集合的基数 (Cardinal Number)

离散数学—集合论

南京大学计算机科学与技术系

集合的基数

- 引言：有限与无限
- 集合的等势关系
- 集合的基数
- 可数集(**Countable set**)
- **Cantor**定理
- 优势关系
- **Bernstein**定理
- 有限集的基数（容斥原理）

我们怎么比较集合的大小

- “数得清”的我们就数元素个数。
- “不可数”的咋办？
  - “常识”不一定经得起追问。

有限与无限：“宇宙旅馆”

啊？客满啦？

没关系，我让现在住在 k 号房间的客人移到 k+1 号。你就住进第1号房间吧！

有限与无限：差别不仅是数量

- 伽利略悖论：
  - 传统公理：“整体大于部分”
  - 伽利略发现：$\{1,2,3,\ldots\}$与$\{1^2,2^2,3^2,\ldots\}$一一对应。

集合的等势关系

- 等势关系的定义
  - 如果存在从集合A到B的双射，则称集合A与B等势。
  - 集合A与B等势记为：A≈B, 否则A≉B。
  - A≈B意味着：A，B中的元素可以“一一对应”。
  - 要证明A≈B，找出一个从A到B的双射即可。
- “等势”的集合就被认为是“一样大”

等势关系是等价关系

- 自反性：I_A:A→A
- 对称性：如果f:A→B是双射，则f的反函数f^{-1}: B→A，也是双射。
- 传递性：如果f:A→B，g:B→C均是双射，则 $g \circ f$ 是从A到C的双射。
- 例子
  - 与自然数集等势的所有集合构成一个等价类。

自然数定义为集合（回顾）

- 设a为集合, 称$a \cup \{a\}$为a的后继, 记为或a^+, 或s(a)。
- 集合N递归定义如下：
  - $\varnothing \in N$
  - $\forall a(a \in N \rightarrow a^+ \in N)$
- N的每一个元素称为一个自然数（自然数集合）
  - $\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}, \ldots$
  - $\varnothing$记为0，0^+记为1，1^+记为2，2^+记为3，余此类推

有限集与无限集

- **S是有限集合，iff. 存在自然数n，使得S与n等势**
- **S不是有限集合(无限集、无穷集)，iff. 存在S的真子集S’，使得S与S’等势**

$\Rightarrow$ **S一定包含一个与自然数集合等势的子集M = {a_0,a_1,a_2,…}**

令S’=S-{a_0}，可以定义f:S→S’如下：

对于任意$a_i \in$M, $f(a_i)= a_{i+1}$； 对于任意$x \in$S-M, $f(x)= x$.

显然这是双射，即S与其真子集S'等势。

$\Leftarrow$ 假设S是有限集，令$|S|=n$, 则对S的任意真子集S', 若$|S'|=m$，必有$m<n$, 因此从S'到S的任一单射不可能是满射。

集合A的基数

- 若A与自然数n等势，则card A = n
- 若A与自然数集合N等势，则card A = $\aleph_0$
- 若A与实数集合R等势，则card A = $\aleph$
- 如果存在从A到N的单射，则称A为可数集，或可列集。[card A $\le \aleph_0$]

无限可数集（无穷可列集）

- 与自然数集等势的集合称为无限可数集
  - 直观上说：集合的元素可以按确定的顺序线性排列，所谓“确定的”顺序是指对序列中任一元素，可以说出：它“前”、“后”元素是什么。
- 整数集(包括负数)与自然数集等势

$$
\begin{array}{cccccccccccccccc}
0, & -1, & 1, & -2, & 2, & -3, & 3, & -4, & \ldots\ldots \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0, & 1, & 2, & 3, & 4, & 5, & 6, & 7, & \ldots\ldots
\end{array}
$$

$$
g(n)=\begin{cases} 2n & n>=0 \\ -2n-1 & n<0 \end{cases}
$$

自然数集的笛卡儿积是可列集

- 所有的自然数对构成的集合与自然数集等势

<0,0>	<1,0>	<2,0>	<3,0>	<4,0>	•••
<0,1>	<1,1>	<2,1>	<3,1>	•••	
<0,2>	<1,2>	<2,2>	•••		
<0,3>	<1,3>	•••			
<0,4>					

