概率论与数理统计笔记

恰冷饭菌

2020 年 11 月 21 日

目录

第一章　事件与概率

1.1　概率论的基本概念

1.1.1　随机试验与随机事件

定义 1.1 基本事件是随机试验中的每一个不可再分的单一结果.

定义 1.2 随机事件 (event) 是 $n(n \geq 1)$ 个基本事件组成的结果.

定义 1.3 样本空间 (sample space,Ω) 是随机试验中所有基本事件（样本点，ω）构成的集合.

因而，一个事件是 Ω 的子集. 根据不同的试验目的，样本空间应该予以不同的选择.

定义 1.4 对于一个事件 A：
(1) 若 $A = \Omega$，则 A 为必然事件.
(2) 若 $A = \varnothing$，则 A 为不可能事件.

1.1.2　事件的运算

定义 1.5 $A \subset B$：A 是 B 的子事件，A 发生蕴含着 B 一定发生.

定义 1.6 若 $A \subset B$ 且 $A \supset B$，则 $A = B$.

定义 1.7 $A \cup B$：事件的和，其含义为至少有一个事件发生. 若 A, B 互斥，则简记为 $A + B$.

定义 1.8 $A \cap B$：事件的积，其含义为两个事件同时发生. 简记为 AB.

定义 1.9 若 $A \cap B = \varnothing$，则称 A, B 互斥（不相容）.

定义 1.10 $\overline{A}(A^c)$：对立事件（余事件），其含义为该事件不发生.

余事件可以理解为集合对全集的补集.

定义 1.11 $A - B$：差事件，其含义为 A 发生且 B 不发生. 显然 $A - B = A\overline{B}$.

定义 1.12 $A \triangle B$：对称差，其含义为 A, B 不同时发生. 显然 $A \triangle B = (A \cup B) - (A \cap B)$.

定理 1.1 (De Morgan 对偶法则)

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i} \tag{1.1.1}$$

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i} \tag{1.1.2}$$

例 1.1 表示以下事件：

(1)$A_1, A_2, \ldots, A_n$ 至少发生两个.

(2)$A_1, A_2, \ldots, A_n$ 至多发生一个.

(3)$A_1, A_2, \ldots, A_n$ 刚好发生 k 个.

解 (1) $\bigcup\limits_{1 \leq i,j \leq n, i \neq j} (A_i A_j)$.

(2) $\overline{\bigcup\limits_{1 \leq i,j \leq n, i \neq j} (A_i A_j)} = \overline{\bigcup\limits_{1 \leq i,j \leq n, i \neq j} (A_i \cap A_j)} = \bigcap\limits_{1 \leq i,j \leq n, i \neq j} \overline{(A_i \cap A_j)} = \bigcap\limits_{1 \leq i,j \leq n, i \neq j} \left(\overline{A_i} \cup \overline{A_j}\right)$.
此外还可以直接根据定义，$\bigcup\limits_{1 \leq i,j \leq n, i \neq j} \left(\overline{A_i} \cap \overline{A_j}\right)$. 代入 $n = 3$ 的情况画 Venn 图可以发现这两种表述的确是等价的.

(3) $\bigcup\limits_{1 \leq a_1, a_2, \ldots, a_n \leq n, a_1 \neq a_2 \neq \cdots \neq a_n} \left(A_{a_1} A_{a_2} \ldots A_{a_k} \overline{A_{a_{k+1}}} \ldots \overline{A_{a_n}}\right)$.

1.1.3　概率的定义及性质

定义 1.13 (古典概型) 记 $\Omega = \omega_1, \omega_2, \ldots, \omega_n, n \in \mathbb{N}_+$，且 $P(\omega_i) = \dfrac{1}{n}$，则 $P(\omega_{a_1}, \omega_{a_2}, \ldots, \omega_{a_k}) = \dfrac{k}{n}(0 \leq k \leq n)$.

定义 1.14 以下是概率的公理化定义：

(1) 随机事件 A 满足 $0 \leq P(A) \leq 1$.

(2)$P(\Omega) = 1$.

(3) 若 $A \cap B = \varnothing$，则 $P(A \cup B) = P(A) + P(B)$.

上述定义的 (3) 推广到无穷情况仍然有效（当 $A_i \cap A_j = \varnothing(i \neq j)$）：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \tag{1.1.3}$$

对于样本空间内所有可以计算概率的事件的集合 $\mathscr{F}$(集合的集合)，再通过定义加法和乘法，不难验证这是 $\sigma-$ 代数. 它有以下性质：

定理 1.2 (容斥原理)$\forall A_1, A_2, \ldots, A_n \in \mathscr{F}$：

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P\left(A_i\right) - \sum_{1\le i<j\le n} P\left(A_iA_j\right) + \sum_{1\le i<j<k\le n} P\left(A_iA_jA_k\right) + \cdots + (-1)^{n-1} P\left(A_1A_2\ldots A_n\right). \tag{1.1.4}$$

1.2　计数原理

1.2.1　排列组合

定义 1.15 若从 n 个互异的元素中一并取出考虑顺序的 k 个 $(k \le n)$，则一共有

$$A_n^k = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!} \tag{1.2.1}$$

种取法.

定义 1.16 若从 n 个互异的元素中一并取出不考虑顺序的 k 个 $(k \le n)$，则考虑到 k 个元素的全排列，一共有

$$C_n^k = \binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!} \tag{1.2.2}$$

种取法.

注意，组合数的计算其实蕴含了产生的各个序列是完全互异的. 若序列之间存在等价情况，也需要考虑在内.

例 1.2 甲、乙、丙、丁四人进行乒乓球双打练习，两人一对地结为对打的双方，有多少种不同的结对方式?

解 因为选定其中两个人另外两个也就确定了，所以一共有一半是重复的. 故应该是 $\dfrac{\binom{4}{2}}{2} = 3$ 种方式.

定理 1.3 若从 n 个互异的元素中依次有放回地取出不考虑顺序的 k 个，一共有 $\binom{n+k-1}{n}$ 种取法.

证明 因为我们可以将这个问题等价地转换成相同元素的分组问题，则此问题等价于将 k 个相同的小球放进 n 个互异的盒子里（允许空盒）. 这类对相同元素的可为零分割一般采取隔板法，也即将 $n-1$ 个隔板和 k 个小球一起排列，从左到右一一对应分组. 由于隔板和小球都是相同的，所以共有 $\dfrac{A_{n+k-1}^{n+k-1}}{A_{n-1}^{n-1}A_k^k} = \dfrac{A_{n+k-1}^k}{A_k^k} = \binom{n+k-1}{k}$ 种取法. ∎

1.2.2　多组组合

定理 1.4 有 n 个互异元素,现在要将其分为 $k(k \leq n)$ 个互异的组,使得各组依次有 $n_1, n_2, \ldots, n_k$ $\left(\sum_{i=1}^{k} n_i = n\right)$ 个元素. 那么一共有

$$\frac{n!}{n_1!n_2!\ldots n_k!} \tag{1.2.3}$$

种分法.

证明 采用逐次取出的方法即可.$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{k-2}}{n_{k-1}} = \frac{n!}{n_1!n_2!\ldots n_k!}$.■

定理 1.5 有 n 个互异元素，现在要将其分为 $k(k \leq n)$ 个性质上相同的组，使得各组依次有 $\underbrace{n_1, n_1, \ldots, n_1}_{m_1 \text{ groups}}, \underbrace{n_2, n_2, \ldots, n_2}_{m_2 \text{ groups}} \cdots, \underbrace{n_k, n_k, \ldots, n_k}_{m_k \text{ groups}}$ $\left(\sum_{i=1}^{k} m_i n_i = n\right)$ 个元素. 那么一共有

$$\frac{n!}{(n_1!)^{m_1}(n_2!)^{m_2}\cdots(n_k!)^{m_k}} \cdot \frac{1}{m_1!m_2!\cdots m_k!} \tag{1.2.4}$$

种分法.

证明 只需要在上一种情况下考虑人数相同的组的全排列即可. 共有 $\frac{n!}{(n_1!)^{m_1}(n_2!)^{m_2}\cdots(n_k!)^{m_k}} \cdot \frac{1}{m_1!m_2!\cdots m_k!}$ 种分法. ■

定理 1.6 有 n 个元素属于 k 个不同的类，同类元素不可辨认，各类元素分别有 $n_1, n_2, \ldots, n_k$ 个，其中 $\sum_{i=1}^{k} n_i = n$. 要把它们排成一列，则共有

$$\frac{n!}{n_1!n_2!\ldots n_k!} \tag{1.2.5}$$

种排法.

证明 只需要考虑全排列后各个类别分别全排列即可. ■

例 1.3 一批产品有 N 个，其中废品有 M 个. 现从中随机取出 n 个，在以下两种情形下，分别求“其中恰好有 m 个废品”这一事件的概率.(1) 有放回.(2) 无放回.

解 (1) 为了考虑问题方便，不妨设元素之间是互异的. 显然 $|\Omega| = N^n$. 分以下步骤：(a) 首先在 n 个样本中确定 m 个废品所占的位置，共有 $\binom{n}{m}$ 种. (b) 其次，废品可以安排 M^m 种，非废品可以安排 N^{n-m} 种. 故最终结果为 $\frac{\binom{n}{m}M^m N^{n-m}}{N^n} = \binom{n}{m}\left(\frac{M}{N}\right)^m\left(\frac{N-M}{N}\right)^{n-m}$.

(2) 为了考虑问题方便，不妨设废品与废品（非废品与非废品）没有区别. 显然 $|\Omega| = \binom{N}{n}$. 在是废品和不是废品的 $M, N-M$ 个元素中分别挑选 $m, n-m$ 个即可. 答案是 $\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$.

例 1.4 n 个男生，m 个女生排成 (1) 一排 (2) 一圈 $(m \le n+1)$，求任意两个女孩不相邻的概率.

解 (1) 假设男女孩都是互异的. 那么采用插空法，$|\Omega| = (m+n)!$，且男孩有 $n!$ 种排列. 而女孩确定如何安插的方式有 A_{n+1}^m 种（注意！这里采用的是一个个安插女孩的方式，所以已经计及了女孩的互异性）. 故答案为 $\frac{n!(n+1)!}{(m+n)!(n+1-m)!}$.

(2) **错误的解法**：只需要将上一问中的插空个数变为 n 个即可. 注意，我们在计算样本空间和事件的模的时候已经默认取定了圆圈的起点和排列的顺序，因而不需要再考虑圆圈的旋转性了. 所以答案为 $\frac{n!A_n^m}{(m+n)!}$

正解：$|\Omega| = \frac{(n+m)!}{n+m} = (n+m-1)!$, 其中除以 $n+m$ 是考虑到旋转重复.$|A| = \frac{C_n^m m!n!}{n}$, 其中除以 n 是考虑到插入前主排列的旋转重复. 故 $P(A) = \frac{(n+1)!C_n^m m!}{(n-m+1)!}$

例 1.5 设有 n 个人随机地坐到礼堂第一排 N 个座位上，试求下列事件的概率:

(1) 任何人都没有邻座，也即任何人的邻座都空着.

(2) 每人恰有一个邻座.

(3) 关于中央对称的两个座位至少有一个空着.

解 为了方便考虑问题，设这些人都是相同的. 显然 $|\Omega| = C_N^n$(如果是互异的那就是 A_N^n).

(1) 要达到这个目的，必须要求 $N \ge 2n-1$(坐在最边上的当然就没有邻座了). 在这个前提下，使用插空法，也即将 n 个人安插在 $N-n+1$ 个空隙中，也即 $|A| = C_{N-n+1}^n$. 故 $p = \frac{(N-n+1)(N-n)\cdots(N-2n+1)}{N(N-1)\cdots(N-n+1)}$

(2) 要达到这个目的，首先需要 n 为偶数，否则无论怎样总会出现没有邻座或者两个邻座的现象. 这里仍然使用插空法，只不过将两两捆绑在一起插空以确保满足条件. 此时需要总座位数 $N \ge \frac{3}{2}n-1$. 满足这个前提之后，直接考虑 $\frac{n}{2}$ 个排列元素. 这里就体现出之前假设人是相同的好处了，因为此时不必考虑分组方式的互异性. 可以得到 $|B| = C_{N-n+1}^{n/2}$.

(3) 分两种情况.(i)n 为奇数. 此时又有两种情况：(a) 中间坐人，那么就是 $n-1$ 个人坐 $\frac{N-1}{2}$ 个座位；(b) 中间不坐人，那么就是 n 个人坐 $\frac{N-1}{2}$ 个座位. 显然 $\frac{N-1}{2} < n-1$ 的时候是不可能满足条件的；有 $\frac{N-1}{2} = n-1$ 的时候，那么 $|C| = 2^{n-1}$；有 $\frac{N-1}{2} \ge n$ 的时候，就有 $|C| = C_{(N-1)/2}^{n-1}2^{n-1} + C_{(N-1)/2}^n 2^n$.

(ii)n 为偶数. 和上面的讨论一致，首先要求 $N \ge 2n$，那么 $|C| = C_{N/2}^n 2^n$.

例 1.6 (配对问题) 将 n 双鞋打乱后随机两两分成 n 组，恰好每一组均为一双鞋的概率为？

解 设组是有序的，则如果考虑一双鞋里的两只是一样的，那么 $|\Omega| = \dfrac{(2n)!}{\underbrace{2!\cdots 2!}_{n}}, |A| = n!$. 即便考虑不一样，那也只不过是将 $|\Omega|$ 的分子乘到 $|A|$ 上而已.

1.3　几何概型

定义 1.17 设 Ω 是 Euclid 空间内确定的集合，且 $0 < \sigma(\Omega) < +\infty$. 那么对于 $\forall A \subset \Omega$，有

$$P(A) = \frac{\sigma(A)}{\sigma(\Omega)}. \tag{1.3.1}$$

这里 $\sigma(\cdot)$ 表示测度.

例 1.7 甲乙两人相约在 $[0, T]$ 时段内会面，规定先到者等候 $t(0 < t \leq T)$ 的时间再离去. 求甲乙两人会面的概率.

解 设 x, y 分别为甲乙两人到的时间.$\Omega = [0, T] \times [0, T], A = \{(x, y) \mid |x - y| \leq t\} \cap \Omega$. 故 $P(A) = \dfrac{\sigma(A)}{\sigma(\Omega)} = 1 - \left(1 - \dfrac{t}{T}\right)^2$.

例 1.8 (蒲丰投针实验) 桌面上画满间隔均为 a 的平行线，现向桌面任意投放一长 $l(l < a)$ 的针，求针与某直线相交的概率.

解 $\Omega = \left\{(\rho, \theta) : 0 \leq \rho \leq \dfrac{a}{2}, 0 \leq \theta \leq \dfrac{\pi}{2}\right\}, E = \left\{(\rho, \theta) : \rho \leq \dfrac{l}{2} \sin\theta\right\} \cap \Omega$. 故 $p = \dfrac{\displaystyle\int_0^{\pi/2} \frac{l}{2}\sin\theta \mathrm{d}\theta}{\dfrac{\pi a}{4}} = \dfrac{2l}{\pi a}$.

1.4　条件概率

1.4.1　全概率公式

定义 1.18 设事件 A, B 是随机试验 Ω 中的两个事件. 若 $P(B) > 0$ 称

$$P(A|B) = \frac{P(AB)}{P(B)} \tag{1.4.1}$$

为事件 B 发生条件下事件 A 发生的条件概率.

上述定义提示我们，样本空间是相对的. 事实上，我们可以考虑条件概率套条件概率的分解

$$P(AB|C) = P(A|BC)P(B|C) \tag{1.4.2}$$

上面的 C 可以理解成“大前提”性质的事件.

例 1.9 有 10 个产品，其中有 3 个次品. 从中一个个不放回抽取检验，问第一次取到次品后第二次再取到次品的概率.

解 本题要注意的陷阱是，检验 9 次就足够了. 因此 $|\Omega| = A_{10}^{9}$, $P(\text{第一次取出的是次品且第二次取出的是次品}) = \frac{6}{90}$, $P(\text{第一次取出的是次品}) = \frac{3}{10}$.

定理 1.7 (条件乘法定理)

$$P(AB) = P(A|B)P(B), \tag{1.4.3}$$

推广到 n 个事件

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\ldots A_{n-1}). \tag{1.4.4}$$

这是将多个事件的交化成条件概率的形式的有效方法. 在未给出事件的相互联系的时候，上式是计算的一个重要公式.

例 1.10 将 n 根短绳的 $2n$ 个端头任意两两连接，试求恰好连成 n 个圈的概率.

解 首先我们可以人为地规定绳子的编号 (即绳子是互异的)，这样就方便我们对事件进行划分. 那么记 A_i 为 i 号绳子自己的两端连接成功的事件. 那么就有 $p = P(A_1A_2...A_n)$，再使用定理1.7，得到 $P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1...A_{n-1}) = \frac{1}{2n-1}\frac{1}{2n-3}\cdots\frac{1}{3}\frac{1}{1} = \frac{1}{(2n-1)!!}$. 其中 !! 表示二阶阶乘，留意.

定义 1.19 设 $B_1, B_2, \ldots, B_n$ 两两不相容的一组事件 (显然，两两不相容等价于全体不相容)，且

$$\bigcup_{i=1}^{n} B_i = \Omega. \tag{1.4.5}$$

则称 $B_1, B_2, \ldots, B_n$ 为 Ω 的一个分割 (又称完备事件,partition).

