \frac{2}{3} & if\ 1 \leqslant x < 2, \\ \frac{11}{12} & if\ 2 \leqslant x < 3, \\ 1 & if\ 3 \leqslant x. \end{cases}$$

求 X 的概率分布和 $P\{X < 3\}, P\{X = 1.5\}, P\{X > \frac{1}{2}\}, P\{2 < X \leqslant 4\}$。

2.3 期望

定义 2.5 (期望 (Expectation) 或期望值 (Expected value))

$$E(X) = \sum_{n=1}^{\infty} x_i p_i = \sum_{n=1}^{\infty} x_i p(x_i)$$

为了简便，$E(X) = EX$。期望是随机变量函数取值的加权平均，权重就是该取值的概率。♣

例题 2.7 X 表示“投掷一枚骰子的点数”，求 $E(X)$.

例题 2.8

$$I = \begin{cases} 1 & if\ A, \\ 0 & if\ A^c. \end{cases}$$

求 $E(I)$

例题 2.9 120 名同学乘坐 3 辆校车去郊游。第一辆车 36 名同学，第二辆 40 名，第三辆 44 名。到站下车后，我随机找了一名同学，问他“你坐的哪辆车呀”，即问他车上有几名同学。X 表示“这名同学乘坐车辆的人数”，求 $E(X)$，与每辆车上同学数量的平均数比，哪个大呢？

2.4 随机变量函数的期望

定义 2.6 (随机变量函数)

X 是一个随机变量，g(X) 是该随机变量的一个函数。g(X) 也是一个随机变量，也有自己概率分布，也有自己的 $E[g(X)]$。♣

例题 2.10 随机变量 X 的取值 {-1,0,1}，且 p(-1)=0.2，p(0)=0.5，p(1)=0.3。通过做表求 $E[X^2]$。

定义 2.7 (随机变量函数期望)

$$E[g(X)] = \sum_{n=1}^{\infty} g(x_i)p(x_i)$$

推论 2.1

$$E[aX + b] = aE[X] + b$$

定义 2.8 (n 阶矩 (nth moment of X))

$$E(X^2) = \sum_{n=1}^{\infty} x_i^2 p(x_i)$$

2.5 方差

定义 2.9 (方差 (Variance) 和标准差 (standard deviation))

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

$$SD(X) = \sqrt{Var(X)}$$

推论 2.2

$$D(X) = Var(X) = E(X^2) - (E(X))^2 = EX^2 - (EX)^2$$

例题 2.11 随机变量 X 代表投掷均匀骰子的点数，求 $E[X]$ 和 $Var[X]$。

推论 2.3

$$Var(aX + b) = a^2 Var(X)$$

2.6 常用离散随机变量

定义 2.10 (两点分布、0-1 分布、伯努利 (Bernoulli) 分布)

$$p(x_i) = \begin{cases} p & if\ x_1, \\ 1 - p & if\ x_2. \end{cases}$$

定义 2.11 (二项 (Binomial) 分布)

n 重伯努利试验，每次试验发生的概率为 p，X 表示试验发生的次数。则

$$X \sim B(n, p) \Leftrightarrow P\{X = k\} = p(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

例题 2.12 投掷 5 枚硬币，硬币相互独立，五枚硬币正面朝上的概率分布。

例题 2.13 某厂生产的螺丝次品率为 0.1，并假设每枚螺丝的质量相互独立。现工厂将 10 个螺丝装一盒，并承诺每盒最多出现一个次品，否则可以退货，问退货率。

例题 2.14 赌徒押注 1 到 6 之间的一个数字，庄家投掷三个骰子，若出现赌徒押的数字一次，则给 1 单位奖励，两次则 2 单位，三次则 3 单位，否则庄家要拿走赌徒 1 单位。问该游戏公平吗？

例题 2.15 某性状由显性基因 d(dominant gene) 和隐形基因 r(recessive gene) 决定。问一对混合型 (hybrid) 基因的父母的四个孩子中，有三个孩子具有显性性状的概率。

例题 2.16 一个通信系统有 n 个元件，超过一半元件工作正常，系统就能正常运行。假设各原件都是独立的，且各元件正常工作的概率为 p。问

1. p 为何值时，5 元件系统比 3 元件系统可靠？
2. 何时 2k+1 个元件的系统比 2k-1 个元件的系统更可靠。

定理 2.1

if $X \sim B(n, p)$, then

$$E(X) = np \quad Var(X) = np(1-p)$$

推论 2.4

如果 $X \sim B(n, p)$, 那么 p(k) 先单调递增，再单调递减，在 $k = \frac{n+1}{p}$ 时取最大值。

定义 2.12 (泊松分布 (Poisson))

$$X \sim P(\lambda) \Leftrightarrow p(k) = P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}$$

例题 2.17 某本书某一页的印刷错误数服从参数为 0.5 的泊松分布。计算该页至少有一处错误的概率。

推论 2.5

当二项分布的 n 足够大，p 足够小，二项分布会趋向 $\lambda = np$ 的泊松分布。

例题 2.18 一台机器生产出的零件的次品率是 0.1，十个这样的零件中至多一个次品的概率。

推论 2.6

if $X \sim P(\lambda)$, then $E(X) = \lambda$, $Var(X) = \lambda$.

