概率论:Week 7

学业辅导中心

1 概率论发展简史

不少材料中都有很详细的介绍[1]. 这里只简单介绍.

概率论作为一门学科起源于 17 世纪中期. 1654 年，一个名叫 A.G.C. de Méré 的法国贵族对赌博以及赌博中的问题很感兴趣，但他对一些问题感到很困惑，为解决自己的困惑，他向著名数学家 B. Pascal (1623—1662) 求助. 为解答 de Méré 提出的问题，Pascal 与另外一位法国著名数学家 P. Fermat (1601—1655) 通信进行了讨论. 1655 年，荷兰科学家 C. Huygens (1629—1695) 首次访问巴黎，期间他学习了 Pascal 与 Fermat 关于概率论的工作. 1657 年，当他回到荷兰后，写了一本小册子，名叫《De Ratiociniis in Ludo Aleae》(可译为《关于机会游戏的计算》), 这是关于概率论的第一本书.

图 1: A.G.C. de Méré

图 2: B. Pascal

图 3: P. Fermat

图 4: C. Huygens

概率论在 18 世纪得到了快速发展，这个期间的主要贡献者是 J. Bernolli (1654—1705) 与 A. de Moivre (1667—1754).Bernoulli 在概率论领域的代表作是《Ars Conjectandi》(可译为《猜测的艺术》)，发表于 1713 年，即他逝世后的第八年. 在此书中他严格地证明了概率论的第一个**极限定理**. De Moivre 开创了概率论的现代方法：1718 年发表了《The Doctrine of Chance》. 在此书中统计**独立性**的定义首次出现.1730 年 de Moivre 的另外一本专著《Miscellanea Analytica Supplementum》(可译为《解析方法》) 正式出版. 其中，关于对称 Bernoulli 试验的中心极限定理首次提出并得到证明.

[1]我参考了http://www.tup.tsinghua.edu.cn/upload/books/yz/058550-01.pdf

图 5: J.Bernolli

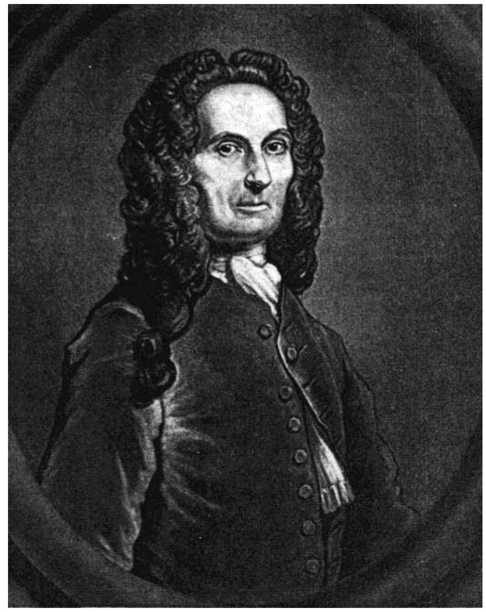


图 6: A. de Moivre

19 世纪，Pierre-Simon Laplace 在他 1812 年的著作 Theorie Analytique des Probabilities中介绍了概率的数学理论及其科学应用. 同时期的主要贡献者是S. D. Poisson (1781—1840), C. F. Gauss (1777—1855), P. L. Chebyshev (1821—1894)等. 这个时期的研究主要围绕极限定理展开.

图 7: P.S.M. Laplace

图 8: S. D. Poisson

图 9: C. F. Gauss

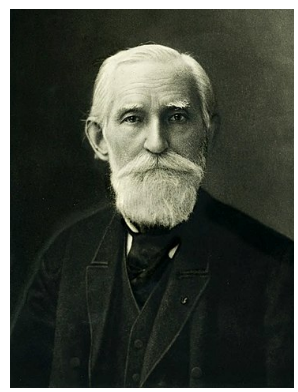


图 10: P. L. Chebyshev

20 世纪可称为概率论发展的现代时期，本时期开始于概率论的公理化. 在这个方向上的早期贡献者有 S. N. Berstein (1880—1968), R. von Mises (1883— 1953) 与 E. Borel (1871—1956). 1933 年，俄罗斯著名数学家 A. N. Kolmogorov 出版了他的专著《Foundations of the Theory of Probability》. 其中，他为概率论建立了目前广泛采纳的**公理化体系**.