定理 1.8 (全概率公式) 若 $\{B_1, B_2, \ldots, B_n\}$ $(1 \le n \le +\infty)$ 是 Ω 的一个分割，且 $P(B_i) > 0 (i = 1, \ldots, n)$(这个条件可有可无)，$A \subset \Omega$，则

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{1.4.6}$$

全概率公式对方便求出条件概率的场合比较适用. 当然，条件概率也有全概率公式

定理 1.9 (条件概率的全概率公式) 若 $\{B_1, B_2, \ldots, B_n\}$ $(1 \le n \le +\infty)$ 是 Ω 的一个分割，且 $P(B_i) > 0 (i = 1, \ldots, n)$(这个条件可有可无)，$A, C \subset \Omega, P(C) > 0$，则

$$P(A|C) = \sum_{i=1}^{n} P(A|B_i C)P(B_i|C). \tag{1.4.7}$$

有时也会利用 B 与 $\overline{B}$ 人为构造一个分割.

例 1.11 设有来自三个地区的考生报名表共 50 份，三个地区分别有 10，15 和 25 份, 其中女生的报名表分别为 3 份，7 份和 5 份, 现随机地选一个地区, 从该地区的报名表中先后抽出 2 份.

(1) 求先抽到的 1 份是女生报名表的概率.

(2) 已知后抽到的 1 份是男生报名表, 求先抽到的 1 份是女生报名表的概率.

解 设 A_i 表示事件"选定地区 i", B_i 表示不同的组合 $\{MM\}, \{MF\}, \{FM\}, \{FF\}$(M=Male, F=Female). 那么使用全概率公式就可以了.

(1) 本题要注意的陷阱是：$P(A_i)$ 应当是均等地等于 $\dfrac{1}{3}$ 的，因为这是从三个地区中随机抽选而并非按比例抽选. 所以

$$P(B_3 + B_4) = \sum_{i=1}^{3} P(B_3 + B_4|A_i)P(A_i) = \frac{1}{3}\left(\frac{3 \times 9}{A_{10}^2} + \frac{7 \times 14}{A_{15}^2} + \frac{5 \times 24}{A_{25}^2}\right) = \frac{29}{90} \tag{1.4.8}$$

(2) 先求得

$$\begin{aligned} P(B_1 + B_3) &= \sum_{i=1}^{3} P(B_1 + B_3|A_i)P(A_i) \\ &= \frac{1}{3}\left(\frac{3 \times 7 + 7 \times 6}{A_{10}^2} + \frac{7 \times 8 + 8 \times 7}{A_{15}^2} + \frac{5 \times 20 + 20 \times 19}{A_{25}^2}\right) = \frac{61}{90} \end{aligned} \tag{1.4.9}$$

再求得

$$P\left((B_1+B_3)\cap(B_3+B_4)\right)=P(B_3)=\frac{1}{3}\left(\frac{3\times 7}{A_{10}^2}+\frac{7\times 8}{A_{15}^2}+\frac{5\times 20}{A_{25}^2}\right)=\frac{2}{9} \tag{1.4.10}$$

那么

$$P(B_1+B_4|B_1+B_3)=\frac{P(B_3)}{P(B_1+B_3)}=\frac{20}{61} \tag{1.4.11}$$

全概率公式在暗含某种敏感问题的时候也有其作用.

例 1.12 四宫辉夜同学想知道有多少人觉得她和白银御行同学是般配的，所以做了一个问卷调查. 问卷内容包含两个题目：(1) 你觉得四宫辉夜同学和白银御行同学般配吗？(2) 你穿平角内裤吗？问卷进行的方式是让石上优同学掷硬币，掷到了正面就回答问题 (1)，否则回答问题 (2). 由于两个问题的答案都只有“是”或者“否”，所以不必担心看到辉夜大小姐牙白脸色的情况. 现在已知秀知院学园内平角内裤率为 p，回答“是”的比率为 q，则有多少人是推这一对的？

解 设推这一对的概率为 r，则回答“是”的概率

$$P(Y)=P(Y|A)P(A)+P(Y|B)P(B)=r\times\frac{1}{2}+p\times\frac{1}{2}=q \tag{1.4.12}$$

那么显然

$$r=2q-p. \tag{1.4.13}$$

例 1.13 (Polya 罐子模型) 罐子中放有 a 个白球和 b 个黑球，每次从罐子中随机抽出一个球，并且连同 c 个同色球一起放回罐子中，如此反复进行. 试证明：在第 n 次取球的时候取出白球的概率为 $\dfrac{a}{a+b}$.

解 这是一个非常巧妙的问题. 我的上帝啊，我保证您在观赏完这个问题的解答之后惊奇于此的，我发誓！否则我就用那肮脏的靴子蘸上隔壁珍妮太太的苹果派狠狠地踢您一脚！先生！

用 A_i 表示在第 i 次取出白球的事件. 这里仍然是使用 $A_1,\overline{A}_1$ 作为参考分划的方式，理由是第一个事件往往是容易计算概率的. 那么

$$P(A_k)=P(A_1)P(A_k|A_1)+P(\overline{A}_1)P(A_k|\overline{A}_1) \tag{1.4.14}$$

这一步又要犯难了：我怎么知道 $P(A_k|A_1)$ 是多少呢？

这个解答的另一个巧妙之处在于采用了数学归纳法.

我们不妨承认假设：设罐子内白球 α 个，黑球 β 个，那么在第 $k-1$ 次取球的时候符合题设条件，取出白球的概率为 $\dfrac{\alpha}{\alpha+\beta}$，也即

$$\alpha=a,\beta=b \xrightarrow{\text{after } k-1 \text{ times}} P(A_{k-1})=\frac{\alpha}{\alpha+\beta}=\frac{a}{a+b} \tag{1.4.15}$$

然后再预先设定 A_1 发生，也即

$$\alpha=a+c,\beta=b \xrightarrow{\text{after } k-1 \text{ times}} P(A_{k-1})=\frac{\alpha}{\alpha+\beta}=\frac{a+b}{a+b+c} \tag{1.4.16}$$

为什么明明是数学归纳法，这里却直接代入了 $\dfrac{\alpha}{\alpha+\beta}$? 因为 $\dfrac{\alpha}{\alpha+\beta}$ 是相对预先设定好的 $k-1$ 次而言的，所以自然就可以直接用了. 这种改变参照事件的方法在之后的无记忆性里面还可以找到影子. 自然地，对于 A_1 不发生的情况

$$\alpha=a,\beta=b+c \xrightarrow{\text{after } k-1 \text{ times}} P(A_{k-1})=\frac{\alpha}{\alpha+\beta}=\frac{a}{a+b+c} \tag{1.4.17}$$

接下来只需要验证 $P(A_k)=\dfrac{a}{a+b}$ 即可. 代回一开始的分割式

$$\begin{aligned} P(A_k)&=P(A_1)P(A_k|A_1)+P(\overline{A}_1)P(A_k|\overline{A}_1)\\ &=\frac{a}{a+b}\frac{a+b}{a+b+c}+\frac{b}{a+b+c}\frac{a}{a+b+c}=\frac{a}{a+b} \end{aligned} \tag{1.4.18}$$

大功告成.

下面同样有一个结合递推与全概率公式的例子.

例 1.14 **一罐内有 a 个黑球和 b 个白球，从中任意取一球，如果是白球则将它放回去，如果是黑球，则从另一罐内取一白球替换它放回去. 在重复 n 次这样的做法后，求第 $n+1$ 次取出的是白球的概率.**

解 这一题没上题刺激. 直接设 A_k 为第 k 次取出的是白球的事件，其对应的概率为 p_k. 显然

$$\begin{aligned} p_{n+1}=P(A_{n+1})&=P(A_{n+1}|A_n)P(A_n)+P(A_{n+1}|\overline{A}_n)P(\overline{A}_n)\\ &=p_np_n+\left(p_n+\frac{1}{a+b}\right)(1-p_n)=\left(1-\frac{1}{a+b}\right)p_n+\frac{1}{a+b} \end{aligned} \tag{1.4.19}$$

结合初值 $p_1 = \dfrac{a}{a+b}$，得到

$$p_{n+1} = 1 - \left(1 - \frac{1}{a+b}\right)^n \frac{b}{a+b} \Rightarrow p_n = 1 - \left(1 - \frac{1}{a+b}\right)^{n-1} \frac{b}{a+b} \tag{1.4.20}$$

(不知道这个递推怎么化成通项的自觉复习高中数列知识).

1.4.2 Bayes 公式

定理 1.10 设 $\{B_1, B_2, ..., B_n\}$ 是样本空间的一个分割 (其中暗含了 $P(B_i) > 0$)，A 是 Ω 中的一个事件，且 $P(A) > 0$，那么

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n} P(A|B_j)P(B_j)} \tag{1.4.21}$$

在因果关系互换的时候，那 Bayes 公式就要登场了. 其实 Bayes 公式本质上就是利用乘法的交换律进行事件的导来导去，以达到因果互换的目的.

例 1.15 一种诊断某癌症的试剂，经临床试验有如下记录: 有癌症病人阳性的概率为 95%，无癌症病人阴性的概率为 95%. 现用这种试剂在某社区进行癌症普查，设该社区癌症发病率为 0.5%，问某人反应为阳性时该人患癌症的概率.

解 这是典型的条件概率因果逆转的问题. 通常来说是“患病了才会有阳性和阴性之分”，但是这里需要通过“阴阳性倒推是否患病”. 而且要注意“有癌症病人阳性的概率为 95%”是条件概率. 那么就使用 Bayes 公式就好了. 答案是 8.7% .

1.4.3 事件的独立性

定义 1.20 设 A, B 是随机试验中的两个事件，若 $P(AB) = P(A)P(B)$，则称事件 A 和 B 相互独立.

定理 1.11 若 A, B 独立且 $P(B) \neq 0$，则 $P(A|B) = \dfrac{P(AB)}{P(B)} = P(A)$.

可以知道上述定理为什么只有第一个是独立性的定义. 因为独立性的定义是没有任何前置条件的，而第二个则要求 $P(B) > 0$. 事实上第二个也是常用的推论. 我们就有以下结论

定理 1.12 $\Omega, \varnothing$ 和概率为 0 的事件与任意事件独立.

肯定地，$\varnothing$ 不等价于概率为 0. 比如考虑零测集上的事件，显然是可以发生的.

上述分开两者是为了强调这一点. 如果考虑其互斥事件的话，我们还可以进行适当的推广.

定义 1.21 对于事件 $A_1, ..., A_n$，若

$$P(A_1...A_n) = P(A_1)...P(A_n) \tag{1.4.22}$$

则称 $A_1, ..., A_n$ 相互独立.

定义 1.22 对于事件 $A_1, ..., A_n$，若

$$\forall i \neq j, P(A_iA_j) = P(A_i)P(A_j) \tag{1.4.23}$$

则称 $A_1, ..., A_n$ 两两独立.

定理 1.13 独立不具有传递性. 也即相互独立 $\dfrac{\Rightarrow}{\not\Leftarrow}$ 两两独立.

定义 1.23 用 $\widetilde{A}$ 表示 A 或 $\overline{A}$.

定理 1.14 事件 $A_1, ..., A_n$ 相互独立当且仅当

$$P(\widetilde{A_1}...\widetilde{A_n}) = P(\widetilde{A_1})...P(\widetilde{A_n}) \tag{1.4.24}$$

其中有 2^n 个式子.

第二章　随机变量及其分布

2.1　随机变量

定义 2.1 对 $\forall A \subset \Omega$，可定义示性函数

$$I_A(\omega) = \begin{cases} 1, \omega \subset A, \\ 0, \omega \not\subset A. \end{cases} \tag{2.1.1}$$

定义 2.2 Ω 为一样本空间, 定义随机变量 X 为一映射 $X : \Omega \to \mathbb{R}$, 使得 $\{\omega : X(\omega) \leq x\} \in \mathscr{F}$.

定义 2.3 定义分布函数

$$F(x) = P(X \leq x). \tag{2.1.2}$$

2.2　离散型随机变量

定义 2.4 设 X 为一随机变量，若 X 只取有限或可数个值，则称 X 为一个 (一维) 离散随机变量. 设其全部可能值为 $\{a_1, a_2, \dots\}$，则称 $p_i = P(X = a_i), i = 1, 2, \dots$ 为 X 的概率函数.

显然，概率函数必须满足 $p_i \leq 0, i = 1, 2, \dots$ 以及 $\sum\limits_i p_i = 1$. 概率函数和取值一一对应的表格可以以矩阵形式列出，也即

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ p_1 & p_2 & \cdots \end{pmatrix} \tag{2.2.1}$$

上式叫做 X 的分布列.

定义 2.5 若一次试验的结果只有 A 与 $\overline{A}$，则称其为一次 Bernoulli 试验.

定义 2.6 若将概率相同的 Bernoulli 试验独立地重复 n 次，则称其为 n 重 Bernoulli 试验.

2.2.1 0-1 分布

定义 2.7 称

$$P(X=1)=p, P(X=0)=q=1-p \tag{2.2.2}$$

为 X 服从 0-1 分布. 又称 Bernoulli 分布.

2.2.2 二项分布

定义 2.8 称

$$P(X=k)=\binom{n}{k}p^k\left(1-p\right)^{n-k}, k=0,1,\ldots,n \tag{2.2.3}$$

为 X 服从二项分布，记为 $X\sim B(n,p)$.

记住，分布规律之后 k 的取值是必写的，否则不声明取值范围的分布规律是无意义的.

二项分布考察的是 n 重 Bernoulli 试验中成功了几次，不考虑这些成功的次数是如何分配的.

2.2.3 Poisson 分布

定义 2.9 称

$$P(X=k)=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, k=0,1,2,\ldots,\lambda>0 \tag{2.2.4}$$

为 X 服从参数为 λ 的 Poisson 分布，记为 $X\sim P(\lambda)$.

定理 2.1 (小数定理) 在 n 重 Bernoulli 试验中，以 p_n 代表事件 A 在试验中出现的概率，它与实验总数 n 有关. 若有 (稳定稀有事件条件)

$$n\to\infty, np_n\to\lambda, \tag{2.2.5}$$

那么

$$n\to\infty, \binom{n}{k}p_n^k\left(1-p_n\right)^{n-k}\to\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}. \tag{2.2.6}$$

关于“稀有事件”的定义，这里从一个更加严谨的角度叙述.

考虑将区间 $[0,1]$ 划分成均等的 n 份 $l_j:\left[\frac{j-1}{n},\frac{j}{n}\right)$(最后一个区间右闭). 那么在 l_j 之间独立的前提下，可以假设有这样一种事件，$P(X_1|l_j)=\frac{\lambda}{n}$，也即发生的概率与选取区间长度成正比，且 $P(X\geq 2|l_j)=o\left(\frac{1}{n}\right)$(逐次发生)，那么就有 $X\sim B\left(n,\frac{\lambda}{n}\right)$，在 $n\to\infty$ 的情况下就有 $X\sim P(\lambda)$.

事实上，如果我们可以确定对某对特定的 n_0,p_0，有 $n_0p_0\approx\lambda$ 的话，我们就可以用 Poisson 分布进行近似.

例 2.1 现在需要 100 个符合规格的元件. 从市场上买的该元件有废品率 0.01. 考虑到有废品存在，我们准备买 $100+a$ 个元件使得从中可以挑出 100 个符合规格的元件. 我们要求在这 $100+a$ 个元件中至少有 100 个符合规格的元件的概率不小于 0.95. 问 a 至少要多大?

解 由于 $np_n=(100+a)0.01=1+0.01a\approx 1$，故可进行近似 $X\sim P(1)$(表示废品个数). 之后再解不等式 $P(X\leq a)\geq 0.95$ 即可.

例 2.2 假设一块放射性物质在单位时间内发射出的 α 粒子数 ξ 服从参数为 λ 的 Poisson 分布。而每个发射出来的 α 粒子被记录下来的概率是 p. 如果各粒子是否被记数器记录是相互独立的，试求记录下来的 α 粒子数 η 的分布.

解 以事件 $\{\xi=n\},n=0,1,2,...$ 作为划分. 故 $P(\eta=k)=\sum\limits_{n=0}^{\infty}P(\eta=k|\xi=n)P(\xi=n)=\sum\limits_{n=k}^{\infty}\binom{n}{k}p^kq^{n-k}\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}=\sum\limits_{n=k}^{\infty}\frac{(\lambda q)^{n-k}\,\mathrm{e}^{-\lambda}\,(\lambda p)^k}{k!\,(n-k)!}=\frac{(\lambda p)^k}{k!}\mathrm{e}^{-\lambda p}$，其中利用了 $\mathrm{e}^x=\sum\limits_{i=1}^{\infty}\frac{x^i}{i!}$.

2.2.4 几何分布

定义 2.10 在 n 重 Bernoulli 试验中，若 $n\to\infty$，则称该试验为可列重 Bernoulli 试验.

定义 2.11 称

$$P(X=k)=q^{k-1}p=(1-p)^{k-1}p,k=1,2,... \tag{2.2.7}$$

为服从几何分布. 记为 $X\sim G(p)$.

几何分布的意义是，进行可列重 Bernoulli 试验时第一次成功所需的次数. 所以有无限次可能的验证前提是必要的.

定理 2.2 若 $X\sim G(p)$，则 $E(X)=\frac{1}{p}$.

定理 2.3 (几何分布的无记忆性) 若 $X\sim G(p)$，则 $P(\xi>m+n|\xi>m)=P(\xi>n)$.

定理 2.4 几何分布是唯一具有无记忆性的离散分布.

2.2.5 Pascal 分布

定义 2.12 称

$$P(X_r = k) = \binom{k-1}{r-1} p^r q^{k-r}, k = r, r+1, ... \tag{2.2.8}$$

为 X_r 服从 Pascal 分布 (负二项分布)，记为 $X \sim NB(r,p)$.