类似的图形显示：
可列个可列集的并集仍然是可列集合

<0,0>, <0,1>, <1,0>, <0,2>, <1,1>, <2,0>, <0,3>, ...

$$l(m,n)=\frac{1}{2}\sum_{i=1}^{m+n} i+(m+1)=\frac{(m+n)(m+n+1)}{2}+(m+1)$$

证明无限集等势的例子

- (0,1)与整个实数集等势
  - 双射：$f:(0,1)\to R : f(x) = \mathrm{tg}(\pi x - \frac{\pi}{2})$
- 对任意不相等的实数a,b($a<b$), [0,1]与[a,b]等势
  - 双射： $f:[0,1]\to[a,b]: f(x) = (b-a)x+a$

  (这实际上意味着：任意长的线段与任意短的线段等势)

实数集不是可列集

- (0,1)不是可列集 //注意：(0,1)与实数集合等势
  - “对角线证明法”

    假设(0,1)中的元素可以线性排列：

    $0.b_{11}b_{12}b_{13}b_{14}\ldots$

    $0.b_{21}b_{22}b_{23}b_{24}\ldots$

    $0.b_{31}b_{32}b_{33}b_{34}\ldots$

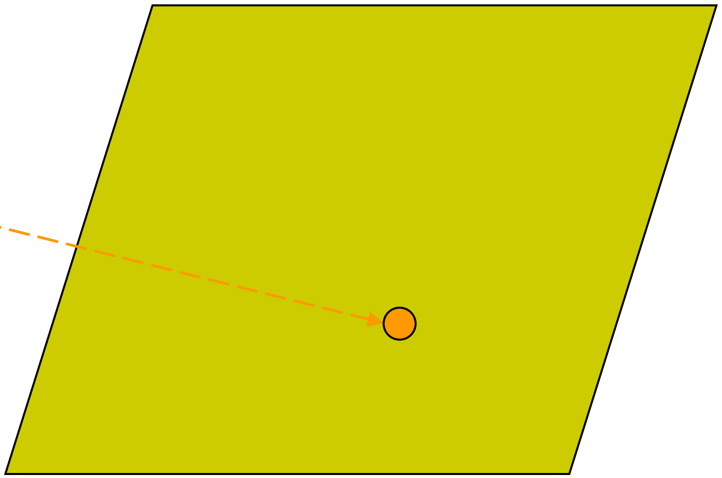
    $0.b_{41}b_{42}b_{43}b_{44}\ldots$

    $\vdots$

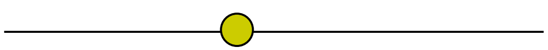
    则0. $b_1b_2b_3b_4\ldots$（$b_i \neq b_{ii}$）不含在上述序列中

直线上的点集与平面上的点集等势

$$\left\{\begin{array}{l} 0.a_1a_2a_3\ldots\ldots \\ 0.b_1b_2b_3\ldots\ldots \end{array}\right.$$



$$0.a_1b_1a_2b_2a_3b_3\ldots\ldots$$



这实际上意味着直线上的点与任意有限维空间的点“一样多”！

康托尔定理

- 任何集合与其幂集不等势，即：**A≉ρ(A)**
  - 证明要点：

设g是从A到ρ(A)的函数，构造集合B如下：

$$B=\{x\in A \mid x\notin g(x)\}$$

则B∈ρ(A)，但不可能存在$x\in A$，能满足$g(x)=B$，因为，如果有这样的x_0, 则$x_0\in B \leftrightarrow x_0\notin B$。

因此，g不可能是满射。

集合的优势关系

- 如果存在从集合A到集合B的单射，则称“集合B优势于集合A”，记为 A≼•B

  [card A ≤ card B]

- 如果集合B优势于集合A，且B与A不等势，则称“集合B真优势于集合A”，记为A≺•B

  [card A < card B]