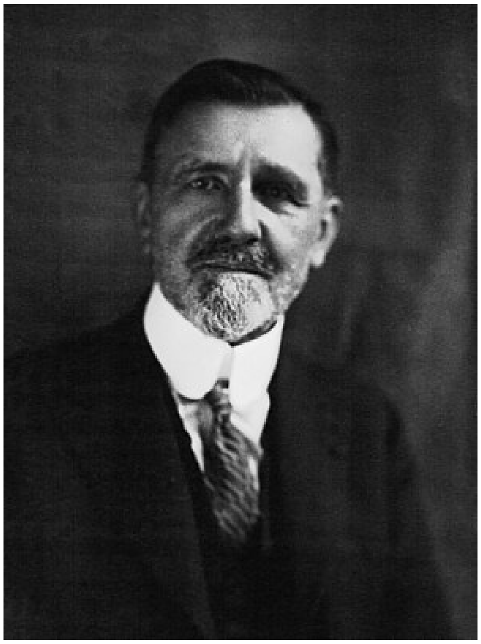


图 11: E. Borel

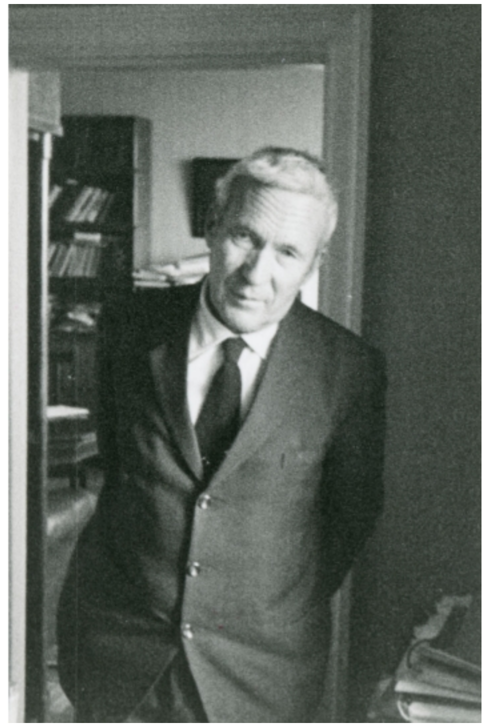


图 12: A. N. Kolmogorov

2 样本空间与事件

定义 (事件)**.** *对于随机现象通常关心的是在**试验或观察中**某个**结果**是否出现, 这些结果称为**随机事件**,简称**事件(event)**.*

以后我们一般都用 A,B, C 等大写英文字母表示随机事件.

为什么研究概率论? 随机现象有其偶然性的一面,也有其必然性的一面.这种必然性表现为大量试验中随机事件出现的频率的稳定性,即一个随机事件出现的频率常在某个固定的常数附近摆动,这种规律性我们称之为**统计规律**性. 频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的、不随人们意志而改变的一种客观属性,因此可以对它进行**度量**.

定义 (随机试验)**.** *随机现象的实现和对它某个特征的观测.*

3 概率空间

接下来我们的目标是:如何用数学的语言, 严格说明"发生某个随机事件的概率是多少"这件事情.