Pascal 分布的意义是第 r 次成功为止所需试验次数 X_r. 这意味着我们将最后一次成功剔除后就可以随意安排之前 $r-1$ 次成功的排列了. 故

$$P(X_r) = k = \binom{k-1}{r-1} p^{r-1} q^{k-r} \cdot p = \binom{k-1}{r-1} p^r q^{k-r}, k = r, r+1, ... \tag{2.2.9}$$

无需特别记忆. 因为 Pascal 分布和二项分布本质相同，都是成功试验在全试验中的排列.

例 2.3 现有 n 种彩券，假设彩券总量相当大，以至于每次抽取不改变取出各类彩券的比例，求全收集到彩券的平均次数.

解 记 X_j 为已收集 $j-1$ 种彩券的条件下收集 j 种彩券的次数. 那么 $X_1 = 1, X_2 \sim G\left(\frac{n-1}{n}\right)$, ..., $X_j \sim G\left(\frac{n-(j+1)}{n}\right)$, ..., $X_n \sim G\left(\frac{1}{n}\right)$. 显然 $X = X_1 + X_2 + \cdots + X_n$，故 $E(X) = n\left(\sum_{i=1}^{n} \frac{1}{n}\right) \sim n \ln n$.

上述例子告诉我们，负二项分布的本质就是 r 个独立的超几何分布的排列 (只不过上述例子是更加一般的情况，也即每次超几何分布的概率不一样).

例 2.4 (Banach 火柴问题) 某人口袋里放有两盒火柴, 每盒装有火柴 n 根. 他每次随机取出一盒, 并从中拿出一根火柴使用. 试求他取出一盒, 发现已空, 而此时另一盒中尚余 r 根火柴的概率.

解 以 A 表示甲盒已空, 而此时乙盒中尚余 r 根火柴的事件. 由对称性知, 所求的概率等于 $2P(A)$. 我们将每取出甲盒一次视为取得一次成功，这样每次成功和失败的概率均为 $\frac{1}{2}$. 注意题目叙述的陷阱：题目其实是要求成功 $n+1$ 次，失败 $n-r$ 次. 因为取出这个动作本身就代表着一次试验的发生——发现已空和试验是否成功没有关系，只是没有火柴记录实验结果罢了. 所以剩下的就是套公式，答案是 $2\binom{2n-r}{n}2^{r-2n-1} = \binom{2n-r}{n}2^{r-2n}$.

例 2.5 (赌本分配问题) 甲乙两人进行赌博，现在他们的得分情况是甲 3 分，乙 4 分（谁赢谁得 1 分，假定输赢概率均为 $\frac{1}{2}$）. 游戏规则规定谁最先到达 6 分谁获得胜利，但是现在因为某些因素不得不终止游戏. 根据接下来甲乙获胜的概率分配 1000 元赌本，应该如何分配?

解 这实际上是要求 n 次成功发生在 m 次失败之前的概率. 注意，这并不是简单的 n 次成功实验和 $m-1$ 次失败实验的简单排列问题，因为一旦成功了那么整个实验就终止了. 因此必须分别考虑再予以求和. 记 $F_k=\{第n次成功发生在第k次试验\}$，则 $E=\bigcup\limits_{k=n}^{n+m-1}F_k$. 注意到各个 F_k 是互斥的，那么 $P(E)=\sum\limits_{k=n}^{n+m-1}P(F_k)=\sum\limits_{k=n}^{n+m-1}\binom{k-1}{n-1}p^nq^{k-n}$. 再将本例的情景代入这个普遍的公式里，也即甲获胜的概率为 $n=3,m=2$ 的情况，则 $P(E_1)=\binom{2}{2}2^{-3}+\binom{3}{2}2^{-4}=\dfrac{5}{16}$.

2.2.6　离散均匀分布

定义 2.13 记

$$P(X=a_k)=\frac{1}{n},k=1,2,...,n \tag{2.2.10}$$

为 X 服从离散均匀分布.

2.3　连续随机变量

我们先来进一步讨论随机变量分布函数的性质. 按照上一节的叙述，分布函数有以下几个等价表述

$$F(x)=P(X\le x)=P\left(\{\omega:X(\omega)\le x\}\right)=P\left(X^{-1}\left(-\infty,x\right]\right). \tag{2.3.1}$$

一般来说，我们要区分分布函数的自变量和对应的随机变量，这在之后的运算中是十分重要的. 以及，分布函数具有如下性质：

(1) 有界性.$0\le F(x)\le 1$.

(2) 规范性.$\begin{cases}\lim\limits_{x\to+\infty}F(x)=1,\\ \lim\limits_{x\to-\infty}F(x)=0.\end{cases}$. 这个性质的重要性在于，对于某些定义在 $\mathbb{R}$ 的子区间上的概率，在写出分布函数的时候，对于没有定义的地方一定要补上 0 和 1，这是尤其重要的！

(3) 右连续性. $\lim\limits_{x\to x_0^+}F(x)=F(x_0)$(注意，因为我们对随机变量的定义为左开右闭的形式，所以这里是右连续).

定义 2.14 若 $\exists f(x)\ge 0,\text{s.t. }F(x)=\displaystyle\int_{-\infty}^{x}f(t)\mathrm{d}t$(也意味着 $F(x)$ 在 $\mathbb{R}$ 上绝对连续，因为没有规定 $f(x)$ 必须连续所以不一定可导.)，那么 $F'(x)=f(x)$，称 $f(x)$ 为 $F(x)$ 的概率密度函数.

显然 $f(x)$ 应该满足以下性质：

(1)$f(x) \geq 0$.

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

反之，满足以上两个条件且可积的函数一定是概率密度函数. 若条件 (2) 的右边是某个实数 a，则 $g(x) = \frac{f(x)}{a}$ 是满足要求的概率密度函数，记变换 $\varphi : f \to g$ 为归一化.

定义 2.15 对可列个形如 $(-\infty, a]$, $a \neq +\infty$ 的集合进行交、并、补运算形成的集合称为 Broel 集.

定理 2.5 概率密度函数在任何实 Broel 集上的积分是这个集合内随机变量取值的概率. 即

$$\forall A \subset \mathscr{B}(\mathbb{R}), P(X \in A) = \int_A f(t)\mathrm{d}t. \tag{2.3.2}$$

定理 2.6 若随机变量分布函数 $F(x)$ 在 x_0 处连续，则 $P(X = x_0) = 0$. 否则

$$P(X = x_0) = \lim_{x \to x_0^+} F(x) - \lim_{x \to x_0^-} F(x). \tag{2.3.3}$$

证明 只证明第一个小命题. 显然，$\{X = x\} = \bigcap_{n=1}^{\infty} \left\{x - \frac{1}{n} < X \leq x\right\}$. 故 $P(x = x_0) \leq P\left(x_0 - \frac{1}{n} < X \leq x_0\right) = F(x_0) - F\left(x_0 - \frac{1}{n}\right)$. 当 $n \to \infty$，且 F 在 x_0 连续，故右边趋于 0. ■

这是一个重要的性质. 在前述离散分布时，各个值有离散地概率分布决定了其概率函数要在这个点“跳变”，并且在没有概率分布的情况下就必须维持不变.

2.3.1 均匀分布

均匀分布是指在某个测度大于 0 小于 ∞ 的区间内概率密度恒为常数的分布. 那么显然它的概率密度函数为 $f(x) = \begin{cases} \frac{1}{b-a}, a < x < b, \\ 0, x \leq a \text{ or } x \geq b \end{cases}$. 但是这样在最后标出取值范围的形式往往不利于计算 (常常会漏掉)，也不利于普适的规则.

所以我们需要使用示性函数来进行表示限定区间上的取值 (这样，所有定义域都可以表示为 $\mathbb{R}$ 了，规律的套用也具有了普适性). 以后没有特别说明都要使用示性函数 (或者在最后一步化成附加范围的形式).

定义 2.16 均匀分布的概率密度函数为

$$f(x) = \frac{1}{b-a} I_{(a<x<b)}. \tag{2.3.4}$$

那么我们就记 $X \sim U(a,b)$.

定理 2.7 可以求得均匀分布的分布函数为

$$F(x)=\begin{cases}0, x \leq a, \\ \dfrac{x-a}{b-a}, a<x<b, \\ 1, x \geq b.\end{cases} \tag{2.3.5}$$

不同的是，由于分布函数是积分的结果，当然就不方便用示性函数来表示了.

一般来说，我们是取区间 (a,b) 的. 这是因为 $U(a,b), U[a,b), U(a,b], U[a,b]$ 的分布函数在本质上是相同的.

一个特例：$X \sim U(0,1), f(x)=I-(0<x<1)(x), F(x)=\begin{cases}0, x \leq 0, \\ x, 0<x<1, \\ 1, x \geq 1.\end{cases}$

2.3.2 指数分布

定义 2.17 指数分布的概率密度函数为

$$\lambda \mathrm{e}^{-\lambda x} I_{(x \leq 0)} \tag{2.3.6}$$

其中 $\lambda>0$ 为常数.

定理 2.8 指数分布的分布函数为

$$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x}, x \geq 0, \\ 0, x<0.\end{cases} \tag{2.3.7}$$

指数分布常用来讨论电子元件的寿命问题. 引入失效率

$$h(x)=\lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x+\Delta x | X>x)}{\Delta x} \tag{2.3.8}$$

也就是在 x 时刻尚能正常工作但是在之后单位时间内失效的概率. 当 $h(x) \equiv \lambda(x>0)$ 时，则 x 服从指数分布，也即指数分布描述了无老化的电子元件的寿命分布.

证明 因为 $h(x)=\lim\limits_{\Delta x\to 0}\dfrac{P(x\le X\le x+\Delta x|X>x)}{\Delta x}=\lambda$，则假设 X 为元件寿命，其分布函数 $F(x)$，则代入 $P(x\le X\le x+\Delta x|X>x)=\dfrac{P(\{\,x\le X\le x+\Delta x\}\{X>x\})}{P(X>x)}=\dfrac{F(x+\Delta x)-F(x)}{1-F(x)}$ 故原极限为 $\dfrac{F'(x)}{1-F(x)}=\lambda$. 解这个微分方程，即得 $F(x)=1-\mathrm{e}^{-\lambda x}$. ■

定理 2.9 指数分布具有无记忆性，也即

$$P(X>s+t|X>s)=P(X>t). \tag{2.3.9}$$

定理 2.10 指数分布是唯一具有无记忆性的连续分布.

2.3.3　正态分布

定义 2.18 正态分布的概率密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{2.3.10}$$

其中 $x\in\mathbb{R}$. 记 X 服从正态分布为 $X\sim N(\mu,\sigma^2)(\sigma>0)$.

可以看出，σ 越大，概率密度曲线越“矮胖”，反之亦然.

定义 2.19 称 $X\sim N(0,1)$ 时的概率密度函数为标准密度函数

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \tag{2.3.11}$$

对应的分布函数称为标准分布函数

$$\Phi(x)=\frac{1}{2}\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \tag{2.3.12}$$

定理 2.11 若 $X\sim N(\mu,\sigma^2)$，则 $\dfrac{X-\mu}{\sigma}\sim N(0,1)$. 此时

$$P\left(X\in[a,b]\right)=P\left(\frac{X-\mu}{\sigma}\in\left[\frac{a-\mu}{\sigma},\frac{b-\mu}{\sigma}\right]\right)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right). \tag{2.3.13}$$

Φ 是可以保留的记号.

此外，对于负数的情况

定理 2.12 $\Phi(-x) = 1 - \Phi(x)$.

记住常用的积分结果：

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-x^2}\mathrm{d}x = \sqrt{\pi} \quad \int_{-\infty}^{+\infty} \mathrm{e}^{-x^2/2}\mathrm{d}x = \sqrt{2\pi} \tag{2.3.14}$$

2.3.4 Laplace 分布

定义 2.20 Laplace 分布 (双指数分布, 重尾分布) 的概率密度函数为

$$f(x) = C\mathrm{e}^{-\lambda|x|}. \tag{2.3.15}$$

其中 $C = \dfrac{\lambda}{2}$ 为非独立参数，须满足归一化条件.

2.3.5 Γ 分布

首先回顾 Γ 函数的一些性质 (Euler 积分).Γ 函数的形式为

$$\Gamma(s) = \int_0^{+\infty} t^{s-1}\mathrm{e}^{-t}\mathrm{d}t, s > 0. \tag{2.3.16}$$

并且，形如幂函数和自然对数底数的幂的乘积的形式的积分 (尽管有因子) 都可以化成 Γ 函数的形式

$$\int_0^{+\infty} t^{\alpha-1}\beta^{\alpha}\mathrm{e}^{-\beta x} = \Gamma(\alpha) \tag{2.3.17}$$

特殊的值

$$\Gamma(1) = 1 \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{2.3.18}$$

并且有递推公式

$$\Gamma(s+1) = s\Gamma(s), s > 0 \tag{2.3.19}$$

根据这个可以得到

$$\Gamma(n) = (n-1)!, n \in \mathbb{N}_+ \tag{2.3.20}$$

且

$$\Gamma\left(n+\frac{1}{2}\right)=\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\cdots\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{(2n-1)!!}{2^n}\sqrt{\pi} \tag{2.3.21}$$

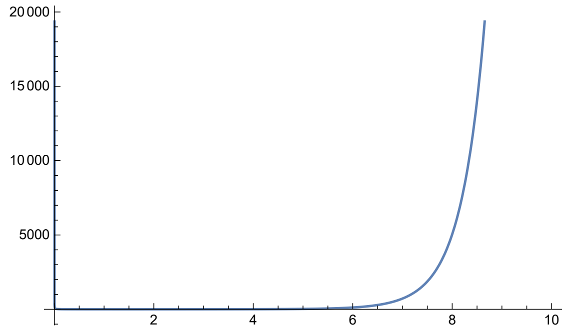


图 2.1　Γ 函数的图像

定义 2.21 Γ 分布的概率密度函数为

$$\gamma(x)=\frac{\beta^\alpha x^{\alpha-1}\mathrm{e}^{-\beta x}}{\Gamma(\alpha)}I_{(x>0)} \tag{2.3.22}$$

其中参数 α,β 独立，因为 f 在 $\mathbb{R}$ 上的积分恒为 1. 其中 α 称为形状参数，β 称为尺度参数.

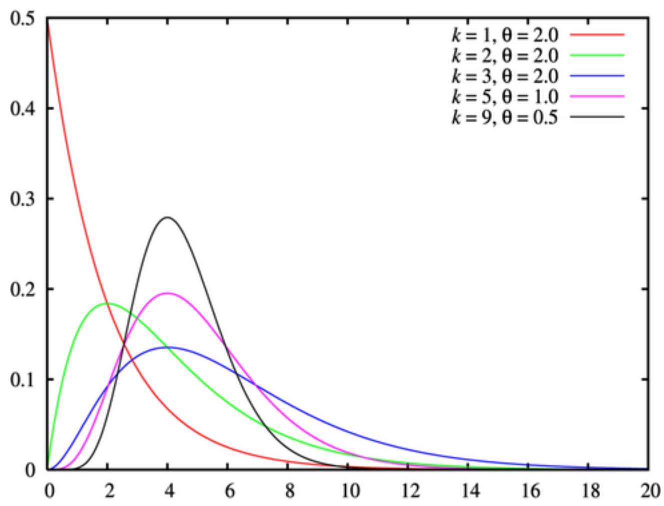


图 2.2　Γ 分布的概率密度函数, 其中 $k=\alpha,\theta=\beta$

定理 2.13 若随机变量 $X_1, ..., X_n$ 服从独立同分布 Exp(λ)，那么

$$S_n = \sum_{k=1}^{n} X_k \sim \Gamma(n, \lambda). \tag{2.3.23}$$

容易看出，$\gamma(x)$ 的最大值点为 $x = \dfrac{\alpha - 1}{\beta}$.

Γ 分布的引入使得之前的离散分布和连续分布串起来了. 有以下关系 (等可能试验情况下):

- 对于离散情况的二项分布，几何分布，负二项分布.
  - 发生次数服从二项分布.
  - 两次之间的间隔服从几何分布.
  - 第 n 次试验发生时之前已经成功的试验次数服从负二项分布.
- 对于连续情况的 Poisson 分布，指数分布，Γ 分布.
  - 发生次数服从 Poisson 分布，二项分布的极限.
  - 两次之间的间隔服从指数分布，几何分布的极限.
  - 第 n 次试验发生时之前已经成功的试验次数服从 Γ 分布，Poisson 分布的共轭先验分布.

2.4　随机向量

2.4.1　离散型随机向量

定义 2.22 对于一个向量 $X = (x_1, x_2, ..., x_n)$，若 $x_1, x_2, ..., x_n$ 都是随机变量，那么称 X 为随机向量.

定义 2.23 如果随机向量中每一个随机变量都是离散的，那么称这个向量为离散型随机向量. 设 $x_i (i = 1, 2, ..., n)$ 的所有可能取值为 $a_{ij} (j = 1, 2, ..., b_i)$，则称

$$p(j_1, j_2, ..., j_n) = P(x_1 = a_{1j_1}, x_2 = a_{2j_2}, ..., x_n = a_{nj_n}) \tag{2.4.1}$$

为 X 的概率函数.

定理 2.14 显然概率函数有如下性质：

(1)$p(j_1, j_2, ..., j_n) = \geq 0$.

(2) $\displaystyle\sum_{1 \leq j_1 \leq b_1} \sum_{1 \leq j_2 \leq b_2} \cdots \sum_{1 \leq j_n \leq b_n} p(j_1, j_2, ..., j_n) = 1.$

2.4.2　多项分布

定义 2.24 若事件 $A_1,A_2,...,A_n$ 是样本空间的一个分割，且记 $p_k=P(A_k)$，那么显然 $p_k\leq 0,\sum_{k=1}^{n}p_k=1$. 现在将试验独立重复地进行 N 次，分别用 X_i 表示 A_i 出现的次数，显然 $i=0,1,...,n$. 那么 $X=(X_1,X_2,...,X_n)$ 为服从多项分布的随机向量,记为 $X\sim M(N;p_1,p_2,...,p_n)$.