- 实数集合真优势于自然数集
- 例子：对任意集合A，A的幂集真优势于集合A

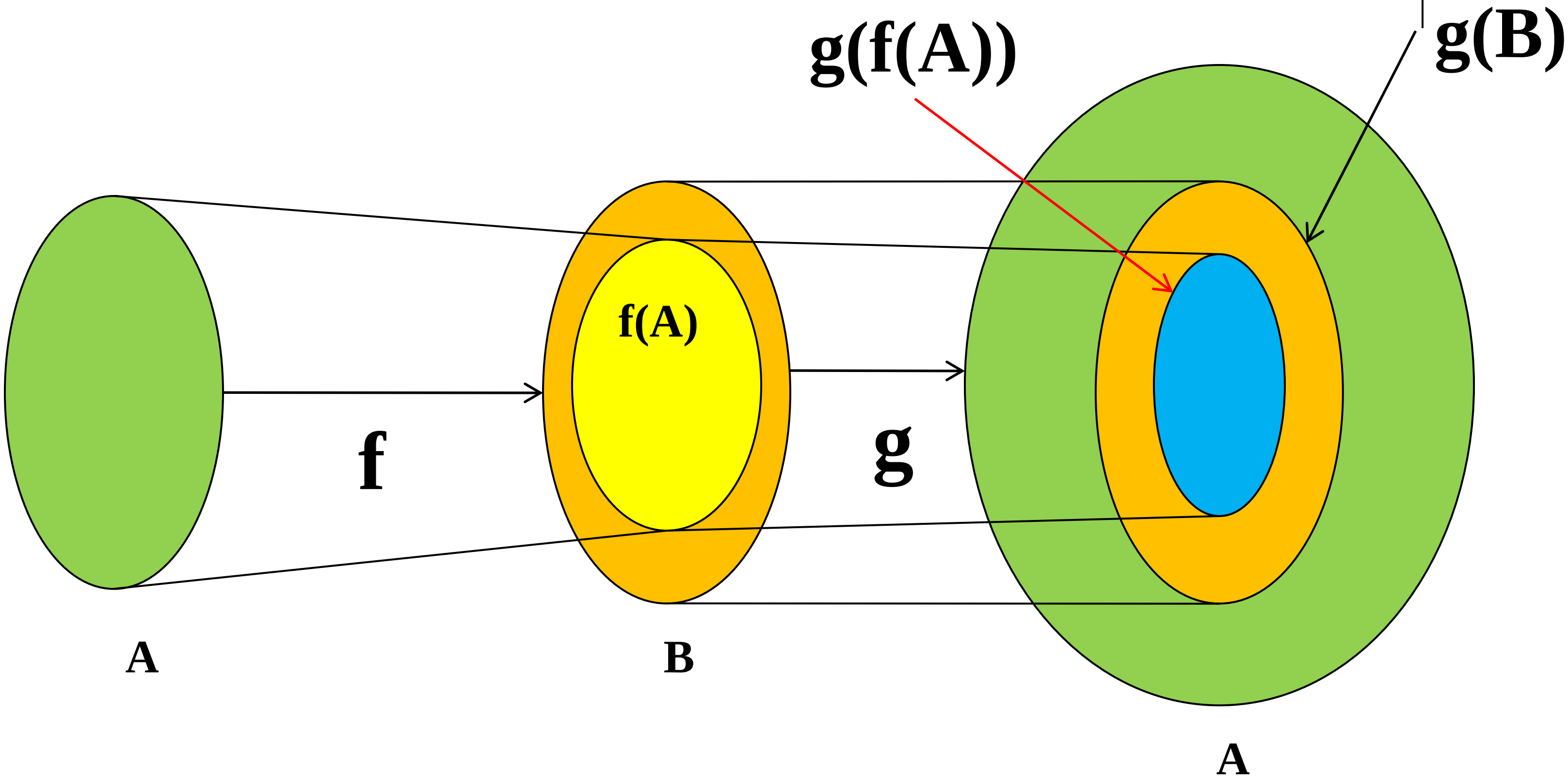
集合优势关系的性质

- 自反性：恒等函数
- 若A≼•B，且B≼•A，则A≈B (比较:反对称性)