定义 2.13 (几何分布 (Geometric distribution))

独立重复试验，直到发生为止，X 表示需要试验的次数。

$$p(k) = P\{X = k\} = (1-p)^{n-1}p \quad with\ E(X) = \frac{1}{p}\ Var(X) = \frac{1-p}{p^2}$$

例题 2.19 坛子里有 N 白 M 黑共（N+M）个球。取出球并放回，直到取出黑球为止。问正好取 n 次和至少取 k 次的概率。

定义 2.14 (负二项分布 (Negative binomial distribution))

独立重复试验，直到成功 r 次为止，X 表示需要试验的次数。

$$p(r) = P\{X = r\} = \binom{n-1}{r-1}p^r(1-p)^{n-r} \quad with\ E(X) = \frac{r}{p}\ Var(X) = \frac{r(1-p)}{p^2}$$

例题 2.20Banach 火柴问题 某烟民有两盒火柴，一盒放左边口袋，一盒放右边口袋，需要火柴时就随机拿一盒并随机取出一根。假设两盒火柴开始时均有 N 根火柴。问当他第一次发现一盒火柴空了的时候，另一盒还有 k 根火柴的概率。

定义 2.15 (超几何分布 (Hypergeometric distribution))

坛子里有 N 个球，其中 M 白，N-M 黑。随机无放回取出 n 个样本。X 表示白球数

$$p(k) = P\{X = k\} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{n}{k}} \quad with\ E(X) = n\frac{M}{N}\ Var(X) = n\frac{M}{N}(1 - \frac{M}{N})(\frac{N-n}{N-1})$$

♣

例题 2.21 假设山里有 N 只大熊猫，但是科学家们并不知道具体数字。于是科学家们先在这座山里抓住 M 只大熊猫，标记后放回。过一段时间后再捉 n 只。X 表示 n 只大熊猫中被标记的只数。若这段实际内大熊猫总数没有变化，每只大熊猫被抓住的概率相等，那么 X 服从超几何分布。这种估计方法被称为极大似然估计 (maximum likelihood estimate)。

2.7 连续随机变量

定义 2.16 (连续随机变量)

$Given\ F(X),\ X \subseteq R,\ \exists f \geqslant 0,\ \forall B \subseteq X\ s.t.$

$$P\{X \in B\} = \int_B f(x)\,dx$$

我们称 X 为连续 (continuous) 随机变量，f 为 X 的概率密度函数 (pdf,probability dense function),F 是 X 的累积分布函数 (cdf,cumulative distribution function)。

♣

命题 2.2

1. $f(x) \geqslant 0$;
2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$;
3. $P\{X = a\} = 0$;
4. $\forall\ a \leqslant b,\ P\{a < X \leqslant b\} = P\{a \leqslant X \leqslant b\} = \int_a^b f(x)\,dx = F(b) - F(a)$;
5. $P\{X < a\} = P\{X \leqslant a\} = \int_{-\infty}^a f(x)\,dx = F(a)$;
6. if f is cts at a, then $F'(a) = f(a)$。

♠

例题 2.22

$$f(x) = \begin{cases} kx & if\ 0 \leqslant x < 3, \\ 2 - \frac{x}{2} & if\ 3 \leqslant x < 4, \\ 0 & otherwise. \end{cases}$$

a) 求 k;
b) 求 F(X);
c) 求 $P\{1 < X \leqslant 3.5\}$

例题 2.23

$$F(x) = \begin{cases} 0 & if\ x < 0, \\ x^2 & if\ 0 \leqslant x \leqslant 1, \\ 1 & if\ 1 \leqslant x. \end{cases}$$

a) 求 $P\{0.3 < X < 0.7\}$;
b) 求 f(x)。

例题 2.24 某电脑死机前连续运行的时长是一个连续随机变量，其密度函数（pdf）为

$$f(x) = \begin{cases} \lambda e^{-x/100}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

a) 求电脑死机前运行 50 至 100 小时的概率；

b) 求电脑运行不到 100 小时就死机了的概率。

例题 2.25 某电子元件寿命为连续随机变量，其密度函数（pdf）为

$$f(x)=\begin{cases}0, & x\leqslant 100\\ \dfrac{100}{x^2}, & x>100\end{cases}$$

现在某电子设备包含 5 个这样的电子元件，且这些电子元件的寿命相互独立。问 150 小时内，这台电子设备 5 个这样的元件中恰好有 2 个需要更换的概率。

2.8 连续随机变量的期望和方差

定义 2.17 (期望)

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

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

♣

例题 2.26 求 $E(X)$ 和 $E(X^2)$ 和 $Var(X)$，当 X 的密度函数（pdf）为

$$f(x)=\begin{cases}2x, & 0\leqslant x\leqslant 1\\ 0, & otherwise\end{cases}$$

例题 2.27 求 $E(e^X)$，当 X 的密度函数（pdf）为

$$f(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & otherwise\end{cases}$$