定理 2.15 多项分布 $X\sim M(N;p_1,p_2,...,p_n)$ 的概率函数为

$$P(X_1=k_1,X_2=k_2,...,X_n=k_n)=\frac{N!}{k_1!k_2!\cdots k_n!}p_1^{k_1}p_2^{k_2}\cdots p_n^{k_n}, \tag{2.4.2}$$

注意其中 $k_1+k_2+\cdots+k_n=N$.

上式的结果是显然的. 考虑定理1.4中所述的多组组合问题的结果，因为发生是有次序先后的，那么可以考虑将按次序先后发生的 N 次试验分成 n 组，也即多组组合. 而对于其中某一个分量 X_i 来说，显然 $X_i\sim B(N,p_i)$，也即多项分布的每一个分量遵从二项分布.

2.4.3　联合分布

如果考虑两个 (或多个) 分量之间的联系的话，那么称其为联合分布 (律). 设离散随机变量 X,Y 的所有可能取值为 $x_1,x_2,...,x_m;y_1,y_2,...,y_n$，那么我们可以用矩阵的形式表示每一种情况 $(X,Y)=(x_i,y_j)$ 的概率函数

$$\begin{pmatrix} p_{11} & p_{21} & \cdots & p_{n1}\\ p_{12} & p_{22} & \cdots & p_{n2}\\ \vdots & \vdots & \ddots & \vdots\\ p_{1m} & p_{2m} & \cdots & p_{nm} \end{pmatrix} \tag{2.4.3}$$

注意上面的记号习惯和矩阵是相反的，因为这里是按照 x,y 的顺序写的. 那么每一列的和就是对应 x_i 的取值概率 $P(X=x_i)$，每一行的和就是对应 y_j 的取值概率 $P(Y=y_j)$.

2.4.4　连续型随机向量

定义 2.25 称 $X=(x_1,x_2,...,x_n)$ 为 n 维连续型随机向量,如果存在 $\mathbb{R}^n$ 上的非负函数 $f(x_1,x_2,...,x_n)$，使得域 $E=[a_1,b_1]\times[a_2,b_2]\times[a_n,b_n]$ 上的积分满足

$$\int\int\cdots\int_E f(x_1,x_2,...,x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n=P(X\in E) \tag{2.4.4}$$

且称 f 为 X 的概率密度函数.

定义 2.26 对于 n 维连续型随机向量 $X=(x_1,x_2,...,x_n)$，称 $F(x_1,x_2,...,x_n)$ 为它的 (联合) 分布函数，当存在 F 使得

$$F(a_1,a_2,...,a_n)=P(-\infty<x_1\le a_1,-\infty<x_2\le a_2,...,-\infty<x_n\le a_n) \tag{2.4.5}$$

对任意 $(a_1,a_2,...,a_n\in\mathbb{R}$ 成立.

定理 2.16 可以验证 F 具有如下性质：

(1) 对每个变元 x_i 单调非降.

(2) 关于每个变元 x_i 连续.

(3) $\begin{cases}\lim\limits_{\bigcap_{i=1}^n\{x_i\to+\infty\}} F=1,\\ \lim\limits_{\bigcup_{i=1}^n\{x_i\to-\infty\}} F=0\end{cases}$

上面的性质 3 是重要的，因为这意味着当每个分量都趋于正无穷的时候，分布函数才能趋于 1；而只要有一个变量趋于负无穷，那么分布函数便趋于 0 了. 这是为了在降维后仍然满足归一化条件而成立的.

定理 2.17 显然

$$F(x_1,x_2,...,x_n)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_n}f(X_1,X_2,...,X_n)\mathrm{d}X_1\mathrm{d}X_2\cdots\mathrm{d}X_n \tag{2.4.6}$$

根据上述定理我们可以看出，分布函数其实就是对密度函数做变上限的积分. 如果令这个上限趋于正无穷，那么自然就变成了边缘化 (参见下一小节) 了.

事实上，已知分布函数求相对某一个变量的分布律 (也即边缘分布) 就是利用了这一性质.

定理 2.18 已知连续型随机向量 $(X_1,X_2,...,X_n)$ 的分布函数 $F(X_1,X_2,...,X_n)$，那么变量 $X_{a_1},X_{a_2},...,X_{a_i}(i<n)$ 的联合密度为

$$\frac{\partial}{\partial(X_{a_1})\partial(X_{a_2})...\partial(X_{a_i})}\left(\lim_{\cap_{k\neq j,\forall j=a_1,...,a_i}\{X_k\to+\infty\}}F(X_1,X_2,...,X_n)\right) \tag{2.4.7}$$

特别地，求其中某个变量 X_i 的边缘分布密度

$$\frac{\partial}{\partial X_i}\left(\lim_{\cap\forall i,i\neq j\{X_j\to+\infty\}}F(X_1,X_2,...,X_n)\right) \tag{2.4.8}$$

如果要求 $(X_1,X_2,...,X_n)$ 的联合分布密度

$$\frac{\partial}{\partial(X_1)...\partial(X_n)}F(X_1,X_2,...,X_n) \tag{2.4.9}$$

上面这个定理告诉我们，要求边缘分布，就必须先把不需要的变量给求极限，这事实上是根据分布函数与密度函数的关系可以得到的. 这个积分关系对于二元函数有很有用的推论.

定理 2.19 对于二元连续随机向量，已知其分布函数 F，则

$$P(X\in[x_1,x_2],Y\in[y_1,y_2])=F(x_1,y_1)+F(x_2,y_2)-(F(x_1,y_2)+F(x_2,y_1)). \tag{2.4.10}$$

其中 x_1,x_2,y_1,y_2 皆可以为无穷.

证明其实就是利用多变量微积分的方法，其实这个定理也挺好记得，只要是"主对角线之和减副对角线之和"就行了.

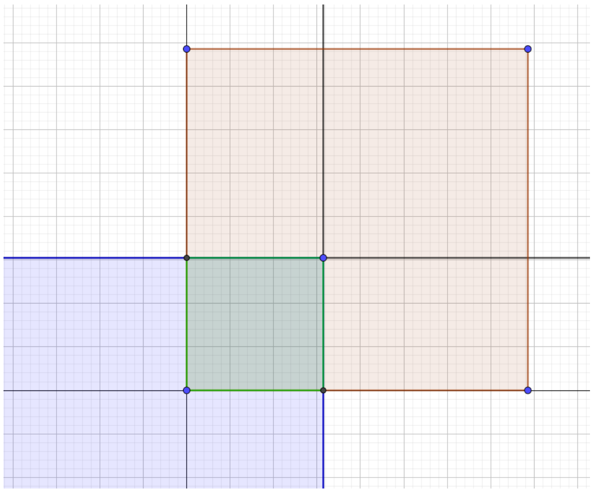
至于限制在一定区域内的分布，实际上和一维的状况是类似的，也即"定义域外分布函数一定不会增长". 但是多元变量的情况就得根据各个变量的情况分别分析了.

例 2.6 设二维随机变量 (X,Y) 的分布函数为

$$f(x,y)=\begin{cases}\cos x\cos y, 0<x,y<\dfrac{\pi}{2}\\ 0, \text{else}\end{cases} \tag{2.4.11}$$

(1) 试求 X,Y 联合分布函数.
(2) 试求 $P\left(0\le X\le\dfrac{\pi}{4}\le Y\le\dfrac{\pi}{2}\right)$.

解 (1) 根据限制域的特点，我们很显然要做分类讨论. 以下设待求函数为 $F(u,v)$. 但是，这是完全没有必要一个一个情况代的. 因为只需要在图上标出积分域，就自然而然清楚需要怎么积分了.



(a)u,v 中有一个小于 0(包括均小于 0). 很明显，只要有这种情况存在，积分域内就没有非 0 的情况.

(b)u,v 均介于 $0,\frac{\pi}{2}$ 之间. 那么 $F(u,v)=\int_0^u\int_0^v\cos x\cos y\mathrm{d}x\mathrm{d}y=\sin u\sin v$.

(c)u,v 中有一个大于 $\frac{\pi}{2}$，另一个介于 $0,\frac{\pi}{2}$ 之间. 那么 $F(u,v)=\int_0^1\int_0^v\cos x\cos y\mathrm{d}x\mathrm{d}y=\sin v$ 或同理，$\sin u$.

(d)u,v 均大于 $\frac{\pi}{2}$. 显然 $F(u,v)=1$.

综上，$F(x,y)=\begin{cases}0,x\leq 0\text{ or }y\leq 0\\ \sin x\sin y,0<x,y<\frac{\pi}{2}\\ \sin x,0<x<\frac{\pi}{2},y\geq\frac{\pi}{2}\\ \sin y,x\geq\frac{\pi}{2},0<y<\frac{\pi}{2}\\ 1,x,y\geq\frac{\pi}{2}\end{cases}$.(2) 只需套公式即可.

对于高维连续随机向量，我们一般不使用分布函数.

定义 2.27 $X=(x_1,x_2,...,x_n)$ 在域 $E=[a_1,b_1]\times[a_2,b_2]\times[a_n,b_n]$ 上服从 n 维均匀分布，则它的密度函数为

$$f(x_1,x_2,...,x_n)=\frac{1}{\sigma(E)}=\frac{1}{(b_1-a_1)(b_2-a_2)\cdots(b_n-a_n)}. \tag{2.4.12}$$

定义 2.28 (二元正态分布的定义，重要!)$X=(x,y)$ 在域 $\mathbb{R}^2$ 上服从二元正态分布，则它的密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{(x-a)^2}{\sigma_1^2}-2\rho\frac{(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\right)\right) \tag{2.4.13}$$

其中参数 $a,b\in\mathbb{R},\sigma_1,\sigma_2\in\mathbb{R}^+,\rho\in[-1,1]$. 记为 $X\sim N(a,b,\sigma_1^2,\sigma_2^2,\rho)$.

2.4.5 边缘分布 (边际分布)

定义 2.29 设 $X=(x_1,x_2,...,x_n)$ 为 n 为随机向量，取其中 $m<n$ 个分量再构成子向量 $(x_{j_1},x_{j_2},...,x_{j_m})$，在这个子空间 $\mathbb{R}^m\subset\mathbb{R}^n$ 内考虑的分布称为 X 的一个边缘分布.

对于二维离散型随机向量，其边际分布律就是每个分量的分布律，也就是列联表的行和分布/列和分布. 类似地，对于 n 维随机向量降维到 m 维随机向量的过程，可以通过对剩余 $n-m$ 个所有分量求和实现概率函数的求解.

显然，因为低维和高维的关系，边际分布律无法决定联合分布律.

> **作者注** 我在做题的时候犯过这样一个傻 * 错误，那就是："既然积分可以得到边际概率密度，那求个导不就得到联合概率密度了么". 大错特错！定积分 (广义定积分) 又不是不定积分，直接给你的变量给整没了你还积个毛线！事实上，这就是"边际分布律无法决定联合分布律"的最直观表述.

对于二维连续型随机向量，其边际分布律就是对另外一个分量进行 $\mathbb{R}$ 上的广义积分得到的，也即

$$P(x_1\leq X\leq x_2)=\int_{u=x_1}^{u=x_2}\int_{v=-\infty}^{v=+\infty}f(u,v)\mathrm{d}u\mathrm{d}v=\int_{x_1}^{x_2}f_X(u)\mathrm{d}u. \tag{2.4.14}$$

称 $f_X(x)=\displaystyle\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y$ 为 X 的边缘密度函数.Y 的边缘密度函数也是同理. 对于多元连续型随机向量，对 $(X_{j_1},X_{j_2},...,X_{j_m})$ 的边缘密度函数自然就是对余下 $n-m$ 个变量积分的结果了. 要注意的是，积分域并不一定是矩域，所以在不是矩域的情况下是需要按区域积分的. 这在之后的条件分布的求解中也会用到！(参见例2.8)

显然，因为低维和高维的关系，边际概率密度无法决定联合概率密度. 这句话的意义相当深刻！其本质就是，另外的变量被积分或求和给弄掉了，所以也就变不回来了. 而后文的条件分布和边际分布因此就是两码事了.

定理 2.20 若 $(X,Y)\sim(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则 $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$.

证明 证明很简单，凑出对 x 或者 y 的积分就可以得到. ■

可以发现，二元降至一元之后，参数 ρ 不见了，这就是边际概率密度无法决定联合概率密度的一个具体例子.

2.5　条件分布和随机变量的独立性

2.5.1　条件分布

离散型随机变量的条件分布

定义 2.30 设二维离散型随机向量 X, Y 的概率函数形式为 $p_{ij} = P(X = x_i, Y = y_j), i = 1, 2, ..., m; j = 1, 2, ..., n$. 对给定的事件 $\{Y = y_j\}$，若其概率 $P(Y = y_j) > 0$，则称

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}} \tag{2.5.1}$$

为给定 $Y = y_j$ 条件下的 X 的分布律. 式中 $p_{\cdot j}$ 代表对第一个下标 $(x_i, i = 1, 2, ..., m)$ 求和的概率，也即 y_j 的边际概率.

例 2.7 设 $X = (x_1, x_2, ..., x_n) \sim M(N; p_1, p_2, ..., p_n)$，试求 X_1 在给定 $X_2 = k$ 的条件下的条件分布律.

解 要明确的是，对于多项分布的其中的分量，满足

$$(x_1, x_2) \sim M(N; p_1, p_2, 1 - p_1 - p_2) \tag{2.5.2}$$

那么 $P(x_1 = i, x_2 = j) = \dfrac{N!}{i!j!(N-i-j)!} p_1^i p_2^j (1 - p_1 - p_2)^{N-i-j} (0 \le i, j, i + j \le N)$. 然后还有一个显然的结论，就是对于多项分布的每一个分量，$x_2 \sim B(N, p_2)$. 因此套用公式得到 $P(x_1 = i | x_2 = k) = \dfrac{(N-k)!}{i!(N-k-i)!} \left(\dfrac{p_1}{1-p_2}\right)^i \left(1 - \dfrac{p_1}{1-p_2}\right)^{N-k-i}, i = 0, 1, ..., N-k$(注意后面的取值是要写的).

连续型随机变量的条件分布

定义 2.31 连续型随机向量 (X, Y) 中的 X 在给定条件 $Y = y$ 下的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, f_Y(y) > 0. \tag{2.5.3}$$

此时记为

$$X|y \sim f_{X|Y}(x|y). \tag{2.5.4}$$

例 2.8 设随机向量 (X,Y) 服从 $\{(x,y)\mid |x+y|\le 1, |x-y|\le 1\}$ 内的均匀分布，

(1) 试求出 X 和 Y 的边缘分布.

(2)X 和 Y 是否相互独立?

(3) 求在 $X=a(0<a<1)$ 时 Y 的条件密度函数.

解 (1) 显然 $f(x,y)=\frac{1}{2}I(D)$. 那么

$$f_X(x)=\begin{cases}\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\int_{-x-1}^{1+x}\frac{1}{2}\mathrm{d}y=1+x, x\in[-1,0]\\ \int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y=\int_{x-1}^{1-x}\frac{1}{2}\mathrm{d}y=1-x, x\in[0,1]\end{cases}=1-|x| \tag{2.5.5}$$

故 $f_Y(y)=1-|y|$.

(2) 由于 $f(x,y)\ne f_X(x)f_Y(y)$，所以不独立.

(3)$f_{Y|X}(y|x)=\dfrac{f(x,y)}{f_X(x)}=\dfrac{\frac{1}{2}}{1-x}=\dfrac{1}{2-2x}$.

一般来说，只要注意积分域的问题，就不太有必要标出示性函数来. 但是在每个函数的结果后面是一定要标出取值范围的.

一般的高维情形

定义 2.32 不论是对于离散型还是连续型随机向量 $(X_1,X_2,...,X_n)$ 以及其中的分量 $(X_{i_1},X_{i_2},...,X_{i_m}), m<n$，若有

$$(X_1,X_2,...,X_n)\sim f(x_1,x_2,...,x_n)\quad (X_{i_1},X_{i_2},...,X_{i_m})\sim g(x_{i_1},x_{i_2},...,x_{i_m}) \tag{2.5.6}$$

那么定义在 $(X_{i_1},X_{i_2},...,X_{i_m})=(x_{i_1},x_{i_2},...,x_{i_m})$ 条件下 $(X_{j_1},X_{j_2},...,X_{j_{n-m}})=(x_{j_1},x_{j_1},...,x_{j_{n-m}})$ 的条件密度为

$$h(x_{j_1},x_{j_1},...,x_{j_{n-m}}|x_{i_1},x_{i_2},...,x_{i_m})=\frac{f(x_1,x_2,...,x_n)}{g(x_{i_1},x_{i_2},...,x_{i_m})} \tag{2.5.7}$$

2.5.2　随机变量的独立性

定义 2.33 称离散型随机变量 $X_1,X_2,...,X_n$ 相互独立，若它们的联合分布律等于各自边缘分布律的乘积

$$P(X_1=x_1,X_2=x_2,...,X_n=x_n)=P(X_1=x_1)P(X_2=x_2)\cdots P(X_n=x_n) \tag{2.5.8}$$

定义 2.34 称连续型随机变量 $X_1, X_2, ..., X_n$ 相互独立，若它们的联合分布律等于各自边缘分布律的乘积

$$f(x_1, x_2, ..., x_n) = f(x_1)f(x_2)\cdots f(x_n), \forall (x_1, x_2, ..., x_n) \in \mathbb{R}^n. \tag{2.5.9}$$

从上面定义可以看出，$X_1, X_2, ..., X_n$ 相互独立便意味着其中部分分量 $X_{i_1}, X_{i_2}, ..., X_{i_m}(m < n)$ 相互独立；但 X_j, X_k 两两独立并不能推出 $X_1, X_2, ..., X_n$ 相互独立.