  (Cantor-Bernstein定理)
- 传递性：单射的复合仍然是单射

Bernstein定理的证明

- 若A≼•B，且B≼•A，则A≈B。
- 由A≼•B可知,存在从A到B的一对一函数*f*，同样，由B≼•A可知，存在从B到A的一对一函数*g*，于是：$g \circ f$ 是从A到A的一对一函数。
- 显然，$g(f(A)) \subseteq g(B) \subseteq A$, 且*f*, *g*的一对一性质保证了：$g(f(A)) \approx A$, $g(B) \approx B$。
- “三明治”引理可推出：$A \approx g(B)$, 从而A≈B。

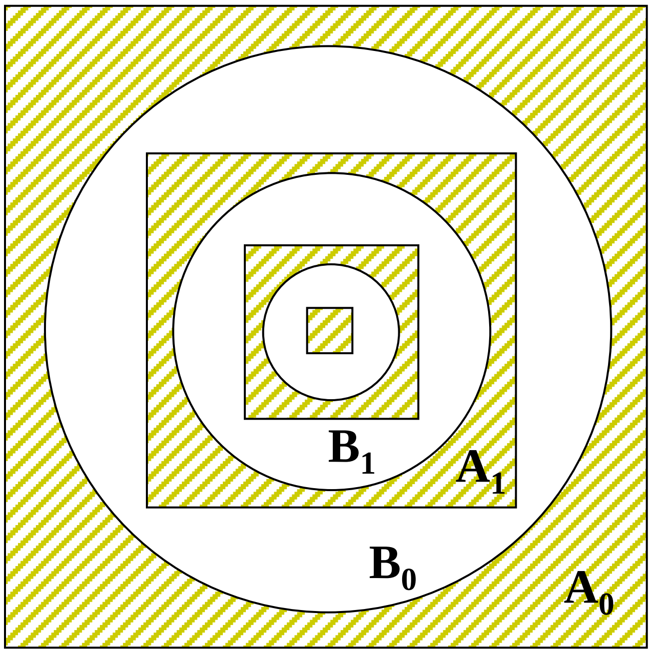


$$g(f(\mathrm{A}))\subseteq g(\mathrm{B})\subseteq \mathrm{A}$$

$$g(f(\mathrm{A}))\approx \mathrm{A},\ g(\mathrm{B})\approx \mathrm{B}$$

“三明治”引理的证明

- 若$A_1 \subseteq B \subseteq A$, 且$A_1 \approx A$, 则：$B \approx A$



1. 令$A_0=A$, $B_0=B$.

2. 设f是从A_0到A_1的一一对应函数（$A_0 \approx A_1$）令$A_{n+1}=f(A_n)$, $B_{n+1}=f(B_n)$, 递归地得到序列:

$A_0, A_1, \ldots, A_n, \ldots$ 以及 $B_0, B_1, \ldots, B_n, \ldots$

3. 由$A_1 \subseteq B_0 \subseteq A_0$, 得$A_{n+1} \subseteq B_n \subseteq A_n$

4. 令$C_n=A_n-B_n$, $\cup C_n=C$（C即左图阴影部分）, $D=A-C$（图中白色部分）

可以定义从A_0到B_0 的一一对应函数g如下:

$$g(x)=\begin{cases} f(x) & 若x \in C \quad \text{阴影部分} \\ x & 若x \in D \quad \text{白色部分} \end{cases}$$

优势关系的反对称性用于证明等势

- 证明实数集的两个子集(0,1)和[0,1]等势。
  - 直接找双射不太容易

关键是如何安排在[0,1]中但不在(0,1)中的0和1。

想象那个“宇宙旅馆”。我们可以取(0,1)的一个与自然数集合等势的子集(一定有)$\{a_1,a_2,a_3,\ldots\}$, “腾出”前两个位置安排0和1

一种证法：
$$f(x)=\begin{cases}\dfrac{1}{2} & x=0\\[2ex] \dfrac{1}{2^2} & x=1\\[2ex] \dfrac{1}{2^{n+2}} & x=\dfrac{1}{2^n},\ n=1,2,3\ldots\\[2ex] x & x\text{为其它值}\end{cases}$$