3.1 概率论背景

定义 (样本空间)**.** *样本空间(Sample Space): 随机试验中所有基本事件所构成的集合, 通常用 Ω 或 $\mathcal{S}$ 表示. 样本空间中的元素, 称为样本点, 通常用 ω 等表示.*

- 掷一枚骰子, 观察出现的点数. 则$\Omega = \{1, 2, 3, 4, 5, 6\}$.
- 考察某一地区的年降雨量, 则$\Omega = \{x|0 \leqslant x < T\}$, 这里 T 表示某个常数, 表示降雨量不会超过 T.

定义 (随机事件)**.** *简称事件 (Event), 在随机试验中我们所关心的可能出现的各种结果, 它由一个或若干个基本事件组成.*

- 必然事件 (Ω): 在试验中一定会发生的事件;
- 不可能事件 (ϕ): 在试验中不可能发生的事件;
- 因此, 我们不严格的说样本空间的子集称为随机事件.

至此, 我们知道(描述)了随机性的来源: 随机性来源于样本空间中的抽样.

3.2 集合的运算

关于事件的运算, 我们知道, 并:

$$\bigcup_{\iota \in T} A_\iota \overset{\text{def}}{=} \{x : \exists \iota \in T \text{使} x \in A_\iota\}$$

交:

$$\bigcap_{i \in T} A_i \overset{\text{def}}{=} \{x : x \in A_i, \forall t \in T\}$$

德摩根法则:

$$\left\{\bigcup_{\iota \in T} A_\iota\right\}^c = \bigcap_{\iota \in T} A_\iota^c; \quad \left\{\bigcap_{\iota \in T} A_\iota\right\}^c = \bigcup_{\iota \in T} A_\iota^c.$$

单调集列的极限: 若$A_n \uparrow$, 则它的极限可以被定义为

$$\lim_{n \to \infty} A_n \overset{\text{def}}{=} \bigcup_{n=1}^{\infty} A_n$$

同理, 若$A_n \downarrow$, 则它的极限可以被定义为

$$\lim_{n \to \infty} A_n \overset{\text{def}}{=} \bigcap_{n=1}^{\infty} A_n$$

因此, 可以定义一般集合的上, 下极限

$$\lim_{\pi \to \infty} \inf A_\pi \overset{\text{def}}{=} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \text{ 和 } \lim_{n \to \infty} \sup A_n \overset{\text{def}}{=} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

当上极限等于下极限时就是集合的极限.

3.3 测度和σ代数

现在我们的问题是: 如何刻画这些事件("集合")的概率("大小").

为此, 我们需要构造一个集函数/set function("测度")来度量这些集合.

定义 (测度)**.** 给定空间Ω上的集合系$\mathscr{F}$. 定义在$\mathscr{F}$上, 取值于$[0,+\infty]$的函数被称为**非负集函数**, 用字母$\mu,\nu,\tau,\cdots$等记之. 若对任意可列个两两不交代集合$A_1,A_2,\cdots\in\mathscr{E}$, 当$\bigcup_{i=1}^{\infty}A_n\in\mathscr{E}$时, 成立

$$\mu\Big\{\bigcup_{n=1}^{\infty}A_n\Big\}=\sum_{n=1}^{\infty}\mu(A_n),$$

则称μ有**可列可加性**.

设$\phi\in\mathscr{E}$. 若$\mathscr{E}$上的非负集函数μ有可列可加性, 满足$\mu(\phi)=0$, 则称μ是$\mathscr{E}$上的测度.

注记.

对于这样的集函数, 有必要设定它的定义域使得这个函数是良定的. 这自然要求我们引入σ代数(σ-algebra).

定义 (σ代数)**.** 称满足下列三个条件的集合系$\mathscr{F}$为σ代数:

$$\begin{aligned}&(i)\Omega\in\mathscr{F};\\&(ii)\text{若}A\in\mathscr{F},\quad\text{则}\overline{A}\in\mathscr{F};\\&(iii)\text{若}A_n\in\mathscr{F},\quad n=1,2,\cdots,\quad\text{则}\bigcup_{s=1}^{\infty}A_n\in\mathscr{F}.\end{aligned}$$

注记. 有两个特殊的σ代数: 幂集(包含集合最多), $\{\phi,\Omega\}$(包含集合最少).