定理 2.21 设 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$，则 (X, Y) 独立当且仅当 $\rho = 0$.

例 2.9 证明：

(1) 若 $(x_1, x_2, ..., x_n)$ 服从矩域 $E : [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ 上的均匀分布，则 $x_1, x_2, ..., x_n$ 相互独立.

(2) 若 $(x_1, x_2, ..., x_n)$ 服从 n 维单位球上的均匀分布，则 $x_1, x_2, ..., x_n$ 不独立.

解 (1) 易知

$$f(x_1, x_2, ..., x_n) = \frac{1}{(b_1 - a_1)(b_2 - a_2)\cdots(b_n - a_n)}, (x_1, x_2, ..., x_n) \in E. \tag{2.5.10}$$

且[1]

$$\forall x_i, f(x_i) = \int_{a_1}^{b_1} \cdots \int_{a_{i-1}}^{b_{i-1}} \int_{a_{i+1}}^{b_{i+1}} \cdots \int_{a_n}^{b_n} f(x_1, x_2, ..., x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n = \frac{1}{b_i - a_i} \tag{2.5.11}$$

显然满足独立性条件.

(2) 易知

$$f(x_1, x_2, ..., x_n) = \frac{1}{\pi}, (x_1, x_2, ..., x_n) \in B(1). \tag{2.5.12}$$

但

$$\begin{aligned} &f(x_i) \\ &= \int_{-1}^{1} \mathrm{d}x_1 \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \mathrm{d}x_2 \cdots \int_{-\sqrt{1-x_1^2-x_2^2-\cdots-x_{n-1}^2}}^{\sqrt{1-x_1^2-x_2^2-\cdots-x_{n-1}^2}} f(x_1, x_2, ..., x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \end{aligned}$$

[1] 以后为了书写简洁，规定积分号的顺序对应被积变量的顺序.

显然这个积分是无法化成单独的积分域的. 因为考虑到各个分量的均匀性，要求 $\forall i, f(x_i) = \dfrac{1}{\sqrt[n]{\pi}}$，这显然不是上面的积分结果.

特别地，考虑单位圆上 (X,Y) 的均匀分布. 显然

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,y)\mathrm{d}y = \frac{2}{\pi}\sqrt{1-x^2} \tag{2.5.13}$$

同理 $F_Y(y) = \dfrac{2}{\pi}\sqrt{1-y^2}$. 则 $f_X(x)f_Y(y) = \dfrac{4}{\pi^2}\sqrt{(1-x^2)(1-y^2)} \neq \dfrac{1}{\pi}$.

定理 2.22 设有 n 个事件 $A_1,...,A_n$，定义 $X_i = I_{A_i}$ 为 A_i 的示性函数，则 $A_1,...,A_n$ 独立当且仅当 $X_1,...,X_n$ 独立.

定义 2.35 若 $(x_1,x_2,...,x_n)$ 的概率密度函数可以分解成

$$f(\boldsymbol{X}) = f(x_1,...,x_n) = g(x_{i_1},...,x_{i_m})h(x_{j_1}...x_{j_{n-m}}) = g(\boldsymbol{Y})h(\boldsymbol{Z}) \tag{2.5.14}$$

的形式，则称 $\boldsymbol{Y},\boldsymbol{Z}$ 相互独立.

定理 2.23 若 X_1,X_2 独立，且 $Y = g(X_1), Z = h(X_2)$. 则 Y,Z 独立.

注意上述命题不可逆.

2.6 随机变量函数的概率分布

2.6.1 离散型随机变量函数的概率分布

非常简单，要做的就是将变换 $Y = f(X)$ 之后将相同的 Y_i 合并，对应的 p_i 合并即可. 对于离散型随机向量也是同理，不过向量相等当且仅当各个分量相等.

定理 2.24 当 ξ,η 是相互独立的非负整值随机变量，且各有分布律 $\{a_k\},\{b_k\}$. 那么 $\zeta = \xi+\eta$ 有分布律

$$P(\zeta = n) = P(\xi+\eta = n) = \sum_{k=0}^{n} a_k b_{n-k}. \tag{2.6.1}$$

上面的公式称为离散卷积公式. 它的含义是将 n 在 ξ,η 中的全部分配遍历一遍.

定义 2.36 若对于某个分布 $X \sim D(\lambda_1, \lambda_2, ..., \lambda_k)$，关于某个参数 λ_i 有：若 $X_1 \sim D(\lambda_1, ..., (i\text{–th})\mu_1, ..., \lambda_k), ..., X_m \sim D(\lambda_1, ..., (i\text{–th})\mu_m, ..., \lambda_k)$ 且

$$\sum_{j=1}^{n} X_j \sim D(\lambda_1, ..., (i\text{–th}) \sum_{j=1}^{n} \mu_j, ..., \lambda_k) \tag{2.6.2}$$

则称分布 D 关于参数 λ_i 具有再生性.

定理 2.25 二项分布关于试验次数有再生性. 也即：对于 $X \sim B(m, p), Y \sim B(n, p)$，有 $X + Y \sim B(m + n, p)$. 对于多个服从相同概率参数的二项分布的变量的场合也是同理.

证明 证明很简单，只需要用离散卷积公式即可.$P(X + Y = k) = \sum_{j=0}^{k} \binom{m}{j} p^j q^{m-j} \binom{n}{k-j} p^{k-j} q^{n-(k-j)} = \binom{m+n}{k} p^k q^{m+n-k}$. ■

定理 2.26 (Poisson 分布关于参数具有再生性). 也即，若 $X \sim P(\lambda), Y \sim P(\mu)$，则 $X + Y \sim P(\lambda + \mu)$.

证明 证明过程仍然是运用离散卷积公式.$P(X+Y = n) = \sum_{k=0}^{n} \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \frac{\mathrm{e}^{-\mu}\mu^{n-k}}{(n-k)!} = \frac{\mathrm{e}^{-(\lambda+\mu)}(\lambda+\mu)^n}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} = \frac{\mathrm{e}^{-(\lambda+\mu)}(\lambda+\mu)^n}{n!} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{\lambda}{\lambda+\mu}\right)^k \left(\frac{\mu}{\lambda+\mu}\right)^{n-k} = \frac{\mathrm{e}^{-(\lambda+\mu)}(\lambda+\mu)^n}{n!}$.■

定理 2.27 (Pascal 分布关于成功次数 r 具有再生性) 若独立的 X, Y 满足 $X \sim NB(r_1, p), Y \sim NB(r_2, p)$，则

$$X + Y \sim NB(r_1 + r_2, p) \tag{2.6.3}$$

对于多个变量的和结论仍然成立.

2.6.2 连续型随机变量函数的概率分布

定理 2.28 (密度变换公式) 若随机变量 X 有概率密度函数 $f(x), x \in (a, b)$(a, b 可以为无穷)，而 $y = g(x)$ 在 $x \in (a, b)$ 上是严格单调的连续函数，存在唯一的反函数 $x = h(y), y \in (\alpha, \beta)$ 且 $h'(y)$ 存在且连续. 那么 $Y = g(X)$ 的概率密度函数

$$f_Y(y) = f_X(x)|h'(y)| = f_X(h(y))|h'(y)|. \tag{2.6.4}$$

这种方法的一般性表述为：利用

$$
\begin{aligned}
&F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le h(y)) = F_X(h(y)) \\
\Rightarrow &f_Y(y) = f_X(h(y))|h'(y)|.
\end{aligned} \tag{2.6.5}
$$

式中绝对值是为了保证概率密度函数恒非负. 注意，其中的 x, y 是函数的自变量，而 X, Y 才表示随机变量，因而可以直接使用 $Y = g(X)$(注意不要弄混了). 对于 g 在全取值域上不单调的情况，一般来说可以使用上面逐步递推的方式求出概率密度函数来，当然，也可以使用下面的定理

定理 2.29 若随机变量 X 有概率密度函数 $f(x), x \in (a, b)(a, b$ 可以为无穷)，且存在分割 T 使得 (a, b) 被分为有限个或可列个互不重叠的子区间 W_i，并且 $u = g(t)$ 分别在每个子区间上严格单调 (不同 W_i 单调性可以不一致) 且连续，则对于随机变量 $Y = g(X)$，在每个区间 W_i 上存在唯一的反函数 $X = h(Y)$，那么在 (α, β) 上 Y 的概率密度函数为各区间结果的和

$$
f_Y(y) = \sum_{i=1}^{n} f_X(h_i(y))|h_i'(y)| \tag{2.6.6}
$$

当然要注意，$Y = g(X)$ 也是有取值范围的，取决于 $g(X)$ 的定义域和值域. 事先一定要标定好其取值范围. 上面的定理的分割是按照 $g(x)$ 的单调性的，而与 f 无关，在使用时务必留意，不要混淆. 而且因为这涉及到变量的转换，所以要预先求出范围，不必使用示性函数了.

以下是几个经典的函数分布：

- Cauchy 分布. 若 $X \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right), Y = \tan X$，则 Y 服从 Cauchy 分布，其概率密度函数为 $f(x) = \frac{1}{\pi(1 + x^2)}$.
- 对数正态分布. 若 $X \sim N(\mu, \sigma^2)$，则 $Y = \mathrm{e}^X \sim LN(\mu, \sigma^2)$.
- Weibull 分布. 若 $X \sim \mathrm{Exp}(\lambda), Y = X^\alpha(\alpha > 0)$，则 Y 服从 Webill 分布.

例 2.10 设 $X \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. 求 $Y = \cos X$ 的概率密度函数.

解 首先要写出 $y \sim (0, 1]$. 发现 $u = g(t) = \cos t$ 是分段单调的，且反函数每段的求导结果的绝对值均为 $\frac{1}{\sqrt{1 - y^2}}$，那么有 $f_Y(y) = \frac{1}{\pi}\frac{2}{\sqrt{1 - y^2}}, y \in (0, 1]$.

例 2.11 若 $X \sim N(0, 1)$，求 $Y = 2X^2 + 1$ 的概率密度函数.

解 注意到 $y \in [1, +\infty)$，且它是分段单调的

$$h(y) = \begin{cases} -\sqrt{\dfrac{y-1}{2}}, y < 0 \\ \sqrt{\dfrac{y-1}{2}}, y \geq 0 \end{cases} \tag{2.6.7}$$

但是在这里是无伤大雅的，因为平方了 (其他情况下可就不一定了) 而且因为只有正负号的差别，所以导数取了绝对值之后是一样的，所以结果的加和只需要代以乘以 2 即可. 代入计算得到结果为 $\dfrac{1}{2\sqrt{\pi(y-1)}}\mathrm{e}^{(1-y)/4}$.

例 2.12 设随机变量 X 服从参数为 λ 的指数分布，以 $Y = \lfloor X \rfloor$ 表示它的整数部分，即不超过 X 的最大整数, 而以 Z 表示它的小数部分, 即 $Z = X - \lfloor X \rfloor$. 试求随机变量 Y 和 Z 各自的分布, 且它们是否相互独立?

解 显然 Y 是离散型随机变量，取值为 $N = 0, 1, 2, ...$

$$P(Y = N) = P(N \leq X \leq N+1) = F(N+1) - F(N) = \mathrm{e}^{-\lambda N}\left(1 - \mathrm{e}^{-\lambda}\right) \tag{2.6.8}$$

而 Z 的分布范围为 $t \in [0, 1)$

$$P(0 \leq Z \leq t) = \sum_{N=0}^{\infty} P(N \leq X \leq N+t) = \sum_{N=0}^{\infty} F(N+t) - F(N) = \frac{1 - \mathrm{e}^{-\lambda t}}{1 - \mathrm{e}^{-\lambda}} \tag{2.6.9}$$

且

$$P(Y = N, 0 \leq Z \leq t) = P(N \leq X \leq N+t) = \mathrm{e}^{-\lambda N}\left(1 - \mathrm{e}^{-\lambda t}\right) = P(Y = N)P(0 \leq Z \leq t) \tag{2.6.10}$$

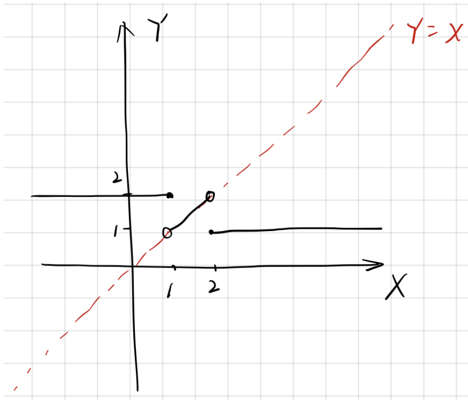
故为独立的.

例 2.13 设随机变量 X 的密度函数为 $f(x) = \dfrac{x^2}{a}, 0 < x < 3$,令随机变量 $Y = \begin{cases} 2, X \leq 1 \\ X, 1 < X < 2 \\ 1, X \geq 2 \end{cases}$

(1) 求随机变量 Y 的分布函数.

(2) 求概率 $P(X < Y)$.

解 这一题的目的是告诉我们分清随机变量本身和函数自变量的关系. 首先确定 a 的值为 9. 画出 $Y = g(X)$ 图像如图所示.



(1) 那么 $F(y) = P(Y \leq y) = P(g(X) \leq y)$.

(a)$y \leq 1, P(g(X) \leq y) = 0$.(注意到此时 Y 的意义就是 $g(X)$)

(b)$1 < y < 2, P(g(X) \leq y) = P(X \geq 2) + P(1 < X \leq y) = \dfrac{19}{27} + \dfrac{y^3 - 1}{27}$.

(c)$y \geq 2, P(g(X) \leq y) = 1$.

故 $F_Y(y) = \begin{cases} 0, y \leq 1 \\ \dfrac{19}{27} + \dfrac{y^3-1}{27}, 1 < y < 2 \\ 1, y \geq 2 \end{cases}$

(2)$P(X \leq Y) = P(X \leq g(X)) = P(X \leq 2) = \dfrac{8}{27}$.

我们在求解的过程中可能会有疑惑，这样是不是有点麻烦了，况且貌似没什么普适性? 事实上，我们有如下定理.

定理 2.30 考虑连续型随机变量 X, Y 的不等式约束 $\varphi(X, Y) < 0$(这里换成其他不等号是一样的)，且 L_1 为 $f_X(x)$ 的定义域，L_2 为 $f_Y(y)$ 的定义域 (因为默认用 f 表示概率密度，所以一般的函数不要再用 f 表示了)，D 为不等式 $\varphi(x, y) < 0$ 成立的区域. 那么

$$P(\varphi(X, Y) < 0) = \iint_{(L_1 \times L_2) \cap D} f_X(x) f_Y(y) \mathrm{d}x \mathrm{d}y \tag{2.6.11}$$

但是，例2.13是不能用这个办法的. 因为 Y 并不是连续型随机变量. 值得一提的是，以上假定了 X, Y 独立. 若 X, Y 是否独立是未知的，也即给定联合分布 $f(x, y)$，那么定理就改变成

定理 2.31 考虑连续型随机向量 (X, Y) 中的一个关于 X, Y 的不等式约束 $\varphi(X, Y) < 0$(这里换成其他不等号是一样的)，且 D_0 为 $f(x, y)$ 的定义域 (因为默认用 f 表示概率密度，所以一般的函数不要再用 f 表示了)，D 为不等式 $\varphi(x, y) < 0$ 成立的区域. 那么

$$P(\varphi(X,Y)<0)=\iint_{D_0\cap D} f(x,y)\mathrm{d}x\mathrm{d}y \tag{2.6.12}$$

以上我们考虑的是 $\mathbb{R}\to\mathbb{R}$ 的映射. 事实上，若考虑 $\mathbb{R}^n\to\mathbb{R}$，甚至是 $\mathbb{R}^n\to\mathbb{R}^n$ 的映射，结果也是类似的.

定理 2.32 对于 n 维随机变量 $(X_1,X_2,...,X_n)$，如果存在在给定域上严格单调的映射 $g_i(X_1,X_2,...,X_n)(i=1,2,...,n):\mathbb{R}^n\to\mathbb{R}$，且 $Y_i=g_i$ 组成 n 维随机变量 $(Y_1,Y_2,...,Y_n)$，那么 g_i 存在反函数 $X_i=h_i(Y_1,Y_2,...,Y_n)$，且 $(Y_1,Y_2,...,Y_n)$ 的概率密度函数为

$$f_{\boldsymbol{Y}}(y_1,y_2,...,y_n)=f_{\boldsymbol{X}}(h_1(y_1,y_2,...,y_n),...,h_n(y_1,y_2,...,y_n))|\det(\boldsymbol{J}_{\boldsymbol{Y}}(\mathscr{H}))| \tag{2.6.13}$$

其中变换 $\mathscr{H}$ 为

$$\mathscr{H}:\mathbb{R}^n\to\mathbb{R}^n \quad \boldsymbol{Y}\mapsto\boldsymbol{X}=(h_1(\boldsymbol{Y}),...,h_n(\boldsymbol{Y})) \tag{2.6.14}$$

意即

$$\det(\boldsymbol{J}_{\boldsymbol{Y}}(\mathscr{H}))=\begin{vmatrix}\dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} & \cdots & \dfrac{\partial h_1}{\partial y_n}\\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} & \cdots & \dfrac{\partial h_2}{\partial y_n}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{\partial h_n}{\partial y_1} & \dfrac{\partial h_n}{\partial y_2} & \cdots & \dfrac{\partial h_n}{\partial y_n}\end{vmatrix} \tag{2.6.15}$$

特别地，对 $(y_1,y_2)\mapsto(x_1,x_2)$ 的情况，有

$$q(y_1,y_2)=p(h_1(y_1,y_2),h_2(y_1,y_2))\left|\det\begin{pmatrix}\dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2}\\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2}\end{pmatrix}\right| \tag{2.6.16}$$

对于 $\mathbb{R}^2\to\mathbb{R}$ 的映射，我们有一些特殊的定理.