优势关系的反对称性用于证明等势 (续)

- 证明实数集的两个子集(0,1)和[0,1]等势。
  - 分别找两个一对一的映射往往比找一个双射容易

$$f:(0,1)\to[0,1]:\ f(x)=x$$

$$g:[0,1]\to(0,1):g(x)=\frac{1}{2}+\frac{x}{4}\qquad \text{注意}: g([0,1])=[\frac{1}{2},\frac{3}{4}]$$

实数集与ρ(N)等势

- $[0, 1) \approx \{0, 1\}^N$ 从而 $R \approx \rho(N)$
  - [0,1)中的数唯一地表示为$0.b_1b_2b_3b_4\ldots$

    不容许连续无数个1，比如1/2=0. 1000… (**NOT** 0.0111…)
  - $f : [0, 1) \to \{0, 1\}^N$

    $0.b_1b_2b_3b_4\ldots \to b_{1,}\, b_{2,}\, b_{3,}\, b_4\ldots$

    f是单射
  - $g : \{0, 1\}^N \to [0, 1)$

    $b_{1,}\, b_{2,}\, b_{3,}\, b_4\ldots \to 0.b_1b_2b_3b_4\ldots$ //看做十进制数

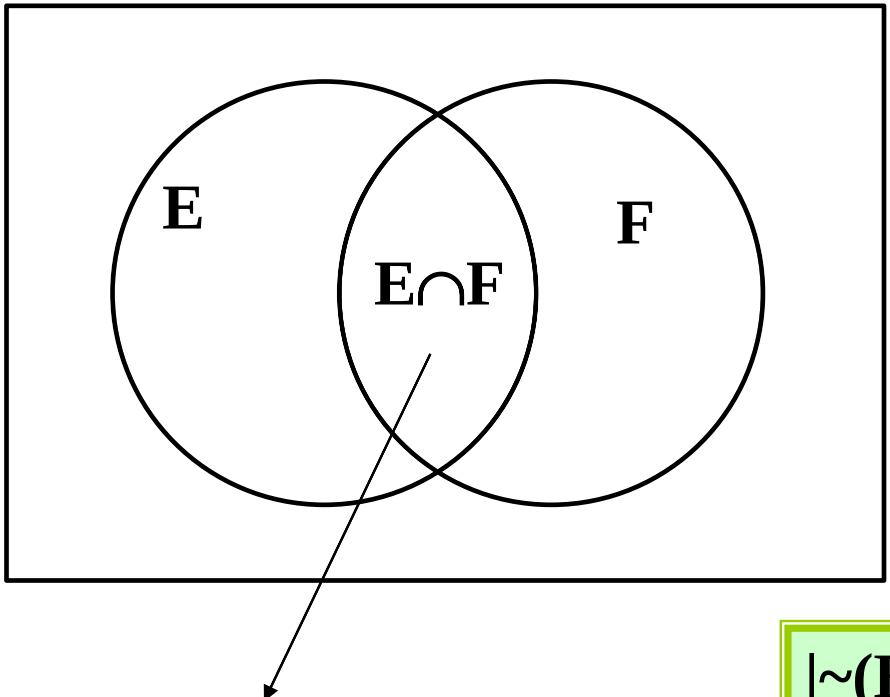
    g是单射
  - 根据Bernstein定理，得证

连续统假设

不存在集合S：

$$\aleph_0 < \mathrm{card}\ S < \aleph$$

有限集的基数（如何计算？）



既学英语，又学法语的同学

假设全班共100人，记为

$$|U| = 100$$

学英语的50人，学法语的30人，分别记为：

$$|E| = 50; |F| = 30$$

既不学英语，也不学法语的人数可能多于20人。

$$|\sim(E\cup F)| = |U| - |E \cup F|$$

$$= |U| - ((|E|+|F|) - |E\cap F|)$$

多少种排法？

- **将0,1,2,...,9排成一列，要求第1个数字大于1，最后一个数字小于8，共有多少种排法？**
  - **这10个数字所有的排法构成全集U, |U|=10!**
  - **第1个数字不大于1的排法构成子集A(即所有以0或者1开头的排法), |A|=2•9!**
  - **最后一个数字不小于8的排法构成子集B(即所有以8或者9结束的排法), |B|=2•9!**
  - **$|A\cap B|=2\cdot 2\cdot 8!$**
  - **题目要求的排法构成子集$(\sim A\cap \sim B)$**
  - **$|(\sim A\cap\sim B)| = |U| - |A\cup B| = |U| - |A| - |B| + |A\cap B| = 10! - 4\cdot 9! + 4\cdot 8! = 2{,}338{,}560$**

三个集合的并集（计算基数）

- 假设定义全集的三个子集A,B,C。则：

$$|A\cup B\cup C|=|A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|$$

- 证明：

$$|A\cup B\cup C|=|A\cup B|+|C|-|(A\cup B)\cap C|$$

$$=|A|+|B|-|A\cap B|+|C|-|(A\cap C)\cup(B\cap C)|$$

$$=|A|+|B|-|A\cap B|+|C|-|(A\cap C)|-|(B\cap C)|+|(A\cap B\cap C)|$$

$$=|A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|$$

关于选课的例子

- 全班共有160个学生
  - 选数学课64人，选计算机课94人，选金融课58人
  - 选数学与金融的28人，选数学与计算机的26人，选计算机与金融的22人
  - 三种课全选的14人。
- 问：这三种课都没选的是多少？只选一门计算机的有多少？