注记. σ代数中, 代数指的是1,2两条(还要补一个关于交封闭), σ指的是第3条.

定义 (可测空间)**.** 非空集合Ω和它上面的σ代数$\mathscr{F}$放在一起写成的$(\Omega,\mathscr{F})$将被称为可测空间.

这时再看书上关于概率的定义, 就是一种特殊的测度:

定义 (概率)**.** 定义在事件域$\mathscr{F}$上的一个集合函数 P 称为概率,如果它满足如下三个要求:

1. $P(A)\geqslant 0$,对一切$A\in\mathcal{F}$;
2. $P(\Omega)=1$;
3. 若$A_i\in\mathscr{F},\quad i=1,2,\cdots$且两两互不相容,则

$$P(\sum_{i=1}^{\infty}A_i)\;=\;\sum_{i=1}^{\infty}P(A_i)$$

注记. σ代数还保证了可以进行测度扩张[2].

[2] 程士宏, P38

3.4 可测空间$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

我们先考虑一个特殊的全空间: $\mathbb{R}$.

定义 (由...生成的σ代数(sigma-algebra generated by...)). *称$\mathscr{F}$是由集合系$\mathscr{A}$生成的σ代数, 若*

1. *$\mathscr{A} \subset \mathscr{F}$;*
2. *任意包含$\mathscr{A}$的σ代数也包含$\mathscr{F}$.*

记$\mathscr{F}$为$\sigma(\mathscr{A})$.

定义 (一维Borel点集). $\mathcal{B}(\mathbb{R}) = \sigma((a, b])$.

注记.

$$\begin{aligned}\mathcal{B}(\mathbb{R}) &= \sigma([a, b)) \\ &= \sigma((a, b)) \\ &= \sigma([a, b]) \\ &= \sigma((-\infty, a)) \\ &= \sigma((-\infty, a]) \\ &= \ldots\end{aligned}$$

上述可测空间在我们定义分布函数和理解随机变量时至关重要.

3.5 概率的性质

1. 不可能事件的概率为 0, 即 $P(\phi) = 0$.
2. 概率具有有限可加性.即若 $A_iA_j = \varnothing(i \neq j)$,则
$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$
3. 对任何事件 A 有$P(\bar{A}) = 1 - P(A)$.
4. 如果 $A \supset B$, 则$P(A - B) = P(A) - P(B)$.
5. (单调性)如果 $A \supset B$, 则 $P(A) \geqslant P(B)$.
6. $P(A \cup B) = P(A) + P(B) - P(AB)$.
7. (Bonferroni 不等式)
$$P(AB) \geqslant P(A) + P(B) - 1 \quad (1.5.8)$$
利用归纳法不难把这两个不等式推广到n个事件的场合.
$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant P(A_1) + P(A_2) + \cdots + P(A_n)$$
$$P(A_1A_2\cdots A_n) \geqslant P(A_i) + P(A_2) + \cdots + P(A_n) - (n - 1)$$

8. (一般加法公式) 若$A_1, A_2, \cdots, A_n$为n 个事件,则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1,\cdots,n} P(A_i) - \sum_{\substack{i<j \\ i,j=1,\cdots,n}} P(A_iA_j)$$

$$+ \sum_{\substack{i<j<k \\ i,j,k=1,\cdots,n}} P(A_iA_jA_k) - \cdots + (-1)^{n-1} P(A_1A_2 \cdots A_n)$$

定义 (集合的上下连续性)**.** 如果对任意 $A_1, A_2, \cdots \in \mathscr{E}, A_n \uparrow A \in \mathscr{E}$, 均有

$$\mu(A) = \lim_{n\to\infty} \mu(A_n),$$

则称 μ 是**下连续**的;

如果对任意 $A_1, A_2, \cdots \in \mathscr{E}, A_n \downarrow A \in \mathscr{E}$ 且 $\mu(A_1) < \infty$,

$$\mu(A) = \lim_{n\to\infty} \mu(A_n),$$

则称 μ 是**上连续**的.