定理 2.33 设 X,Y 的联合概率密度为 $f(x,y)$，则 $Z=X+Y$ 的概率密度 $p(z)$ 为

$$p(z)=\int_{-\infty}^{+\infty} f(x,z-x)\mathrm{d}x=\int_{-\infty}^{+\infty} f(z-y,y)\mathrm{d}y \tag{2.6.17}$$

证明 证明很简单，考虑到积分域 $x+y\leq z$ 即可. 这是定理2.31的一个直接应用. ■

定义 2.37 定义两个在 $\mathbb{R}$ 上可积且绝对可积的函数 $f(x), g(x)$ 的卷积为

$$f * g = \int_{-\infty}^{+\infty} f(x-t)g(t)\mathrm{d}t. \tag{2.6.18}$$

那么，对于两个独立的随机变量，我们就有如下结论

定理 2.34 对定理2.33中的 X, Y，若它们独立，则有

$$p(z) = p(z) = \int_{-\infty}^{+\infty} g(x)h(z-x)\mathrm{d}x = \int_{-\infty}^{+\infty} g(z-y)h(y)\mathrm{d}y = g * h = h * g \tag{2.6.19}$$

上面的公式又称为卷积公式. 要注意的是，这是对 $Z = X+Y$ 而言的. 如果是 $Z = X-Y$，上面这个式子就要进行适当的变换，也即

$$p(z) = \int_{-\infty}^{+\infty} f(x, x-z)\mathrm{d}x = \int_{-\infty}^{+\infty} f(z+y, y)\mathrm{d}y \tag{2.6.20}$$

例 2.14 设 X 和 Y 相互独立且分别服从均值为 1 和 $\frac{1}{4}$ 的指数分布，求 $P(X<Y)$.

解 这里我们分别使用定理2.31,2.33的结果做一遍.

先写出 $f_X(x) = \mathrm{e}^{-x}I(x>0), f_Y(y) = 4\mathrm{e}^{-4y}I(y>0)$, 那么

(A)

$$\begin{aligned} P(X<Y) &= \iint_{y>x>0} \mathrm{e}^{-x}I(x>0)\cdot 4\mathrm{e}^{-4y}I(y>0)\mathrm{d}x\mathrm{d}y \\ &= \iint_{y>x>0} 4\mathrm{e}^{-(x+4y)}\mathrm{d}x\mathrm{d}y = \frac{1}{5} \end{aligned} \tag{2.6.21}$$

(B) 令 $Z = X-Y$，则 $P(X<Y) = P(Z<0)$. 先求 $f_Z(z)$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x)f_Y(x-z)\mathrm{d}x = \int_{-\infty}^{+\infty} \mathrm{e}^{-x}I(x>0)\cdot 4\mathrm{e}^{-4(x-z)}I(x>z)\mathrm{d}x \\ &= \int_{-\infty}^{+\infty} 4\mathrm{e}^{4z-5x}I(x>\max\{z,0\})\mathrm{d}x \\ &= \begin{cases} \int_0^{+\infty} 4\mathrm{e}^{4z-5x}\mathrm{d}x = \frac{4}{5}\mathrm{e}^{4z}, z \le 0 \\ \int_z^{+\infty} 4\mathrm{e}^{4z-5x}\mathrm{d}x = \frac{4}{5}\mathrm{e}^{-z}, z > 0 \end{cases} \end{aligned} \tag{2.6.22}$$

故

$$P(Z<0)=\int_{-\infty}^{0} f_Z(z)\mathrm{d}z=\frac{1}{5} \tag{2.6.23}$$

相比之下的确是 (A) 方法更加简单. 但是如果要求 Z 的密度函数，那么只能使用 (B) 方法了.

对于两个随机变量之商，我们也有类似的结论.

定理 2.35 对二维连续型随机变量 (ξ,η)，其联合密度为 $f(x,y)$，那么其商 $\frac{\xi}{\eta},\frac{\eta}{\xi}$ 的密度函数为

$$p_{\xi/\eta}(x)=\int_{-\infty}^{+\infty} f(xt,t)|t|\mathrm{d}t \tag{2.6.24}$$

$$p_{\eta/\xi}(x)=\int_{-\infty}^{+\infty} f(t,xt)|t|\mathrm{d}t \tag{2.6.25}$$

可以简记为“谁是分子谁多要”.

证明 这个定理是可以现场推出来的,不需要记忆.只需要考虑映射 $\begin{cases} Z=\dfrac{X}{Y}, \\ W=Y \end{cases}$,也即 $\begin{cases} X=ZW, \\ Y=W \end{cases}$, 再使用定理2.32即可. 对 Y/X 也是同样的办法. ■

根据定义2.36中所述，连续型随机变量的分布也具有再生性.

定理 2.36 (正态分布关于 μ,σ^2 具有再生性) 若独立的 X,Y 有 $X\sim N(\mu_1,\sigma_1^2),Y\sim N(\mu_2,\sigma_2^2)$，那么

$$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2) \tag{2.6.26}$$

这个结果可以推广到多个变量的和的形式.更一般地,对 $\forall i,X_i\sim N(\mu_i,\sigma_i^2)$,且 $X=\sum_{i=1}^{n}a_iX_i+b$，有 $X\sim N\left(\sum_{i=1}^{n}a_i\mu_i+b\,,\,\sum_{i=1}^{n}a_i^2\sigma_i^2\right)$.

2.6.3 最值分布

对于 n 个随机变量 $\xi_1,...,\xi_n$，我们可以考虑其最大值和最小值

$$\eta_1=\max\{\xi_1,...,\xi_n\} \tag{2.6.27}$$

$$\eta_2=\min\{\xi_1,...,\xi_n\} \tag{2.6.28}$$

显然这样定义的 η_1, η_2 也是随机变量.现在令 $\xi_1, ..., \xi_n$ 是独立的,其分布函数分别为 $F_1(x), ...,$ $F_n(x)$，那么我们利用事件的运算关系

$$\{\eta_1 \leq x\} = \bigcap_{k=1}^{n} \{\eta_k \leq x\} \tag{2.6.29}$$

$$\{\eta_2 > x\} = \bigcap_{k=1}^{n} \{\eta_k > x\} \tag{2.6.30}$$

就可以得到

$$F_{\eta_1}(x) = P(\eta_1 \leq x) = \prod_{k=1}^{n} P(\xi_k \leq x) = \prod_{k=1}^{n} F_k(x) \tag{2.6.31}$$

$$F_{\eta_2}(x) = P(\eta_2 \leq x) = 1 - P(\eta_2 > x) = 1 - \prod_{k=1}^{n} P(\xi_k > x) = 1 - \prod_{k=1}^{n} (1 - F_k(x)) \tag{2.6.32}$$

特别地，当 $\xi_1, ..., \xi_n$ 独立同分布，也即 $\forall k, F_k = F$ 的时候，上式改写为

$$F_{\eta_1}(x) = F^n(x) \tag{2.6.33}$$

$$F_{\eta_2}(x) = 1 - (1 - F(x))^n \tag{2.6.34}$$

那么概率密度函数

$$f_{\eta_1}(x) = nF^{n-1}(x)f(x) \tag{2.6.35}$$

$$f_{\eta_2}(x) = n\left(1 - F(x)\right)^{n-1} f(x) = n\overline{F}^{n-1}(x)f(x) \tag{2.6.36}$$

其中定义了生存函数 $\overline{F}(x) = 1 - F(x)$.

接下来考虑更复杂的情况. 在独立同分布时，考虑 η_1, η_2 的联合分布. 利用 $P(A) = P(AB) + P(A\overline{B})$ 的变形形式，可以得到

$$\begin{aligned} P(\eta_2 \leq x, \eta_1 \leq y) &= P(\eta_1 \leq y) - P(\eta_2 > x, \eta_1 \leq y) \\ &= P\left(\bigcap_{k=1}^{n} \{\eta_k \leq y\}\right) - P\left(\bigcap_{k=1}^{n} \{x < \eta_k \leq y\}\right) \\ &= F^n(y) - (F(y) - F(x))^n \end{aligned} \tag{2.6.37}$$

当然，得考虑到这个联合分布是有定义域的，也就是 $x \leq y$，所以要在最后的概率密度函数上加上示性函数. 也即

$$f_{\eta_2,\eta_1}(x,y) = \frac{\partial^2}{\partial y \partial x}\left(F^n(y) - (F(y)-F(x))^n\right) = n(n-1)\left(F(y)-F(x)\right)^{n-2} f(x)f(y)I_{(x\leq y)} \tag{2.6.38}$$

对于两个随即变量的最值分布，有个 trivial 的结论用来验算

$$\max a, b + \min a, b = a + b \tag{2.6.39}$$

第三章　数字特征

3.1　中心特征

3.1.1　均值与期望

定义 3.1 对于离散型随机变量 X，其分布律为 $\{X_i : p_i\}$. 若它的概率对对应取值的加权级数绝对收敛，也即

$$\sum_{k=1}^{n} |X_k| p_k \leq +\infty \tag{3.1.1}$$

那么这个随机变量就可以定义数学期望

$$E(X) = \sum_{k=1}^{n} X_k p_k \tag{3.1.2}$$

定义 3.2 对连续型随机变量 X，若 $X \sim f(x)$，且 $f(x)$ 是绝对可积的，也即

$$\int_{-\infty}^{+\infty} |x| f(x) \mathrm{d}x \leq +\infty \tag{3.1.3}$$

那么这个随机变量就可以定义数学期望

$$E(X) = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x \tag{3.1.4}$$

注意，数学期望可以定义的前提是绝对收敛和绝对可积. 这是至关重要的.

例 3.1 试问 $P\left(X = \dfrac{(-1)^k 2^{k-1}}{k}\right) = \dfrac{1}{2^k}, k = 1, 2, ...$ 是否可以求数学期望?

解 因为 $\sum_{k=1}^{\infty}\frac{2^k}{k}\frac{1}{2^k}=\sum_{k=1}^{\infty}\frac{1}{k}=+\infty$，显然这是发散的. 而我们可以求得“$E(X)$”(这里打引号是因为压根就没有定义)$=\sum_{k=1}^{n}\frac{(-1)^k}{k}=-\ln 2$，可以知道这是条件收敛的级数，所以即便收敛也无法定义数学期望.

例 3.2 证明：Cauchy 分布是没有期望的.

解 显然 $\int_{-\infty}^{+\infty}\frac{2x}{\pi(1+x^2)}\mathrm{d}x=+\infty$. 虽然 $\int_{-\infty}^{+\infty}\frac{1}{\pi(1+x^2)}\mathrm{d}x=1$

下面讨论几个常见分布的数学期望.

定理 3.1 若 $X\sim B(n,p)$，则 $E(X)=np$.

证明 用定义证明即可. 其中重要的步骤是将组合数展开，再提出 np，即可凑出新的组合数的形式. ■

定理 3.2 若 $X\sim P(\lambda)$，则 $E(X)=\lambda$.

证明 用定义证明即可. 其中重要的步骤是将 e^x 展开，再提出 λ，即可凑出新的 Poisson 分布形式. ■

例 3.3 (超几何分布) 一个箱子里有 n 个除了颜色外相同的球，其中 r 个红球，$n-r$ 个黑球. 从中随机摸 $m>r$ 个球，其中红球个数记为 X，求 $E(X)$.

解 $P(X=k)=\frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$. 那么 $E(X)=\sum_{k=0}^{r}kP(X=k)=\frac{mr}{n}$，其中使用了

$$\binom{r}{k}=\frac{r}{k}\binom{r-1}{k-1}. \tag{3.1.5}$$

例 3.4 (几何分布) 求几何分布 $X\sim G(p),P(X=k)=q^{k-1}p$ 的期望.

解 容易得到 $EX=\sum_{k=1}^{\infty}kp(1-p)^{k-1}$. 我们可以采用求和 $S_n=\sum_{k=n}^{\infty}kp(1-p)^{k-1}$ 后再取极限得到. 这个求和是容易的，因为是等差数列乘等比数列所以只需要错位相减即可. 最后得到 $EX=\frac{1}{p}$.

例 3.5 (负二项分布) 求负二项分布 $X\sim NB(r,p),P(X=k)=C_{k-1}^{r-1}p^r(1-p)^{k-r}$ 的数学期望.

解 求这个的数学期望既可以根据定义式求，也可以采用事件分解的思想.

方法一　因为求和范围上限是正无穷，所以自然是很难采用凑出二项式展开的方法计算的. 那么，我们不妨利用递推式

$$
\begin{aligned}
E\left(X^n\right) &= \sum_{k=r}^{\infty} k^n C_{k-1}^{r-1} p^r (1-p)^{k-r} = \sum_{k=r}^{\infty} k^n \frac{(k-1)!}{(r-1)!(k-r)!} p^r (1-p)^{k-r} \\
&= \frac{r}{p} \sum_{k=r}^{\infty} k^{n-1} \frac{k!}{r!(k-r)!} p^{r+1} (1-p)^{k-r} = \frac{r}{p} \sum_{k=r}^{\infty} k^{n-1} C_k^r p^{r+1} (1-p)^{k-r} \\
&= \frac{r}{p} \sum_{k=r}^{\infty} k^{n-1} P(Y=k+1)
\end{aligned} \tag{3.1.6}
$$

其中 $Y \sim NB(r+1,p)$. 那么上式可以写为

$$
E\left(X^n\right) = \frac{r}{p} E\left((Y-1)^{n-1}\right). \tag{3.1.7}
$$

这样我们就可以利用它来计算期望了. 当 $n=1$ 时，$EX = \frac{r}{p}$.

方法二　将 $X \sim NB(r,p)$ 分解成 $X = I_1 + \cdots + I_r, \forall j = 1,...,r, I_j \sim G(p)$. 那么 $EX = \frac{r}{p}$.

定理 3.3 若 $X \sim \text{Exp}(\lambda)$，则 $E(X) = \frac{1}{\lambda}$.

证明 用定义证明即可. 其中用到了 $\Gamma(2)=1$ 以及 Γ 函数的收敛因子表达法. ■

定理 3.4 若 $X \sim U(a,b)$，则 $E(X) = \frac{a+b}{2}$.

定理 3.5 若 $X \sim N(\mu,\sigma)$，则 $E(X) = \mu$.

3.1.2　数学期望的性质

定理 3.6 数学期望是线性变换，即

$$
E(\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_n X_n) = \lambda_1 E(X_1) + \lambda_2 E(X_2) + \cdots + \lambda_n E(X_n) \tag{3.1.8}
$$

利用这个性质可以发现更本质的性质. 例如设 $X \sim B(n,p)$, 那么可以令 $I_i \sim B(1,p), i = 1,...,n$，则显然 $E(I_i) = p$，而 $E(X) = E(I_1 + \cdots + I_n) = E(I_1) + \cdots + E(I_n) = np$. 这样的做法是将一个事件分解成多个事件的和，在数学期望的计算中是常见的.

例 3.6 n 个人各有一顶帽子，每个人随机拿一顶，求平均拿对帽子的人数.

解 这可以分解成：拿对帽子的人数 = 拿到自己帽子的人数，那么对这 n 个人编号，记 $I_i = \begin{cases} 1, & \text{if one gets his hat} \\ 0, & \text{if not} \end{cases}$，这样就可以分解成每个人是否拿到自己帽子的情况了. 显然 $\forall i, P(I_i = 1) = \dfrac{1}{n}, E(I_i) = \dfrac{1}{n}, E(X) = E(I_1 + \cdots + I_n) = 1$.

定理 3.7 若 $X_1, X_2, ..., X_n$ 独立，则

$$E(X_1 ... X_n) = E(X_1) ... E(X_n) \tag{3.1.9}$$

对于经过函数变换的随机变量，数学期望的存在性验证和求法只不过是将变换后的变量看作是参与计算的随机变量，对应的概率函数 (概率密度函数) 还是一样的.

定理 3.8 对离散型随机变量 $P(X = a_i) = p_i, i = 1, ..., n$($n$ 可以为 ∞)，若

$$\sum_{i=1}^{n} |g(a_i)| p_i \leq +\infty \tag{3.1.10}$$

那么

$$E(g(X)) = \sum_{i=1}^{n} g(a_i) p_i \tag{3.1.11}$$

定理 3.9 对连续型随机变量 $X \sim f(x)$，若

$$\int_{-\infty}^{+\infty} |g(x)| f(x) \mathrm{d}x < +\infty \tag{3.1.12}$$

那么

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) \mathrm{d}x \tag{3.1.13}$$

例 3.7 将 1 m 棍子随机截成两段，求较短那段的长度的期望.

解 设 X 为其中一段长度，另一段为 $1 - X$. 显然 $X \sim U(0, 1)$. 这道题可能会想到用 min 分布变换来做，但其实不用那么麻烦. 因为 $g(X) = \min\{X, 1 - X\}$，那么 $E(g(X)) = \displaystyle\int_{-\infty}^{+\infty} g(x) I_{(0<x<1)} \mathrm{d}x = \frac{1}{4}$.