问题的解

- **M-数学、C-计算机、F-金融**
- 包含-排斥原理

$|M\cup C\cup F|=|M|+|C|+|F|-|M\cap F|-|M\cap C|-|C\cap F|+|M\cap C\cap F|$

$=64+94+58-28-26-22+14$

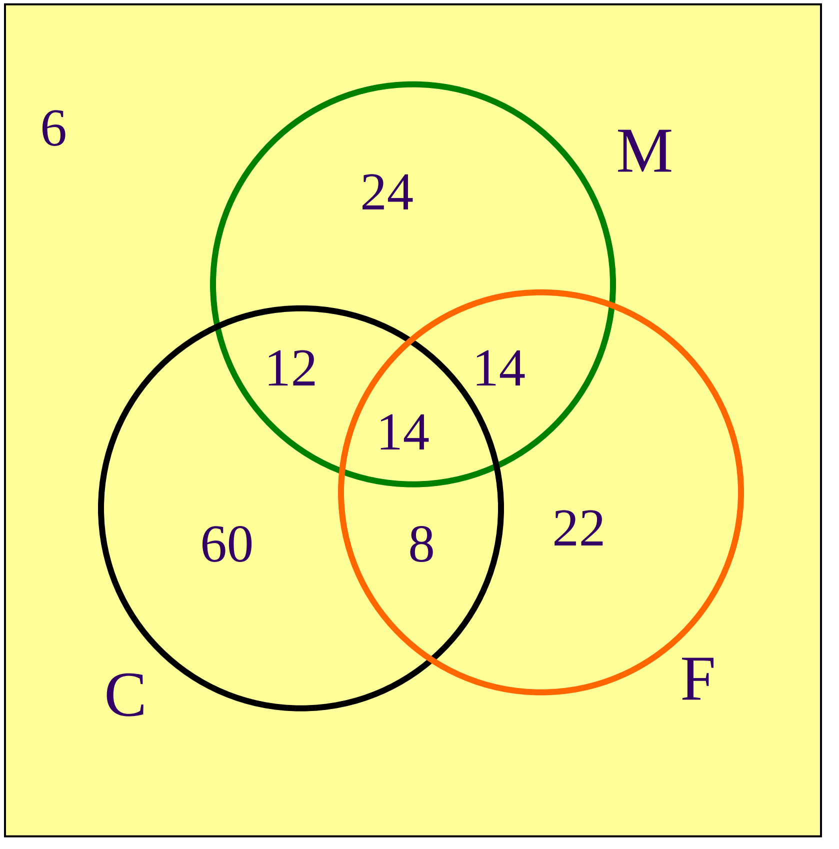
$=154$

未选课的**6**人。

只选了计算机课的**60**人

$|C|-|C\cap(M\cup F)|=$

$|C|-|M\cap C|-|C\cap F|+|M\cap C\cap F|$



容斥原理 (Inclusion-Exclusion Principle)

假设全集含N个元素，$A_1, A_2, \ldots, A_n$是分别满足相应性质的元素构成的子集合。则不满足任何性质的集合的元素个数是：

$$N(\overline{A_1}\,\overline{A_2}\ldots\overline{A_n}) = N - S_1 + S_2 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

其中，$S_k = \sum_{1 \le i_1 \le i_2 \le \ldots \le i_k \le n} | A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k} | \quad k = 1, 2, \ldots, n$

例如：4个子集的公式为：

N - (|S$_1$|+ |S$_2$|+ |S$_3$|+ |S$_4$|)

+ (|S$_1$∩S$_2$|+|S$_1$∩S$_3$|+|S$_1$∩S$_4$|+|S$_2$∩S$_3$|+|S$_2$∩S$_4$|+|S$_3$∩S$_4$|)

- (|S$_1$∩S$_2$∩S$_3$|+|S$_1$∩S$_2$∩S$_4$|+|S$_1$∩S$_3$∩S$_4$|+|S$_2$∩S$_3$∩S$_4$|)

+ |S$_1$∩S$_2$∩S$_3$∩S$_4$|

容斥原理的证明

- 计数公式：$\bigcup_{i=1}^{n} A_i = S_1 - S_2 + S_3 - \ldots + (-1)^{k-1} S_k + \ldots + (-1)^{n-1} S_n$
- 证明：满足1个或多个性质的元素恰好被计数1次.
  - 设对象*a*出现在m个(A_i)集合中
  - *a*在S_1中被计数C_1^m次，S_k中被计数恰好C_k^m次
  - 将上述分析带入计数公式可得：

$$C_1^m - C_2^m + \ldots + (-1)^{k-1} C_k^m + \ldots + (-1)^{m-1} C_m^m$$

    - 该计算式值为1，因为当*x*=1时下式为0：

$$(1-x)^m = 1 - C_1^m x + C_2^m x^2 + \ldots + (-1)^k C_k^m x^k + \ldots + (-1)^m C_m^m x^m$$