2.9 常用连续随机变量

定义 2.18 (均匀分布 (Uniformly distribution))

$$f(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & otherwise\end{cases}$$

♣

例题 2.28 求 (α,β) 上均匀分布的 $f(x),F(x),E(X),Var(X)$。

例题 2.29 $X\sim U(0,10)$，求 $X<3$，$X>6$ 和 $3<X<8$ 的概率。

例题 2.30 某公交车每 15 分钟一班，例如 7:00、7:15 和 7:30 到站，现在我到车站的时间服从 7 点到 7 点半之间的均匀分布。求

a) 我等公交车的时间不超过 5 分钟；

b) 我等公交车的时间超过了 10 分钟。

定义 2.19 (指数分布 (Exponential distribution))

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

♣

例题 2.31 $X\sim E(\lambda)$，求 $F(x),E(X),Var(X)$。

例题 2.32 某电话通话时长服从参数为 0.1 的指数分布。现在有人在我之前来到电话亭打电话。问

a) 我等待时间超过 10 分钟的概率；

b) 我等待时间在 10 到 20 分钟的概率。

例题 2.33 无记忆性 (memoryless) 你去政务大厅办业务，有两个窗口办理你要办理的业务，但两个窗口分别正在接到张三和李四，他们俩个任一办完业务就轮到你了。现在，每位客户的服务时长都服从参数为 λ 的指数分布。问你是三个人中最后一个办完业务的概率。

线索：指数分布具有无记忆性，即 $P\{X > s+t|X > s\} = P\{X > t\}$

例题 2.34Hazard rate or failure rate 某零件已经使用了 t 小时，求它不能再继续使用 dt 小时的概率。为了方便，我们定义 $\bar{F}(x) = \{X > x\}$。

$$P\{X \in (t, t+dt)|X > t\} = \frac{[P\{X \in (t, t+dt), X > t\}}{[P\{X > t\}} = \frac{[P\{X \in (t, t+dt)\}}{[P\{X > t\}} = \frac{f(t)}{\bar{F}(x)} dt$$

我们定义 $\frac{f(t)}{\bar{F}(x)}$ 为 hazard rate $\lambda(x)$。

如果零件寿命服从指数分布，就具有无记忆性。那么对于年龄为 t 的零件，它剩下的寿命和新的完全一样。那么 hazard rate 就是常数 λ。

例题 2.35 瑞利分布 (Rayleigh distribution) $\lambda(t) = a + bt$，其 $F(t) = 1 - e^{-at-bt^2/2}$，$f(t) = (a+bt)e^{-(at+bt^2/2)}$，当 $a = 0$，则是瑞利分布。

例题 2.36 有种说法，各年龄段吸烟者的死亡率是非吸烟者死亡率的两倍。这是什么意思呢？$\lambda_s(t)$ 表示年龄 t 的吸烟者的 hazard rate，$\lambda_n(t)$ 表示年龄 t 的非吸烟者的 hazard rate。$\lambda_s(t) = 2\lambda_s(t)$。根据计算，我们就省略计算过程，结论是一个吸烟者能存活到一定年龄的概率是非吸烟者的相应概率的平方。

定义 2.20 (正态分布 (Normal distribution))

$$X \sim N(\mu, \sigma^2) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

♣

命题 2.3

1. 密度函数关于 $x = \mu$ 对称；
2. $x = \mu$，密度函数取得最大值 $f(x) = \frac{1}{\sqrt{2\pi}\sigma}$；
3. 密度函数曲线在 $x = \mu \pm \sigma$ 处有拐点，并以 x 轴为渐近线；
4. 固定 σ，只改变 μ，会使密度函数曲线左右移动，而不改变曲线形状；
5. 固定 μ，只改变 σ，不会改变密度函数的对称轴，但会改变最大值。
6. if $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

例题 **2.37** $X \sim N(\mu, \sigma^2)$，求 $E(X), Var(X)$。

♠

定义 2.21 (标准正态分布 (standard Normal distribution))

$$X \sim N(0, 1) \Leftrightarrow f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

♣

命题 2.4

1. $\Phi(-x) = 1 - \Phi(x)$
2. if $X \sim N(\mu, \sigma^2)$,then $Z = \frac{(X-\mu)}{\sigma} \sim N(0, 1)$

♠

例题 2.38 $X \sim N(3, 9)$，查表得到 $P\{2 < X < 5\}, P\{X > 0\}, P\{|X-3| > 6\}$。

例题 2.39 某负心汉被指控为某新生儿的父亲，但他申诉说在孩子出生前 290 天到 240 天之前均在国外，并未与孩子的母亲见面。某医生提供证言称，母亲的孕期近似服从参数为 $\mu = 270$ 和 $\sigma^2 = 100$ 的正态分布。如果你是法官，你怎么看？

推论 2.7 (3σ 准则)

对于正态分布，99.74% 的取值位于 $(\mu-3\sigma,\mu+3\sigma)$，95.44% 的取值位于 $(\mu-2\sigma,\mu+2\sigma)$，68.26% 的取值位于 $(\mu-\sigma,\mu+\sigma)$。