例 3.8 设随机变量 X 的期望存在，试证明

(1) 若 X 为可以取非负整值的随机变量，则

$$EX=\sum_{n=1}^{\infty}P(X\geq n)=\sum_{n=0}^{\infty}P(X>n) \tag{3.1.14}$$

(2) 若 X 为可以取非负值的连续型随机变量，则

$$EX=\int_0^{+\infty}(1-F(x))\,\mathrm{d}x \tag{3.1.15}$$

(3) 若 X 为非负随机变量，则 (2) 中的结论依然成立.

证明 (1)(2) 没有什么特别的，分别凑出期望的形式即可. 要注意的是，一般处理无穷级数是比较难的，所以不妨先规定离散型随机变量的取值上限为 n，那么自然就可以导出来结论了. 而 (2) 则是利用了不等式链在积分换序下不变的性质. ■

(3) 的证明则需要考虑间断点，而不能归于具体形式的推演了. 我们有如下定理

定理 3.10 若 $f(X)$ 为非负可测函数，那么

$$E\left(f(X)\right)=f\left(EX\right) \tag{3.1.16}$$

证明 让我们接着 (3) 的证明. 事实上

$$\begin{aligned}EX&=E\left(\int_0^X 1\mathrm{d}x\right)=E\left(\int_0^{+\infty}I(x\leq X)\mathrm{d}x\right)=\int_0^{+\infty}E\left(I(x\leq X)\right)\mathrm{d}x\\&=\int_0^{+\infty}P(X>x)\mathrm{d}x=\int_0^{+\infty}(1-F(x))\mathrm{d}x\end{aligned} \tag{3.1.17}$$

3.1.3　条件期望

定理 3.11 设 X,Y 为离散型随机变量，若在给定 $X=x_0$ 的情况下，Y 的分布律为 $P(Y=b_i|X=x_0)$，那么其条件期望 (已经假设了其可以定义，验证步骤略去)

$$E(Y|X=x_0)=\sum_{i=1}^{n}b_ip_i \tag{3.1.18}$$

定理 3.12 设 X,Y 为连续型随机变量, 若在给定 $X=x_0$ 的情况下, Y 的分布律为 $f_{Y|X}(y|x_0)$，那么其条件期望 (已经假设了其可以定义，验证步骤略去)

$$E(Y|X=x_0)=\int_{-\infty}^{+\infty}yf_{Y|X}(y|x_0)\mathrm{d}y \tag{3.1.19}$$

类似地，我们可以定义条件方差，这里不再举例.

上小节所述性质条件期望仍然满足. 计算条件期望的方法是，先求出条件概率密度函数，再看看这个函数是否可以凑成某个特殊的形式 (若不能就老老实实算). 一般情况下，先分析分析条件分布是否服从某分布规律，会比硬算省一些时间.

例 3.9 假设有 $n(n \geq 3)$ 个不同的盒子与 m 个相同的小球, 每个小球独立地以概率 p_r 落入第 r 个盒子 $(r=1,2,...,n)$. 分别以 $X_1,...,X_n$ 表示落入各个盒子的球数. 试求

(1)$E\left(X_2|X_1=k\right)$ 和 $\operatorname{Var}(X_2|X_1=k)$.

(2)$E\left(X_1+X_2\right)$ 和 $\operatorname{Var}(X_1+\cdots+X_r), r=1,..,n$.

解 (1) 如果套用公式的话，那么计算量会相当之大，毕竟是需要凑组合数的. 这里就要采用分析是哪种分布的办法. 可以发现，$X_2|X_1=k \sim B\left(m-k, \dfrac{p_2}{1-p_1}\right)$ 一定要注意，概率参数应该采用条件概率进行计算！那么显然 $E\left(X_2|X_1=k\right)=p_2\dfrac{m-k}{1-p_1}, \operatorname{Var}(X_2|X_1=k)=\dfrac{(m-k)p_2}{1-p_1}\dfrac{1-p_1-p_2}{1-p_1}$.

(2) 因为 X_1,X_2 是否独立无关加减期望的分解，所以直接得到

$$E(X_1+X_2)=E\left(X_1\right)+E\left(X_2\right)=m(p_1+p_2) \tag{3.1.20}$$

而 $X_1+\cdots+X_r \sim B(m,p_1+\cdots+p_r)$，故 $\operatorname{Var}(X_1+\cdots+X_r)=m(p_1+\cdots+p_r)(p_{r+1}+\cdots+p_n)$.

显然，条件期望 $E(Y|X=x)$ 为 x 的函数，那么我们将 x 换成 X，$E(Y|X)$ 就是一个随机变量了.

定理 3.13 (全期望公式) 设 X,Y 为两个随机变量，则

$$E(X)=E(E(X|Y)) \tag{3.1.21}$$

这在数学期望不易计算但条件期望容易计算的场合下有重要的作用.

例 3.10 一窃贼被关在有 3 个门的地牢里, 其中第一个门通向自由. 出这门走 3 个小时便可以回到地面; 第 9 个门通向另一个地道, 走 5 个小时将返回到地牢; 第 3 个门通向更长的地道, 走 7 个小时也回到地牢. 若窃贼每次选择 3 个门的可能性总相同, 求他为获得自由而奔走的平均时间.

解 设这个窃贼需要走 X 小时才回到地面，且 Y 表示他打开门的号数 (编号 1,2,3). 显然 Y 的每个取值均为 $\dfrac{1}{3}$. 所以

$$E(X)=E(E(X|Y))=\sum_{i=1}^{3}E(X|Y=i)p(Y=i) \tag{3.1.22}$$

显然 $i=1, E(X|Y=i)=3; Y=2, E(X|Y=i)=5+EX; i=3, E(X|Y=i)=7+EX$. 所以

$$E(X)=\frac{15+2E(X)}{3} \Rightarrow E(X)=15 \tag{3.1.23}$$

定理 3.14 设 X, Y 为两个随机变量，且 $g(X)$ 为绝对可积的随机变量，则

$$E(g(X))=E(E(g(X)|Y)) \tag{3.1.24}$$

这个公式可以胜任更复杂的情况. 也即先求解 $h(x)=E(Y|X=x)$，此时再将 $h(x)$ 中代 x 以 X，也就是 $X=x$ 的“逆向变换”(姑且这样理解)，这样，在已知 EX 的前提下，$E(h(X))$ 可以很便利地求出，那么 $E(Y)=E(E(Y|X))=E(h(X))$.

例 3.11 设 $(X,Y)\sim N(a,b,\sigma_1^2,\sigma_2^2,\rho)$，计算 $E(XY)$.

解 先求得 $E(XY|X=x)=E(X|X=x)E(Y|X=x)=x\left(b+\rho\frac{\sigma_2}{\sigma_1}(x-a)\right)$. 那么 $E(XY)=E(E(XY|X))=E\left(bX+\rho\frac{\sigma_2}{\sigma_1}X^2-\rho\frac{\sigma_2}{\sigma_1}aX\right)=ab+\rho\sigma_1\sigma_2$.

3.1.4　中位数

定义 3.3 对于离散型随机变量，若 $\exists k$, s.t.

$$\begin{cases} P(X\le k)\ge \dfrac{1}{2}, \\ P(X\ge k)\ge \dfrac{1}{2}. \end{cases} \tag{3.1.25}$$

则称 k 为 X 的中位数，记作 $\mathrm{Med}(X)=k$.

例 3.12 举例说明中位数可以不唯一.

解 对于两点分布 $X\sim\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$，显然对于 $\forall x\in[0,1]$，都有中位数的条件成立.

定义 3.4 对连续型随机变量 X，给定其概率密度函数为 $f(x)$，定义中位数为满足

$$P(X\le t)=\int_{-\infty}^{t} f(x)\mathrm{d}x=\frac{1}{2} \tag{3.1.26}$$

的 t，记作 $\mathrm{Med}(X)=t$.

定义 3.5 设 $0 < p < 1$，则称 μ_p 是随机变量 ξ 的 p 分位数，若

$$\begin{cases} P(\xi \le \mu_p) \ge p, \\ P(\xi \ge \mu_p) \ge 1 - p. \end{cases} \tag{3.1.27}$$

显然，中位数是 $\dfrac{1}{2}$ 分位数 $\mu_{0.5}$.

3.1.5　众数

定义 3.6 定义离散型随机变量 X 的众数为其最大取值概率点，也即为满足以下条件的 k

$$\forall j \ne i, P(X = x_i = k) \le P(X = x_j) \tag{3.1.28}$$

记作 $\mathrm{Mode}(X) = k$.

定义 3.7 定义连续型随机变量 X 的众数为其概率密度函数的最大值点.

3.2　方差、标准差与矩

3.2.1　方差与标准差

定义 3.8 设随机变量 X，则定义其方差为

$$\mathrm{Var}(X) = E((X - EX)^2) \tag{3.2.1}$$

展开上式的右边，可以得到一种等价的表述

$$\mathrm{Var}(X) = E(X^2) - (EX)^2 \tag{3.2.2}$$

定义 3.9 随机变量的标准差为其方差的算术平方根

$$\sigma = \sqrt{\mathrm{Var}(X)} \tag{3.2.3}$$

定理 3.15 方差有以下性质：

(1)$\mathrm{Var}(X) \ge 0$，因此 $E(X^2) \ge (EX)^2, E(X^2) \ge \mathrm{Var}(X)$.

(2)$\mathrm{Var}(X) = 0$ 当且仅当 $P(X = EX) = 1$. 此时我们称 X 退化到 EX.

(3)$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.

(4) 方差是对平方残差和的最小估计，也即 $\forall C$，有 $\mathrm{Var}(X) \le E((X-C)^2)$，其中等号成立当且仅当 $C = EX$.

(5) 一般地，我们并没有 $\mathrm{Var}(aX+bY) = a\mathrm{Var}(X) + b\mathrm{Var}(Y)$，这是因为

$$\begin{aligned}\mathrm{Var}(X+Y) &= E\left((X-EX+Y-EY)^2\right) \\ &= E\left((X-EX)^2\right) + E\left((Y-EY)^2\right) + 2E\left((X-EX)(Y-EY)\right) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) \end{aligned} \tag{3.2.4}$$

有最后一项交叉项的存在. 特别地，当 X, Y 独立时，才有 (类) 线性性

$$\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \tag{3.2.5}$$

数学期望是没有这个要求的.

以下是常见分布的方差.

定理 3.16 若 $X \sim B(n,p)$，则 $\mathrm{Var}(X) = npq = np(1-p)$.

定理 3.17 若 $X \sim P(\lambda)$，则 $\mathrm{Var}(X) = \lambda$.

例 3.13 (几何分布) 求几何分布 $X \sim G(p), P(X=k) = q^{k-1}p$ 的方差.

解 此处的难点在于求解 $E\left(X^2\right)$. 容易得到 $E\left(X^2\right) = \sum\limits_{k=1}^{\infty} k^2 p(1-p)^{k-1}$. 可以考虑分解的办法，也即

$$\begin{aligned}E\left(X^2\right) &= \sum_{k=1}^{\infty} k(k-1)pq^{k-1} + \sum_{k=1}^{\infty} kpq^{k-1} = pq\sum_{k=1}^{\infty} k(k-1)q^{k-2} + \frac{1}{p} \\ &= pq\sum_{k=1}^{\infty} \frac{\mathrm{d}^2 q^k}{\mathrm{d}q^2} + \frac{1}{p} = pq\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\sum_{k=1}^{\infty} q^k\right) + \frac{1}{p} = pq\frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\frac{1}{1-q} - 1\right) + \frac{1}{p} \\ &= \frac{2q}{p^2} + \frac{1}{p}\end{aligned} \tag{3.2.6}$$

之后不难求出 $\mathrm{Var}(X) = \dfrac{q}{p^2}$.

例 3.14 (负二项分布) 求负二项分布 $X \sim NB(r,p), P(X=k) = C_{k-1}^{r-1}p^r(1-p)^{k-r}$ 的方差.

解 求这个的方差既可以根据定义式求，也可以采用事件分解的思想.

方法一 我们在例3.5中已经求出了递推式 $E\left(X^n\right)=\frac{r}{p}E\left((Y-1)^{n-1}\right)$，此时令 $n=2$，那么 $E\left(X^2\right)=\frac{r}{p}E\left(Y-1\right)=\frac{r}{p}(EY-1)$. 而 $EY=\frac{r+1}{p}$. 那么代入即可得到 $\mathrm{Var}(X)=\frac{rq}{p^2}$.

方法二 将 $X\sim NB(r,p)$ 分解成 $X=I_1+\cdots+I_r,\forall j=1,...,r,I_j\sim G(p)$. 那么 $\mathrm{Var}(X)=\frac{rq}{p^2}$.

定理 3.18 若 $X\sim U(a,b)$，则 $\mathrm{Var}(X)=\frac{(b-a)^2}{12}$.

定理 3.19 若 $X\sim \mathrm{Exp}(\lambda)$，则 $\mathrm{Var}(X)=\frac{1}{\lambda^2}$.

定理 3.20 若 $X\sim N(\mu,\sigma^2)$，则 $\mathrm{Var}(X)=\sigma^2$.

定义 3.10 我们称

$$X^*=\frac{X-EX}{\sqrt{\mathrm{Var}(X)}} \tag{3.2.7}$$

为 X 的标准化随机变量. 易知 $E(X^*)=0,\mathrm{Var}(X^*)=1$.

3.2.2 矩

定义 3.11 设 X 为随机变量，$C\in\mathbb{R},r\in\mathbb{N}_+$，则

$$E\left((X-C)^r\right) \tag{3.2.8}$$

称为 X 关于 C 点的 r 阶矩 (若存在的话). 特别地：

(1) 若 $C=0$，此时

$$\alpha_r=E\left(X^r\right) \tag{3.2.9}$$

称为 X 的 r 阶原点矩.

(2) 若 $C=EX$，此时

$$\mu_r=E\left((X-EX)^r\right) \tag{3.2.10}$$

称为 X 的 r 阶中心矩.

定理 3.21 正态分布的矩有如下性质. 若 $X \sim N(0,1)$，则

$$E\left(X^{2k+1}\right) = 0 \tag{3.2.11}$$
$$E\left(X^{2k}\right) = (2k-1)!! \tag{3.2.12}$$

也即：标准正态分布的奇数阶原点矩为 0，偶数阶原点矩为阶数-1 的二阶阶乘. 由于正态分布的可标准化性，其 n 阶矩也很容易求出.

证明 首先有一个错解，那就是

$$\begin{aligned} E(X^r) &= \int_{-\infty}^{+\infty} x^r \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x = \int_0^{+\infty} y^{(r-1)/2} \frac{1}{2\sqrt{2\pi}} \mathrm{e}^{-y/2} \mathrm{d}y \\ &= \frac{1}{2^{(2-r)/2}\sqrt{\pi}} \int_0^{+\infty} y^{(r+1)/2-1} \frac{1}{2^{(r+1)/2}} \mathrm{e}^{-y/2} \mathrm{d}y = \frac{1}{2^{(2-r)/2}\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \end{aligned} \tag{3.2.13}$$

显然，这违背了积分换元自变量必须单调的规则. 所以千万不要犯这种低级错误. 正确的解法是

$$E(X^r) = \int_{-\infty}^{+\infty} x^r \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x \tag{3.2.14}$$

当 r 为奇数的时候，被积函数是奇函数，所以积分为 0. 当 r 为偶数时

$$\begin{aligned} E(X^r) &= \int_{-\infty}^{+\infty} x^r \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x = 2\int_0^{+\infty} x^r \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \mathrm{d}x = 2\int_0^{+\infty} y^{(r-1)/2} \frac{1}{2\sqrt{2\pi}} \mathrm{e}^{-y/2} \mathrm{d}y \\ &= \frac{1}{2^{-r/2}\sqrt{\pi}} \int_0^{+\infty} y^{(r+1)/2-1} \frac{1}{2^{(r+1)/2}} \mathrm{e}^{-y/2} \mathrm{d}y = \frac{1}{2^{-r/2}\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \end{aligned} \tag{3.2.15}$$

令 $r = 2n$，显然结果为

$$E(X^r) = \frac{1}{2^{-r/2}\sqrt{\pi}} \Gamma\left(n + \frac{1}{2}\right) = \frac{1}{2^{-r/2}\sqrt{\pi}} \frac{(r-1)!!}{2^{r/2}} \sqrt{\pi} = (r-1)!! \tag{3.2.16}$$

证毕. ■

3.3 协方差和相关系数

3.3.1 协方差

定义 3.12 我们称

$$\text{Cov}(X,Y) = E\left((X-EX)(Y-EY)\right) \tag{3.3.1}$$

为 X 与 Y 的协方差 (Covariance).

显然，协方差是二阶矩.

定理 3.22 协方差具有如下性质：

(1) 无序性.$\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

(2) 退化性.$\text{Cov}(X,X) = \text{Var}(X)$.

(3) 一般地，$\text{Cov}(X,Y) = E(XY) - (EX)(EY)$(按照定义的展开式，可以对比方差的展开式 $\text{Cov}(X) = E(X^2) - (EX)^2$). 特别地，若 X, Y 独立，$\text{Cov}(X,Y) = 0$.