  - *a*恰好被计数1次

埃拉托色尼筛选法(Sieve of Eratosthenes)

- 用筛选法求质数（以25以内的为例）

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

[2] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

[3] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

[5] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

100以内有多少质数

- 100以内的任意合数必有不大于其平方根的质数为其因子。这样的质数只有4个：{2, 3, 5, 7}
- 设A_2, A_3, A_5, A_7 分别是可被相应质数整除的100以内大于1的自然数的集合。则100以内质数的数量为：

$$
\begin{aligned}
N(\overline{A_2}\,\overline{A_3}\,\overline{A_5}\,\overline{A_7})+4 &= 99-\left\lfloor\frac{100}{2}\right\rfloor-\left\lfloor\frac{100}{3}\right\rfloor-\left\lfloor\frac{100}{5}\right\rfloor-\left\lfloor\frac{100}{7}\right\rfloor \\
&+\left\lfloor\frac{100}{2\cdot3}\right\rfloor+\left\lfloor\frac{100}{2\cdot5}\right\rfloor+\left\lfloor\frac{100}{2\cdot7}\right\rfloor+\left\lfloor\frac{100}{3\cdot5}\right\rfloor+\left\lfloor\frac{100}{3\cdot7}\right\rfloor+\left\lfloor\frac{100}{5\cdot7}\right\rfloor \\
&-\left\lfloor\frac{100}{2\cdot3\cdot5}\right\rfloor-\left\lfloor\frac{100}{2\cdot3\cdot7}\right\rfloor-\left\lfloor\frac{100}{2\cdot5\cdot7}\right\rfloor-\left\lfloor\frac{100}{3\cdot5\cdot7}\right\rfloor+\left\lfloor\frac{100}{2\cdot3\cdot5\cdot7}\right\rfloor\underline{+4} \\
&= 99-50-33-20-14+16+10+7+6+4+2-3-2-1-0+0+4 \\
&= 25
\end{aligned}
$$