2.10 随机变量函数的分布

例题 2.40 $X \sim U(0,1)$，求 $Y=X^n$ 的分布。

例题 2.41 X 为连续型随机变量，其概率密度为 f_X，求 $Y=X^2$ 的分布。

例题 2.42 X 为连续型随机变量，其概率密度为 f_X，求 $Y=|X|$ 的分布。

定理 2.2

随机变量 X 具有概率密度函数 $f_X(x)$。假设 g(x) 严格单调且可微（g'(x)>0 或 g'(x)<0），因此 g(x) 必定连续，Y=g(X) 为连续型随机变量，其密度函数为

$$f_Y(y)=\begin{cases} f_X[g^{-1}(y)]|\frac{d}{dx}g^{-1}(y)|, & \forall\, x\ s.t.\ y=g(x) \\ 0, & otherwise \end{cases}$$

其中 $g^{-1}(y)$ 为 y=g(x) 的反函数，即满足 y=g(x) 的 x。

第 3 章 多维随机变量

3.1 联合分布函数

定义 3.1 (联合分布函数 (Jointly distribution function))

$$F(a,b) = P\{X \leqslant a, Y \leqslant b\}$$

命题 3.1

1. $F(x,y)$ is nondecreasing for x and y;
2. $0 \leqslant F(x,y) \leqslant 1$;
3. $F(-\infty,y) = 0, F(x,-\infty) = 0$;
4. $F(x,y)$ is right continuous both for x and y;
5. $P\{a_1 < X \leqslant a_2, b_1 < Y \leqslant b_2\} \geqslant 0$。

定义 3.2 (边缘分布或边际分布 (Marginal distribution))

$$F_X(a) = P\{X \leqslant a\} = P\{X \leqslant a, Y \leqslant \infty\} \equiv F(a,\infty)$$

$$F_Y(b) = P\{Y \leqslant b\} = P\{X \leqslant \infty, Y \leqslant b\} \equiv F(\infty,b)$$

定义 3.3 (二维离散随机变量)

当 X 和 Y 均为离散随机变量时，(X,Y) 为二维离散随机变量。

X 和 Y 的联合概率分布律 (Joint probability mass function) 定义为

$$p(x,y) = P\{X = x, Y = y\}$$

X 和 Y 的边缘概率分布律 (marginal probability mass function) 分别为

$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x,y)$$

$$p_Y(y) = P\{Y = y\} = \sum_{x:p(x,y)>0} p(x,y)$$

例题 3.1 坛子里有 3 红 4 白 5 蓝共 12 个球，从中随机取 3 个。X 和 Y 分别表示取出的红球和白球数。做表展示 X 和 Y 的联合分布律和边缘分布律。

定义 3.4 (二维连续随机变量)

$$Given\ F(X,Y)\ of\ (X,Y), \exists\, f(x,y) \geqslant 0,\ \forall\, x,\ y,\ s.t.$$

$$F(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)\, dy\, dx$$

称 (X,Y) 为二维连续型随机变量，f 为 (X,Y) 为概率密度函数，或 X 和 Y 的联合密度函数 (Joint probability density function)。X 和 Y 的边缘概率密度函数 (marginal probability density function) 分别为

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

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

并且

$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$$

♣

例题 3.2

$$f(x,y) = \begin{cases} 2e^{-(x+2y)}, & x>0, y>0 \\ 0, & otherwise \end{cases}$$

a) 求 F(x,y)
b) 求 $P\{X>1, Y<1\}$
c) 求 $P\{X<Y\}$
d) 求 $P\{X<a\}$

例题 3.3

$$f(x,y) = \begin{cases} cy(2-x), & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant x \\ 0, & otherwise \end{cases}$$

a) 求 c
b) 求两个边缘密度函数

3.2 随机变量的条件分布与独立性

定义 3.5 (离散随机变量的条件分布)

在给定 Y=y 的条件下，X 的条件概率分布律 (conditional probability mass function) 为

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$

在给定 Y=y 的条件下，X 的条件概率分布函数 (conditional probability distribution function) 为

$$F_{X|Y}(x|y) = P\{X \leqslant x | Yy\} = \sum_{a \leqslant x} p_{X|Y}(a|y)$$

♣

定理 3.1

$X, Y\ independent \Leftrightarrow p_{X|Y}(x|y) = P\{X = x\}$

♡

例题 3.4 X 和 Y 的联合概率分布律 p(x,y) 符合

$$p(0,0) = 0.4, p(0,1) = 0.2, p(1,0) = 0.1, p(1,1) = 0.3$$

给定 Y=1，求 X 的条件概率分布律。

例题 3.5 X 和 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 的分布。

定义 3.6 (连续随机变量的条件分布)

在给定 Y=y 的条件下，X 的条件概率密度函数 (conditional probability density function) 为

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

在给定 Y=y 的条件下，X 的条件概率分布函数 (conditional probability distribution function) 为

$$F_{X|Y}(a|y) = P\{X \leqslant a | Y = y\} = \int_{-\infty}^{a} f_{X|Y}(x|y)\,dx$$