性质. 可列可加 = 有限可加 + 下连续

注记. 概率是下连续的, 也是上连续的. 且有

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

3.6　概率空间的例子

例 1 (离散概率空间)**.** 设$X = \{x_i\}_{i=1}^{\infty}$, $\mathscr{F} = 2^X$. 若每个x_i对应一个非负实数a_i, 则

$$\mu(A) = \sum_{a_i \in A} a_i, \quad \forall A \in \mathscr{F}$$

是$\mathscr{F}$上的测度, 特别的, 若$\sum\limits_{a_i \in X} a_i = 1$, 则$\mu$就是一个概率测度, $(X, \mathscr{F}, \mu)$形成一个概率空间.

例 2 (古典概型)**.** 设X是一个有限集, $\mathscr{F} = 2^X$, 定义

$$P(A) = \frac{\#(A)}{\#(X)}, \quad \forall A \in \mathscr{F},$$

则$(X, \mathscr{F}, P)$形成一个概率空间, 这就是古典概型.

例 3 (几何概型)**.** 考虑$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mu)$是一个测度空间, 设$\Omega \subset \mathbb{R}^n$, 且有$0 < \mu(\Omega) < +\infty$, 对$A \in \mathcal{B}(\mathbb{R}^n)$, 称

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

为事件 A 的几何概率。这里等可能性体现“落在区域 A 的概率与区域 A 的测度成正比并且与其形状位置无关”.