why?

Euler's totient (ϕ函数, Phi)

- $\phi(n) = |\{ k \mid 1 \le k \le n, \gcd(k, n) = 1\}|, n \in Z^+$
  - $\phi(3) = 2, \phi(4) = 2, \phi(12) = 4$
- 设 $n = p_1^{\alpha 1} p_2^{\alpha 2} \ldots p_k^{\alpha k}$
- 令 $A_i = \{ x \mid 1 \le x \le n, p_i$整除$x\}$
- $\phi(n) = |\sim A_1 \cap \sim A_2 \cap \ldots \cap \sim A_k|$

$$= n - (n/p_1 + \ldots + n/p_k) + (n/p_1 p_2 + \ldots + n/p_{k-1} p_k)$$

$$- \ldots + (-1)^k n/p_1 p_2 \ldots p_k$$

$$= n(1 - 1/p_1)(1 - 1/p_2) \ldots (1 - 1/p_k)$$

粗心的衣帽间管理员

- 剧场的衣帽管理间新来了一个粗心的管理员,他忘了给每个客人的帽子夹上号码牌。散场时他只好随意地将帽子发还给客人。没有任何人拿到自己的帽子的概率是多少？
- 这可以看作一个排列问题：对标号为1,2,3,…,n的n个帽子重新排列，新的序号为i_1, i_2, i_3,…,i_n。上述问题即：满足对任意k ($1 \le k \le n$), $i_k \ne k$的排列出现的概率是多少？
- 这样的排列称为“错位排列”(derangement)。
- 适当的集合模型使问题得到简化。

错位排列的个数 – 推导

- 我们将$i_k=k$称为“性质A_k”。满足性质A_k的排列构成所有排列的一个子集A_k。

错位排列的个数为：

$$N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = N - S_1 + S_2 - S_3 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

其中：$N = n!$

S_k如前面的定义，即 $\sum_{1\le i_1 \le i_2 \ldots \le i_k \le n} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}|$

注意：保持k项不变的置换，即其余$n-k$项可任意排列。

所以：

$$S_1 = \binom{n}{1}(n-1)!;\ S_2 = \binom{n}{2}(n-2)!;\ \ldots,\ S_k = \binom{n}{k}(n-k)! = \frac{n!}{k!}$$

错位排列的个数 – 计算

我们已经知道错位排列的个数为：

$$N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = N - S_1 + S_2 - S_3 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

其中：$N = n!$

将诸$S_k = \binom{n}{k}(n-k)! = \dfrac{n!}{k!}(k = 1,2,3,\ldots,n)$代入上面的式子：

$\therefore N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = n!\sum_{k=1}^{n}\dfrac{(-1)^k}{k!}$; $\therefore$ 要求的概率是:$\sum_{k=1}^{n}\dfrac{(-1)^k}{k!}$

注意:$\sum_{k=1}^{\infty}\dfrac{(-1)^k}{k!} = e^{-1}$, 所以这概率值与$e^{-1} \approx 0.367879$误差小于$\dfrac{1}{n!}$;

换句话说，除了较小的n, 所求概率约为0.36788。

作业

- 教材[2.4.5, 5.1.4, 7.5, 7.6]
  - p. 120： 32, 35, 38
  - p. 265： 16, 22, 40
  - p. 386： 8, 21, 22
  - p. 392： 2, 15