♣

定理 3.2

$X, Y\ independent \Leftrightarrow f_{X|Y}(x|y) = f_X(x)$

例题 3.6 随机变量 (X,Y) 的联合概率密度函数为

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y), & 0<x<1, 0<y<1 \\ 0, & otherwise \end{cases}$$

求已知 Y=y 的条件下，X 的条件密度。

定义 3.7 (随机变量的独立性)

随机变量 X 和 Y 相互独立 (independent)，iff

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

命题 3.2

1. $P\{X \leqslant a, Y \leqslant b\} = P\{X \leqslant a\}P\{Y \leqslant b\}$，即 $F(a,b) = F_X(a)F_Y(b)$；
2. 当 X 和 Y 是离散型随机变量，$p(x,y) = p_X(x)p_Y(y)$;
3. 当 X 和 Y 是连续型随机变量，$f(x,y) = f_X(x)f_Y(y)$;

例题 3.7 从 (0,1) 中任取两个数，分别记为 X 和 Y，求两数之和小于 1.2 的概率，求两数之积小于 1/4 的概率。

例题 3.8 两人约定在某地见面，且到达时间相互独立，且服从中午 12 点到下午 1 点的均匀分布，求先到的人等候十分钟以上的概率。

例题 3.9Buffon 投针问题 桌面上有一些距离为 D 的平行线，向桌面随机投掷长度为 L（L⩽D）。问该针与平行线相交的概率。

例题 3.10 随机变量 (X,Y) 的概率密度函数为

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

a) 求 X 和 Y 的边缘概率密度函数，并判断 X 和 Y 的独立性

b) 给定 Y=y, 求 X 的条件概率密度

例题 3.11

$$f(x,y) = \begin{cases} 8xy, & 0 \leqslant x \leqslant y \leqslant 1 \\ 0, & otherwise \end{cases}$$

问 X 和 Y 是否相互独立。

例题 3.12

$$f(x,y) = \begin{cases} 6xy^2, & 0<x<1, 0<y<1 \\ 0, & otherwise \end{cases}$$

问 X 和 Y 是否相互独立。

定理 3.3

$X, Y\ independent \Leftrightarrow f_{X,Y}(x,y) = h(x)g(y)$

3.3 多维随机变量函数的分布

3.3.1 Z=X+Y 的分布

定理 3.4 (卷积公式)

若 X 和 Y 是两个相互独立的离散型随机变量，其密度函数分别为 $p_X(x)$ 和 $p_Y(y)$，则其和 Z=X+Y 的分布律为

$$P(Z=k)=\sum_{x+y=k} p_x(x)p_Y(y)=\sum_{x=0}^{k} p_x(x)p_Y(k-x)$$

若 X 和 Y 是两个相互独立的连续型随机变量，其密度函数分别为 $f_X(x)$ 和 $f_Y(y)$，则其和 Z=X+Y 的密度函数为

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

例题 3.13 可加性：同一类独立随机变量和的分布仍属于此类分布。

泊松分布的可加性 $X\sim P(\lambda_1)$，$Y\sim P(\lambda_2)$，且 X 和 Y 相互独立，证明 $Z=X+Y\sim P(\lambda_1+\lambda_2)$。

二项分布的可加性 $X\sim B(n,p)$，$Y\sim B(m,p)$，且 X 和 Y 相互独立，则 $Z=X+Y\sim B(n+m,p)$

正态分布的可加性 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且 X 和 Y 相互独立，则 $Z=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

3.3.2 最大值和最小值的分布

定理 3.5

若 X 和 Y 是两个相互独立的随机变量，其累积分布函数分别为 $F_X(x)$ 和 $F_Y(y)$，随机变量 $M=max\{X,Y\}$，$N=min\{X,Y\}$，则

$$F_M(z)=P(M\leqslant z)=P(X\leqslant z,Y\leqslant z)=F_X(x)F_Y(y)$$

$$F_N(z)=P(N\leqslant z)=1-P(N>z)=1-P(X>z,Y>z)==1-[1-F_X(x)][1-F_Y(y)]$$

3.3.3 变量变换法（Change-of-variable Method）

二维随机变量 (Y_1,Y_2) 的联合密度函数为 $f(y_1,y_2)$，如果

$$\begin{cases} y_1=g_1(x_1,x_2),\\ y_2=g_2(x_1,x_2)\end{cases}$$

有连续偏导数，且存在唯一的反函数

$$\begin{cases} x_1=h_1(y_1,y_2),\\ x_2=h_2(y_1,y_2)\end{cases}$$

该变换的 Jacobian 为

$$J=\frac{\partial(g_1,g_2)}{\partial(x_1,x_2)}=\begin{vmatrix}\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2}\\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2}\end{vmatrix}=\left(\frac{\partial(x_1,x_2)}{\partial(g_1,g_2)}\right)^{-1}\neq 0$$