4 概率空间的应用

4.1 公理化的应用: 解决Bertrand悖论

分别对应了不同的概率空间$(\Omega_i, \mathscr{F}_i, P_i), i = 1, 2, 3.$

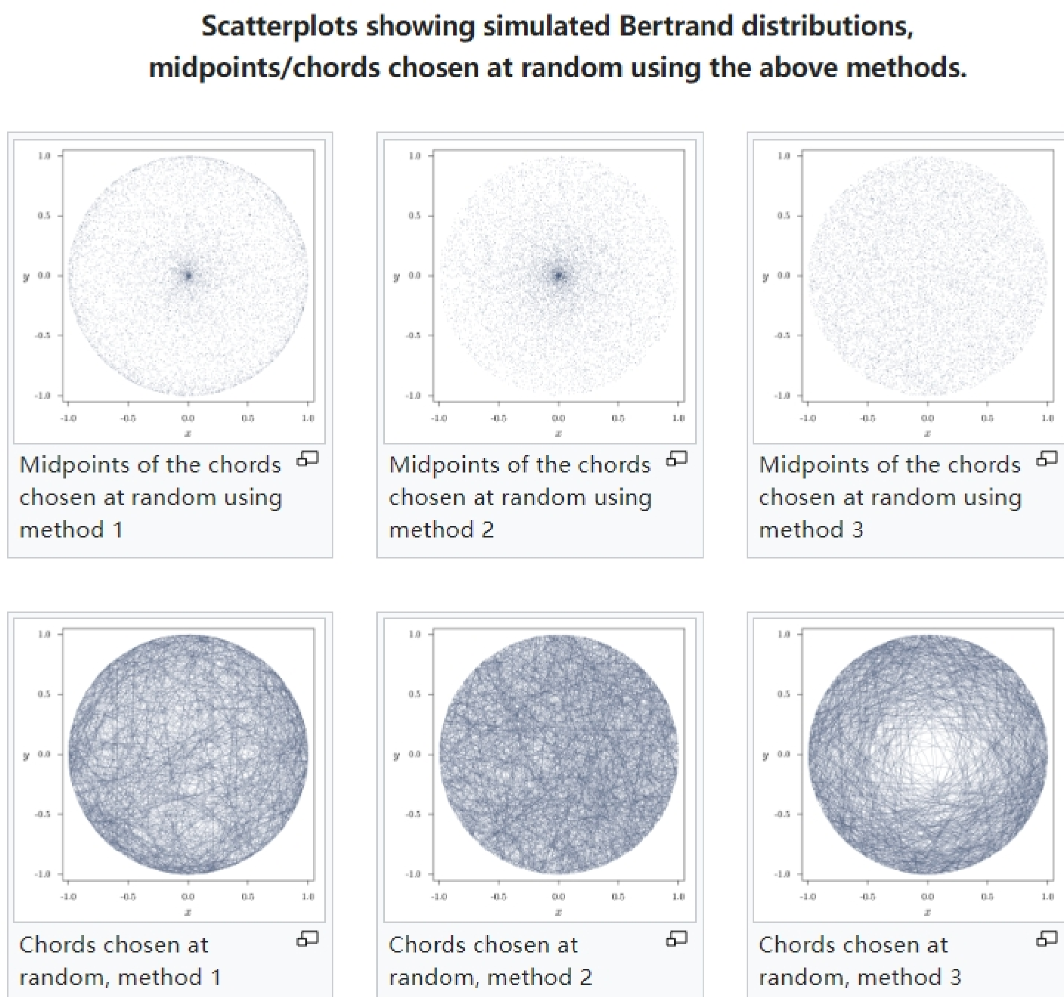


图 13: 三种情况

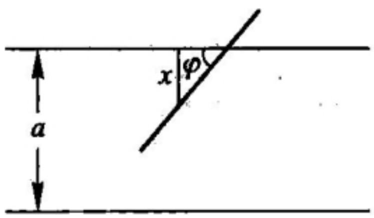
4.2 Buffon’s needle与积分的Monte Carlo方法

桌面上画满间隔均为 a 的平行线，现向桌面任意投放一长为 $l(l < a)$ 的针，求事件 E ={针与某直线相交} 的概率。

Monte Carlo 优点:

1. 误差与维数无关, 不存在“维数灾难”;
2. 适用于复杂函数积分.

［解］　以x表示针的中点到最近的一条平行线的距离，φ表示针与平行线的交角．针与平行线的位置关系见图1.4.2.



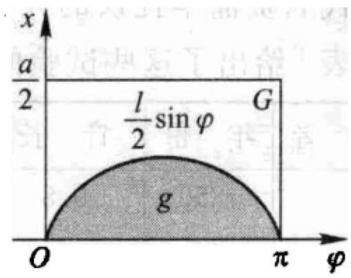


图1.4.2　蒲丰问题

显然有$0\leqslant x\leqslant\frac{a}{2}$，$0\leqslant\varphi\leqslant\pi$，以$G$表示边长为$\frac{a}{2}$及$\pi$的长方形．为使针与平行线相交，必须$x\leqslant\frac{l}{2}\sin\varphi$，满足这个关系式的区域记为$g$，在图1.4.2中用阴影表出，所求的概率为

$$p=\frac{g\text{的面积}}{G\text{的面积}}=\frac{\frac{1}{2}\int_0^{\pi}l\sin\varphi\mathrm{d}\varphi}{\frac{1}{2}a\pi}=\frac{2l}{\pi a}\qquad(1.4.2)$$

4.3 会面问题

[例 4](会面问题) 两人相约 7 点到 8 点在某地会面,先到者等候另一人 20 分钟,过时就可离去,试求这两人能会面的概率.