(4) 系数可以这样提出来：$\text{Cov}(\lambda x, \mu y) = \lambda\mu\text{Cov}(x,y)$. 有一个最直接的推论：遍历分解.

$$\text{Cov}((aX+bY),(cZ+dW)) = ac\text{Cov}(X,Z) + ad\text{Cov}(X,W) + bc\text{Cov}(Y,Z) + bd\text{Cov}(Y,W) \tag{3.3.2}$$

8

定义 3.13 若 $\xi_1,...,\xi_n$ 是定义在同一个概率空间下的随机变量，且其中每个随机变量都是平方可积的，那么称

$$\begin{aligned}
\Sigma = (\text{Cov}(\xi_i,\xi_j))_{n\times n} &= \begin{pmatrix} \text{Cov}(\xi_1,\xi_1) & \text{Cov}(\xi_1,\xi_2) & \cdots & \text{Cov}(\xi_1,\xi_n) \\ \text{Cov}(\xi_2,\xi_1) & \text{Cov}(\xi_2,\xi_2) & \cdots & \text{Cov}(\xi_2,\xi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\xi_n,\xi_1) & \text{Cov}(\xi_n,\xi_2) & \cdots & \text{Cov}(\xi_n,\xi_n) \end{pmatrix} \\
&= \begin{pmatrix} \text{Var}(\xi_1) & \text{Cov}(\xi_1,\xi_2) & \cdots & \text{Cov}(\xi_1,\xi_n) \\ \text{Cov}(\xi_2,\xi_1) & \text{Var}(\xi_2) & \cdots & \text{Cov}(\xi_2,\xi_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\xi_n,\xi_1) & \text{Cov}(\xi_n,\xi_2) & \cdots & \text{Var}(\xi_n) \end{pmatrix}
\end{aligned} \tag{3.3.3}$$

为 $\xi_1,...,\xi_n$ 的协方差矩阵. 显然 $\det(\Sigma) \geq 0$.

定理 3.23 若 $(X,Y) \sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$，那么 (X,Y) 的协方差矩阵为

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{3.3.4}$$

3.3.2　相关系数

定义 3.14 称

$$\rho_{X,Y} = \mathrm{Cov}(X^*,Y^*) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{3.3.5}$$

为随机变量 X,Y 的相关系数.

定理 3.24 若 $X,Y \sim N(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$，则 $\rho_{X,Y} = \rho$.

定理 3.25 相关系数有如下性质：

(1) 若 X,Y 独立，则 $\rho_{X,Y} = 0$.

(2)$|\rho_{X,Y}| \in [0,1]$. 且当 $|\rho_{X,Y}| = 1$ 时 X,Y 存在严格的线性关系

$$\begin{cases} \rho_{X,Y} = 1 & \exists a > 0, b \in \mathbb{R}, \text{s.t.} X = aY + b \quad (\text{正线性相关}) \\ \rho_{X,Y} = 1 & \exists a < 0, b \in \mathbb{R}, \text{s.t.} X = aY + b \quad (\text{负线性相关}) \end{cases} \tag{3.3.6}$$

因而 $\rho_{X,Y}$ 也称作 X,Y 的线性相关系数，只能刻画 X,Y 的线性相关程度.$\rho_{X,Y} = 0$ 并不意味着 X,Y 不相关，只意味着它们不线性相关. 所以定理 (1) 是不可逆的.

例 3.15 对定理3.25的举例：设 $X \sim U\left(-\frac{1}{2},\frac{1}{2}\right), Y = \cos X$. 那么 $EX = EY = 0$，易计算得到 $\mathrm{Cov}(X,Y) = E(XY) = \int_{-1/2}^{1/2} x\cos x \mathrm{d}x = 0$. 故 X,Y 相关系数为 0，但存在非线性的函数关系.

通常来说，我们就直接用 $\rho_{X,Y}$ 来衡量相关程度了. 这是因为 $\mathrm{Cov}(X,Y)$ 会受到两个随机变量量纲的影响，但是仍然可以定性地判断相关程度. 一般来说，不相关不意味着独立 (参见命题3.29).

定理 3.26 (期望形式的 Cauchy-Schwartz 不等式) 对两个平方可积的随机变量 ξ,η，

$$(E(\xi\eta))^2 \le (E\xi)^2 (E\eta)^2 \tag{3.3.7}$$

等号成立当且仅当 $P\left(\frac{\xi}{\eta} = t_0\right) = 1$，其中 t_0 为一常数.

定理 3.27 (方差形式的 Cauchy-Schwartz 不等式) 对两个平方可积的随机变量 ξ, η,

$$\mathrm{Cov}(\xi, \eta) \leq \sqrt{\mathrm{Var}(\xi)\mathrm{Var}(\eta)} \tag{3.3.8}$$

等号成立当且仅当 $P\left(\frac{\xi}{\eta} = t_0\right) = 1$，其中 t_0 为一常数.

定理 3.28 对任何非退化的随机变量 ξ, η 平方可积，如下四个命题相互等价：

(1)ξ, η 不相关.

(2)$\mathrm{Cov}(\xi, \eta) = 0$.

(3)$E(\xi, \eta) = (E\xi)(E\eta)$.

(4)$\mathrm{Var}(\xi + \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta)$.

注意上面定理的 (3),(4). 之前说过这两个式子在 ξ, η 独立时成立 (当然也不可逆，因为总是可以凑出良好的两个随机变量满足式子)，事实上现在对不相关变量也可以这么做了.

定理 3.29 X, Y 独立 $\underset{\not\Leftarrow}{\Rightarrow} X, Y$ 不相关.

例 3.16 易求得若 X, Y 服从单位圆内的均匀分布，则 X, Y 不独立但 X, Y 也不相关.

定理 3.30 只有在正态分布情形下，不相关与独立等价.

证明 这是显然的，因为正态分布的 $\rho = \rho_{X,Y}$. ■

例 3.17 将一根棍子截成两段, 设 X 为短的那段的长度, Y 为长的那段的长度, 则 $\rho_{X,Y} = \underline{-1}$.

解 看似需要使用定义证明，实则不必. 因为有关系式 $X + Y = L$，所以是显然的负线性相关.

例 3.18 掷 n 次骰子，记其中正面朝上的次数为 T_n，反面朝上的次数为 F_n，则 $\rho_{T_n,F_n} = \underline{-1}$.

解 和上题相同，考虑 $T_n + F_n = n$.

3.4 其他数字特征

3.4.1 平均绝对差

定义 3.15 随机变量 X 的平均绝对差为

$$E|X - EX|. \tag{3.4.1}$$

3.4.2 生成函数与矩母函数

定义 3.16 定义离散型随机变量的生成函数

$$g_X(s) = E\left(s^X\right) = \sum_{k=1}^{n} s^k P(X = k) \tag{3.4.2}$$

显然，生成函数与分布是一一对应的.

定理 3.31 离散型随机变量 X, Y 独立当且仅当

$$g_{X+Y}(s) = g_X(s) + g_Y(s) \tag{3.4.3}$$

证明 这是显然的. 因为上式等价于 $E\left(s^X s^Y\right) = E\left(s^X\right) E\left(s^Y\right)$. ■

定义 3.17 连续型随机变量 X 的矩母函数为

$$\varphi_X(t) = E\left(\mathrm{e}^{tX}\right), t \in \mathbb{R}. \tag{3.4.4}$$

显然，矩母函数与分布也是一一对应的. 其独立性也有类似的性质

定理 3.32 连续型随机变量 X, Y 独立当且仅当

$$\varphi_{X+Y}(t) = \varphi_X(t) + \varphi_Y(t) \tag{3.4.5}$$

矩母函数还可以用来求解原点矩.

定理 3.33

$$E\left(X^n\right) = \varphi_X^{(n)}(t)\Big|_{t=0} \tag{3.4.6}$$

特别地，$n = 1$ 时

$$E(X) = \varphi_X'(0) \tag{3.4.7}$$

3.4.3 特征函数

定义 3.18 随机变量 X 的特征函数为

$$E\left(\mathrm{e}^{\mathrm{i}tX}\right), t \in \mathbb{R}. \tag{3.4.8}$$

根据数学期望的定义，很容易计算出对于离散情况 $P(X=a_i)=p_i$ 而言

$$E\left(\mathrm{e}^{\mathrm{i}tX}\right)=\sum_{i=1}^{n}\mathrm{e}^{\mathrm{i}ta_i}p_i \tag{3.4.9}$$

对于连续情况 $X\sim f(x)$ 而言

$$E\left(\mathrm{e}^{\mathrm{i}tX}\right)=\int_{-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}tx}f(x)\mathrm{d}x. \tag{3.4.10}$$

3.4.4 偏度系数

偏度系数是用来检验对称性的一个数字特征.

定义 3.19 定义偏度系数为标准化后的三阶矩

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) \tag{3.4.11}$$

且

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)\begin{cases}>0 & \text{右偏}\\<0 & \text{左偏}\\=0 & \text{对称}\end{cases} \tag{3.4.12}$$

3.4.5 峰度系数

峰度系数是用来检验重尾性的一个数字特征.

定义 3.20 定义峰度系数为标准化后的四阶矩

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right) \tag{3.4.13}$$

且

$$E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right)\begin{cases}\geq 3 & \text{重尾}\\\leq 3 & \text{轻尾}\end{cases} \tag{3.4.14}$$

这是因为标准正态分布的峰度系数为 3.

3.5　常见分布的数字特征

下表给出了常见分布的一些数字特征.

分布名称	参数	概率密度函数 (分布律)	期望	方差	特征函数
退化分布	c	$\begin{pmatrix} c \\ 1 \end{pmatrix}$	c	0	e^{ict}
两点分布	$p \in (0,1)$	$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$	p	pq	$q + p\mathrm{e}^{\mathrm{i}t}$
二项分布 $B(n,p)$	$n \geq 1$ $p \in (0,1)$	$k = 0,1,...,n; P(X=k) = C_n^k p^k q^{n-k}$	np	npq	$\left(q + p\mathrm{e}^{\mathrm{i}t}\right)^n$
几何分布 $G(p)$	$p \in (0,1)$	$k = 1,2,...; P(X=k) = q^{k-1}p$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}$
Pascal 分布 $NB(r,p)$	$r = 1,2,...$ $p \in (0,1)$	$k = r, r+1, ...; P(X=k) = C_{k-1}^{r-1} p^r q^{k-r}$	$\dfrac{r}{p}$	$\dfrac{rq}{p^2}$	$\left(\dfrac{p\mathrm{e}^{\mathrm{i}t}}{1-q\mathrm{e}^{\mathrm{i}t}}\right)^r$
Poisson 分布 $P(\lambda)$	$\lambda > 0$	$k = 0,1,...; P(X=k) = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$	λ	λ	$\mathrm{e}^{\lambda(\exp(\mathrm{i}t)-1)}$
超几何分布 $H(n,M,N)$	$n, M < N \in \mathbb{N}$	k的取值范围根据k,M,N关系确定 $P(X=k) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	$\dfrac{nM}{N}$	$\dfrac{nM}{N}\dfrac{N-M}{N}\dfrac{N-n}{N-1}$	爬
均匀分布 $U(a,b)$	$a,b; a<b$	$\dfrac{1}{b-a}I(a<x<b)$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{\mathrm{e}^{itb}-\mathrm{e}^{ita}}{\mathrm{i}t(b-a)}$
正态分布 $N(\mu,\sigma^2)$	μ, σ^2(注意平方)	$\dfrac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$	μ	σ^2	$\mathrm{e}^{\mathrm{i}at-(\sigma t^2/2)}$
指数分布 $\mathrm{Exp}(\lambda)$	$\lambda > 0$	$\lambda\mathrm{e}^{-\lambda x}I(x>0)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(1-\dfrac{\mathrm{i}t}{\lambda}\right)^{-1}$
χ^2 分布 $\chi^2(n)$	$n \geq 1$	$\dfrac{x^{n/2-1}\mathrm{e}^{-x/2}}{2^{n/2}\Gamma(n/2)}$	n	$2n$	$(1-2\mathrm{i}t)^{-n/2}$

表 3.1 常见分布的一些数字特征

第四章　大数定理与中心极限定理

4.1　大数定理

定义 4.1 若 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P(|\xi_n - \xi| \geq \varepsilon) = 0 \tag{4.1.1}$$

那么我们就称随机变量序列 $\{\xi_n\}$ 依概率收敛到随机变量 ξ，记作 $\xi_n \xrightarrow{p} \xi$.

定理 4.1 设 $\{X_n\}$ 是一列独立同分布的随机变量序列，具有公共的数学期望 μ 和方差 σ^2，则

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu. \tag{4.1.2}$$

即 $\{X_n\}$ 服从大数定理.

事实上，只需要均值存在就可以有大数定理成立，上述定理中加上了方差存在的条件，只是为了证明的方便. 甚至不必同分布，不独立也是可以的. 大数定律的直观含义是：当 n 很大时 $\overline{X}_n$ 接近于 X_i 的期望值.

引理 4.1 (Markov 不等式) 对于只取非负值的随机变量 X，$\forall \varepsilon > 0$，有

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon} \tag{4.1.3}$$

证明 若 X 为连续型变量，设其概率密度 $f(x)$，显然

$$E(X) = \int_0^{+\infty} xf(x)\mathrm{d}x \leq \int_\varepsilon^{+\infty} xf(x)\mathrm{d}x \tag{4.1.4}$$

因为在 $[\varepsilon, +\infty)$ 内总有 $x \geq \varepsilon$，故

$$P(X \geq \varepsilon) = \frac{1}{\varepsilon}\int_{\varepsilon}^{+\infty} \varepsilon f(x)\mathrm{d}x \leq \int_{\varepsilon}^{+\infty} x f(x)\mathrm{d}x \leq E(X) \tag{4.1.5}$$

得证.

对于离散型随机变量的情况，其证明亦是如此思路，只不过是将积分号化成求和号. ■

Markov 不等式的直观含义是，既然一个随机变量的期望是存在的，那么我们就可以认为，它取值趋于无限大的概率是趋于 0 的. 那么其偏离期望的程度应该用什么刻画呢？我们这时就有 Chebyshev 不等式

引理 4.2 (Chebyshev 不等式) 若 $\mathrm{Var}(X)$ 存在，则

$$P\left(|X - EX| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \tag{4.1.6}$$

证明 只需要将引理4.1中的 $(X-EX)^2$ 代替 X，那么 $E\left((X-EX)^2\right) = \mathrm{Var}X$. 而且 ε 是任意的正实数，所以只需要将 ε^2 代以 ε 即可，并注意到 $P((X-EX)^2 \leq \varepsilon^2) = (|X-EX| \leq \varepsilon)$，即可得证. ■

可以看出，Chebyshev 不等式是 Markov 不等式的一个特例，其意义为：一个存在方差的随机变量，其偏离期望的程度是有限的，且被方差以一个比例系数的形式进行了限制.

以上两个不等式可以用来证明大数定理.

4.2 中心极限定理

定理 4.2 (独立同分布中心极限定理, Lindburg-Levy 定理) 设 $X_1, X_2, ..., X_n$ 为独立同分布的随机变量，其有共同的期望 μ 与方差 σ^2，那么 $\forall x \in \mathbb{R}$，有

$$\lim_{n\to\infty} P\left(\frac{1}{\sigma\sqrt{n}}(X_1 + X_2 + \cdots + X_n - n\mu) \leq x\right) = \Phi(x) \tag{4.2.1}$$

换句说法，就是

$$\frac{1}{\sigma\sqrt{n}}(X_1 + X_2 + \cdots + X_n - n\mu) \xrightarrow{d} N(0,1) \tag{4.2.2}$$

也就是标准化后的随机变量组服从标准正态分布.

注意，标准化除以的实则是各个变量方差之和的算术平方根 $\sqrt{n\sigma^2}$.

定理 4.3 (De Moivre-Laplace 定理) 设 $X_1, X_2, ..., X_n$ 为独立同分布的随机变量，且 X_i 服从概率为 p 的两点分布，记 $S_n = X_1 + \cdots + X_n$. 那么显然 $E(S_n) = np, \sigma(S_n) = \sqrt{npq}$，代入定理4.2有

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{d} N(0,1) \tag{4.2.3}$$

显然这是定理4.2的特例.

定理 4.4 (Liapunov 中心极限定理) 设 $X_1, X_2, ..., X_n$ 为独立的随机变量，且记 $S_n = X_1 + \cdots + X_n$. 那么有

$$\mu(S_n) = E(X_1) + \cdots + E(X_n) = \mu_1 + \cdots + \mu_n \tag{4.2.4}$$

$$\sigma(S_n) = \sqrt{\operatorname{Var}(S_n)} = \sqrt{\operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n)} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2} \tag{4.2.5}$$

且

$$\frac{S_n - \mu(S_n)}{\sigma(S_n)} \xrightarrow{d} N(0,1) \tag{4.2.6}$$

例 4.1 有 100 道选择题 (4 选 1). 我们考虑两类学生:

(1) 一点也不懂，完全靠蒙 (即随机选)，求选对 20 题以上的概率.

(2) 从以前的模拟考试知道，选对的概率为 80%，求选对 90 题以上的概率.

解 (1) 这种情况下就需要用到 De Moivre-Laplace 定理

$$\frac{X - 100 \times \frac{1}{4}}{\sqrt{100 \times \frac{1}{4} \times \frac{3}{4}}} = \frac{2X - 50}{5\sqrt{3}} \sim N(0,1) \tag{4.2.7}$$

故 $X \sim N\left(25, \frac{75}{4}\right)$. 那么

$$P(X > 20) = P\left(\frac{2X - 50}{5\sqrt{3}} > -\frac{2}{\sqrt{3}}\right) = 1 - \Phi\left(-\frac{2}{\sqrt{3}}\right) = \Phi\left(\frac{2}{\sqrt{3}}\right) \tag{4.2.8}$$

(2) 与第一问本质是一样的，只不过 p 换成了 0.85.

参考文献

[1] 陈希孺. 陈希孺文集-概率论与数理统计[M]. 合肥: 中国科学技术大学出版社, 2009.

[2] 陈昱. 讲义[M]. 合肥: 自印, 2020.

[3] 程艺, 陈卿, 李平. 数学分析讲义. 第二册[M]. 北京: 高等教育出版社, 2020.