那么：

$$f_{Y_1,Y_2}(y_1,y_2)=f_{X_1,X_2}(x_1,x_2)|J^{-1}|$$

例题 3.14 X_1 和 X_2 独立同分布，求 Y_1 和 Y_2 的联合密度函数。

$$\begin{cases} Y_1 = X_1 + X_2, \\ Y_2 = X_1 - X_2 \end{cases}$$

3.3.4 Z=XY 的分布

定理 3.6

若 X 和 Y 独立，密度分别为 $f_X(x)$ 和 $f_Y(y)$，则 $Z = XY$ 的密度函数为：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(\frac{z}{y}) f_Y(y) |\frac{1}{y}| \, dy$$

更一般地，若 X 和 Y 不独立：

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, \frac{z}{x}) |\frac{1}{x}| \, dx$$

3.3.5 Z=X/Y 的分布

定理 3.7

若 X 和 Y 独立，密度分别为 $f_X(x)$ 和 $f_Y(y)$，则 $Z = X/Y$ 的密度函数为：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(zy) f_Y(y) |y| \, dy$$

更一般地，若 X 和 Y 不独立：

$$f_Z(z) = \int_{-\infty}^{\infty} f(zy, y) |y| \, dy$$

3.4 多维随机变量的数字特征

3.4.1 期望、方差

定理 3.8 (多维随机变量的期望)

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) p(x,y), & discrete \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) \, dx \, dy, & cts \end{cases}$$

- 当 $g(X,Y) = X$，则可以得到 X 的期望:

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

- 当 $g(X,Y) = [X - E(X)]^2$，则可以得到 X 的方差:

$$\begin{aligned} Var(X) = E[X - E(X)]^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))^2 f(x,y) \, dy \, dx \\ &= \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) \, dx \end{aligned}$$

例题 3.15 在长度为 a 的线段上任取两点 X 与 Y，求两点的平均距离。

命题 3.3

- $E(X+Y) = E(X) + X(Y)$
- 当 X 和 Y 相互独立，则 $E(XY) = E(X)X(Y)$
- 当 X 和 Y 相互独立，则 $Var(X \pm Y) = Var(X) + Var(Y)$

3.4.2 协方差

定义 3.8 (协方差)

对于二维随机变量 (X,Y) 的 E[(X-EX)(Y-EY)] 存在，则称其为 X 与 Y 的协方差 **(covariance)**、相关中心矩或混合中心矩，记为

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$$

命题 3.4

- $Cov(X,X) = Var(X)$
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX,bY) = abCov(X,Y)$
- $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$
- 当 X 和 Y 相互独立，则 $Cov(X,Y) = 0$；其逆命题不成立。
- $Cov(X,Y) = E(XY) - E(X)E(Y)$
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$

3.4.3 相关系数

定义 3.9 (相关系数（Correlated coeffecient）)

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{Cov(X,X)}{\sigma_X \sigma_Y}$$

命题 3.5

- $|\rho_{XY}| \leqslant 1$
- 当 X 和 Y 相互独立，则 X 和 Y 不相关。
- $If\ \exists a(\neq 0), b\ s.t.\ P(Y = aX + b) = 1,\ then\ |\rho_{XY}| = 1$。$\rho_{XY} = 1\ as\ a > 0$，$\rho_{XY} = -1\ as\ a < 0$
  - ρ_{XY} 刻画了 X 和 Y 之间的线性关系强弱，因此又被成为“线性相关系数”；
  - $\rho_{XY} = 0$ 说明两者线性不相关，但不能排除其他关系；
  - $0 < |\rho_{XY}| < 1$ 说明 X 和 Y 有一定线性相关关系。越接近 1，两者线性相关程度越高。而协方差没有这个特点。
- $e = E[Y - (aX+b)]^2$ 称为用 aX+b 近似 Y 的均方误差（MSE，Mean-Square Error）。

定义 3.10 (协方差矩阵)

n 维随机向量 $\mathbf{X} = (X_1, X_2, \cdots, X_n)'$，若每个分量的数学期望都存在，则称

$$E(\mathbf{X}) = (E(X_1), E(X_2), \cdots, E(X_n))'$$

为随机向量 $\mathbf{X}$ 的期望向量，简称 $\mathbf{X}$ 的期望。称

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

称为随机向量 $\mathbf{X}$ 的方差-协方差矩阵或协方差阵，记为 $Cov(\mathbf{X})$。

定理 3.9
$Cov(\mathbf{X})$ 是一个对称的非负定矩阵。

3.5 多维正态分布

定义 3.11 (二维正态分布)
$(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{q}{2}}$$

with $\sigma_1 > 0,\ \sigma_2 > 0,\ -1 < \rho < 1,\ and$

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

命题 3.6 (二维正态分布的性质)

- f(x,y) is joint pdf;
- $X \sim N(\mu_1, \sigma_1)$，$Y \sim N(\mu_2, \sigma_2)$;
- ρ is the correlated coeffecient of X and Y;
- 边缘分布相同的多维联合分布可以是不同的。

定义 3.12 (多维正态分布的矩阵表示)