[解] 以 x,y 分别表示两人到达时刻,则会面的充要条件为

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

这是一个几何概率问题,可能的结果全体是边长为 60 的正方形里的点,能会面的点的区域用阴影标出(图 1.4.1).所求概率为

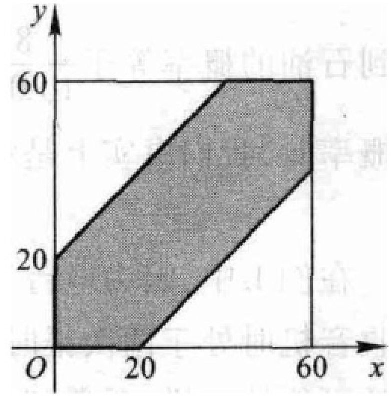


图 1.4.1 会面问题

$$p=\frac{60^2-40^2}{60^2}=\frac{5}{9}$$

4.4 匹配问题

某人写好n 封信，又写好n 只信封然后在黑暗中把每封信放入一只信封中,试求至少有一封信放对的概率.

解答. 若以A_i记第i 封信与信封符合,则所求事件为$A_1\cup A_2\cup\cdots\cup A_n$,所以可以用一般加法公式,不难求得

$$P(A_i)=\frac{(n-1)!}{n!}=\frac{1}{n}$$
$$P(A_iA_j)=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)}$$
$$P(A_iA_jA_k)=\frac{1}{n(n-1)(n-2)},\cdots,P(A_1A_2\cdots A_n)=\frac{1}{n!}$$

因此,

$$
\begin{aligned}
&P(A_1 \cup A_2 \cup \cdots \cup A_n)\\
=&\binom{n}{1}\frac{1}{n}-\binom{n}{2}\frac{1}{n(n-1)}\\
&+\binom{n}{3}\frac{1}{n(n-1)(n-2)}-\cdots+(-1)^{n-1}\frac{1}{n!}\\
=&1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}
\end{aligned}
$$

4.5 最大车牌号

［例5］（最大车牌号） 某城有 N 辆卡车，车牌号从1到 N，有一个外地人到该城去，把遇到的 n 辆车子的牌号抄下(可能重复抄到某些车牌号)，求抄到的最大号码正好为 k 的概率($1 \leqslant k \leqslant N$).

［解］ 可以看作古典概型问题，即设每辆卡车被遇到的机会相同. 若以 A_k 记抄到的最大号码为 k 这一事件，又以 B_k 记抄到的最大号码不超过 k 这一事件，则明显有 $A_k = B_k - B_{k-1}$，而且 $B_k \supset B_{k-1}$，所以由性质4知 $P(A_k) = P(B_k) - P(B_{k-1})$，而由直接计算可得 $P(B_k) = \dfrac{k^n}{N^n}$，因此最后得到

$$P(A_k) = \frac{k^n - (k-1)^n}{N^n}$$

5 排列组合

5.1 基本公式

排列数: $A_n^r = \dfrac{n!}{(n-r)!}$.

组合数(二项系数): $\dbinom{n}{r} = \dfrac{A_n^r}{r!} = \dfrac{n(n-1)\cdots(n-r+1)}{r!} = \dfrac{n!}{r!(n-r)!}$

多项系数: $\dfrac{n!}{r_1!r_2!\cdots r_k!}$(分法)

多类取法: $\dbinom{n_1}{r_1}\dbinom{n_2}{r_2}\cdots\dbinom{n_k}{r_k}$.

有放回抽r个(重复组合数):$\binom{n-r+1}{r}$

注记. 设n个元素分别为$a_1, a_2, \cdots, a_n$，被抽出的次数分别为$x_1, x_2, \cdots, x_n (x_i \geq 0)$，则$x_1 + x_2 + \cdots + x_n = r$，令$y_i = x_i + 1$ 则$y_1 + y_2 + \cdots + y_n = n + r$, y_i为正整数；

这相当于$n + r$个球, 插$n - 1$块板, 由于只有n +*r-1*个空隙, 插法共有$\binom{n-r+1}{n-1} = \binom{n-r+1}{r}$.

Stirling's Formula: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$.