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \qquad \mu = \begin{pmatrix} EX_1 \\ EX_2 \end{pmatrix} \qquad \mathbf{C} = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

$$q = (\mathbf{X} - \mu)'\mathbf{C}^{-1}(\mathbf{X} - \mu)$$

二维正态分布可以记为：

$$f(x, y) = \frac{1}{(2\pi)^{2/2}|\mathbf{C}|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{X} - \mu)'\mathbf{C}^{-1}(\mathbf{X} - \mu)\right\}$$

n 维正态分布可以记为：

$$f(x, y) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{X} - \mu)'\mathbf{C}^{-1}(\mathbf{X} - \mu)\right\}$$

第 4 章 数字特征和大数定律

4.1 数学期望

定义 4.1 (数学期望)

$$E(X) = \sum_x xp(x)\ , \quad when\ X\ is\ discrete\ random\ variable$$

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx, \quad when\ X\ is\ continuous\ random\ variable$$

定理 4.1

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)\ , \quad for\ mass\ function\ p(x,y)$$

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

命题 4.1 (数学期望的性质)

1) $E[aX+b] = aE[X]+b$
   - $E[aX] = aE[X]$
   - $E[X+b] = E[X]+b$
2) $E[X+Y] = E[X]+E[Y]$
   - 推广：$E[X_1+X_2+\cdots+X_n] = E[X_1]+E[X_2]+\cdots+E[X_n]$
3) If X and Y are independent, then $E[XY] = E[X]E[Y]$.
   - 推广 1：If $X_1, X_2, \cdots, X_n$ are independent, then $E[X_1X_2\cdots X_n] = E[X_1]E[X_2]\cdots E[X_n]$.
   - 推广 2：If X and Y are independent, then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

例题 4.1 在长度为 L 的一段路上，在 X 点发生了一起车祸，在 Y 点有一辆救护车，X 和 Y 均匀地分布在这段路上，并且相互独立。问救护车距离车祸地点的平均距离。

例题 4.2boole's inequality

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

例题 4.3 詹森不等式（Jensen's inequality） 如果 $g(x)$ 为凹函数，例如 $g(x) = ln(x)$，那么下面的不等式成立

$$E[g(X)] \leqslant g[E(X)]$$

例题 4.4 二项分布、负二项分布和超几何分布 求二项分布、负二项分布和超几何分布的期望。

例题 4.5 配对问题 N 个人扔帽子，充分混合后，每人随机拿一顶帽子，求选中自己帽子人数的期望值。

例题 4.6 点券收集问题 收集点券，共有 N 种不同的点券，我每次只能收集到一张点券，且每次的点券在 N 种点券中均匀分布。问收集全套 N 种点券时，我收集的点券总数的期望。

例题 4.7 10 个猎人打猎，现在有 10 只野鸭同时起飞，10 个猎人同时开枪随机瞄准并射击一只野鸭。如果猎人击中的概率为 p，问逃过这一劫的野鸭数量的期望。

4.2 方差、协方差和相关系数

定义 4.2 (方差)

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

$$Var(X) = E(X^2) - (EX)^2$$

定义 4.3 (协方差)

$$Cov(X,Y) = E[(X - EX)(Y - EY)]$$

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

命题 4.2 (方差性质)

1) $Var[aX + b] = a^2Var[X]$
   - $Var[aX] = a^2Var[X]$
   - $Var[X + b] = Var[X]$
2) $Var[X \pm Y] = Var[X] + Var[Y] \pm 2Cov(X,Y)$
   - If X and Y are independent, then $Var[X \pm Y] = Var[X] + Var[Y]$

命题 4.3 (协方差性质)

1) $Cov(X,X) = Var(X)$
2) $Cov(X,Y) = Cov(Y,X)$
3) $Cov(aX,bY) = abCov(X,Y)$
4) $Cov(C,X) = 0, C is a constant$
5) $Cov(X + Y, Z) = Cov(X,Z) + Cov(Y,Z)$
6) if X and Y are independent, then $Cov(X,Y) = 0$；其逆命题不成立。

例题 4.8 计算 $Cov(aW + bX, cY + dZ)$。

推论 4.1

协方差的性质有如下的推广

$$U = a + \sum_{i=1}^{n} b_i X_i,\ V = c + \sum_{j=1}^{m} d_j Y_j,\ Cov(V,U) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i b_j Cov(X_i, Y_j)$$

$$Var\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_j Cov(X_i, X_j)$$

定义 4.4 (相关系数)

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

命题 4.4 (相关系数性质)

1) $|\rho(X,Y)| \leqslant 1$
2) $\rho(X,Y) = 0$，称为不相关 (uncorrelated)。$independent \Rightarrow uncorrelated$。

3) $\rho(X,Y)=1 \Leftrightarrow Y=aX+b$。$If\ a>0,\ then\ \rho(X,Y)=1, if\ a<0,\ then\ \rho(X,Y)=-1$。