5.2 组合数公式

1. $\binom{n}{k} = \binom{n}{n-k}$
2. 当$k > n$时, $\binom{n}{k} = 0$
3. $\binom{a}{0}\binom{b}{n} + \binom{a}{1}\binom{b}{n-1} + \cdots + \binom{a}{n}\binom{b}{0} = \binom{a+b}{n}$
4. 一般情况的二项式系数: $\binom{x}{r} = \frac{A_x^r}{r!} = \frac{x(x-1)}{r!}\frac{(x-2)\cdots(x-r+1)}{r!}$
5. $\binom{-a}{k} = (-1)^k\binom{a+k-1}{k}$

5.3 生日问题

在第二个问题中，n 个格子可以任意，即可以从 N 个格子中任意选出 n 个来，这种选法共有$\binom{N}{n}$种，对于每种选定的 n 个格子，有利场合正如第一个问题一样为 $n!$，故所求概率为

$$P_2=\frac{\binom{N}{n}n!}{N^n}=\frac{N!}{N^n(N-n)!}$$

这个例子是古典概型中一个很典型的问题，不少实际问题都可以归结为它.

例如，若把球解释为粒子，把格子解释为相空间中的小区域，则这个问题便相应于统计物理学中的麦克斯韦-玻尔兹曼（Maxwell-Boltzmann）统计.

这也联系着概率论历史上有名的**生日问题**：求参加某次集会的 n 个人中至少有两个人生日相同的概率 p_n. 若把 n 个人看作上面问题中的 n 个球，而把一年的 365 天作为格子，则 $N=365$，这时所求的概率就是 $1-P_2$，即

$$p_n=1-365\times364\times\cdots\times(366-n)/365^n$$

5.4 抽签与顺序无关

［例 5］ 口袋中有 a 只黑球，b 只白球，它们除颜色不同外，其他方面没有差别，现在把球随机地一只只摸出来，求第 k 次摸出的一只球是黑球的概率（$1\leqslant k\leqslant a+b$）.

［第一种解法］ 把 a 只黑球及 b 只白球都看作是不同的（例如设想把它们进行编号），若把摸出的球依次放在排列成一直线的 $a+b$ 个位置上，则可能的排列法相当于把 $a+b$ 个元素进行全排列，总数为 $(a+b)!$，把它们作为样本点全体. 有利场合数为 $a\times(a+b-1)!$，这是因为第 k 次摸得黑球有 a 种取法，而另外 $(a+b-1)$ 次摸球相当于 $a+b-1$ 只球进行全排列，有 $(a+b-1)!$ 种构成法，故所求概率为

$$P_k=\frac{a\times(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$$

这个结果与 k 无关！

5.5 不放回抽样与超几何分布

［例 6］ 如果某批产品中有 a 件次品 b 件合格品，我们采用有放回及不放回抽样方式从中抽 n 件产品，问正好有 k 件是次品的概率各是多少？

所求的概率显然与抽样方式有关，下面我们分别来讨论.

［有放回抽样场合］ 把$a+b$ 件产品进行编号，有放回抽 n 次，把可能的重复排列全体作为样本点，总数为$(a+b)^n$，其中有利场合（即次品正好出现 k 次）的数目是$\binom{n}{k}a^k b^{n-k}$，故所求概率为

$$b_k=\frac{\binom{n}{k}a^k b^{n-k}}{(a+b)^n}=\binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k} \qquad (1.3.8)$$

b_k是二项式$\left(\frac{a}{a+b}+\frac{b}{a+b}\right)^n$展开式的一般项，上述概率称为**二项分布**（分布一词的意义将在第三章阐明）. 关于二项分布更一般的讨论在以后各章陆续进行.

［不放回抽样场合］ 从$a+b$ 件产品中取出 n 件产品的可能组合全体作为样本点，总数为$\binom{a+b}{n}$，有利场合数为$\binom{a}{k}\binom{b}{n-k}$，故所求概率为

$$h_k=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} \qquad (1.3.9)$$

这个概率称为**超几何分布**.