4.3 条件期望

定义 4.5 (条件期望)

$$E[X|Y=y]=\sum_x xP\{X=x|Y=y\}=\sum_x xp_{X|Y}(x|y)$$

$$E[X|Y=y]=\sum_x xP\{X=x|Y=y\}=\int_{-\infty}^{+\infty} xf_{X|Y}(x|y)\,dx$$

例题 4.9

$$f(x,y)=\frac{e^{-x/y}e^{-y}}{y} \qquad -\infty<x<\infty, \infty<y<\infty$$

求 E(X|Y=y)

定理 4.2

$$E[X]=E[E(X|Y)]=\begin{cases} \sum_y E[X|Y=y]p\{Y=y\} & for\ discrete\ case, \\ \int_{-\infty}^{+\infty} E[X|Y=y]f_Y(y)\,dy & for\ continuous\ case. \end{cases}$$

例题 4.10 一个矿工迷路了，他面前有三个条路。现在已知，第一个条路要经过 3 小时才能到达安全地点，第二路要经过 5 小时后会回到原地，第三条了要经过 7 小时会回到原地。假如迷路的矿工每次都是随机算一条路，问这个矿工到达安全地点平均需要几小时。

例题 4.11Crap 游戏 每次投两个筛子，点数之和为 2,3 或 12 则玩家输，7 或 11 则玩家赢。若为其他点数 i 则要一直玩下去，一直到投出 7 或者 i 为止。若最后为 7 则玩家输，若最后为 i 这玩家赢。记 R 为投骰子的次数。求：E(R)，E(R| 玩家赢) 和 E(R| 玩家输)。

定义 4.6 (条件方差)

$$Var(X|Y)\equiv E[(X-E(X|Y))^2|Y]=E[X^2|Y]-(E(X|Y))^2$$

定理 4.3

$$Var(X)=E[Var(X|Y)]+Var(E[X|Y])$$

例题 4.12 在 (0,t) 时段中，达到车站的人数是一个服从泊松分布的随机变量均值为 λt。地铁在时段 (0,T) 时段内随机达到，即服从 (0,T) 上的均匀分布。且地铁到达与候车人数独立。求地铁到站后上车人数的期望和方差。

4.4 其他数字特征

定义 4.7 (矩 (Moments))

$$kth\ Moments:\ E(X^k)=\int_{-\infty}^{+\infty} x^k f(x)\,dx$$

$$kth\ Centered\ Moments:\ E[(X-EX)^k]=\int_{-\infty}^{+\infty}(x-\mu)^k f(x)\,dx$$

定义 4.8 (矩母函数、矩生产函数 (Moments Generating Function,MGF))

$$M(t) = E(e_t X) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx$$

命题 4.5 (矩母函数性质)

1) $M(0) = 1$
2) $M^{(k)}(0) = E(X^k)$

定义 4.9 (α 分位数 (α quantitle))

$$P\{X \leqslant x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)\, dx = \alpha$$

我们称为 x_α 为 α 分位数。

定义 4.10 (中位数 (median))

$$P\{X \leqslant x_{\frac{1}{2}}\} = \int_{-\infty}^{x_{\frac{1}{2}}} f(x)\, dx = \frac{1}{2}$$

我们称为 $x_{\frac{1}{2}}$ 为 $\frac{1}{2}$ 分位数，或者中位数。

定义 4.11 (偏度系数 (Skewness))

我们用 v_k 表示 k 阶矩。

$$Skewness = \frac{v_3}{v_2^{3/2}} = \frac{E(X - EX)^3}{SD(X)^3}$$

为正，称为右偏，说明有大极端值；为负，称为左偏，说明有小极端值。

定义 4.12 (峰度系数 (Kurtosis)

$$Kurtosis = \frac{v_4}{v_2^2} - 3$$

正态分布峰度为 0,；为正，表示比正态分布更尖，尾部更粗；为负，比正态分布更平，尾部更细。

4.5 大数定律与中心极限定理

定理 4.4 (马尔科夫不等式 (Markov's inequality))

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}, \quad \forall a > 0$$

定理 4.5 (切比雪夫不等式 (Chebyshev's inequality))

$$P\{|X - \mu| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}, \quad \forall \varepsilon > 0$$

例题 4.13 若 Var(X)=0，则 $P\{X = E[X]\} = 1$。

定理 4.6 (弱大数定律 (The weak law of large numbers))

Let $X_1, X_2, \cdots X_n$ are iid with $E(X_i) = \mu$.Then $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\{|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| \geqslant \varepsilon\} = 0$$

♡

定理 4.7 (伯努利定理 (Bernoulli's theorem))

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

♡

定理 4.8 (中心极限定理 (The central limit theorem))

Let $X_1, X_2, \cdots X_n$ are iid with $E(X_i) = \mu\ and\ Var(X_i) = \sigma^2$.Then $\forall a \in R$,

$$\lim_{n\to\infty} P\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leqslant a\} = \int_{-\infty}^{a} e^{x^2/2}\, dx$$

♡

定理 4.9 (强大数定律 (The strong law of large numbers))

Let $X_1, X_2, \cdots X_n$ are iid with $E(X_i) = \mu$.Then,

$$\lim_{n\to\infty} P\{\frac{X_1 + X_2 + \cdots + X_n}{n}\} = \mu$$

♡