生日问题

问题描述：假设同学的出生日期都是相互独立的，并且每个人都等可能地出生在一年中的任何一天 (2 月 29 日除外)，那么房间里有多少人才能保证其中至少两个人的生日在同一天的概率不小于 50%？

解决问题：基本思想，计算**对立事件的概率** A 不发生的概率有时要比直接计算事件 A 的概率容易很多。因此，此处可先计算房间里面的每个人生日都不相同的概率。

我们发现“在 n 个人中，所有人的生日都互不相同”的概率是

$$\frac{365}{365}\frac{364}{365}\cdots\frac{365-(n-1)}{365}$$

使用**连乘符号**，这个结果可以改写成

$$\prod_{k=0}^{n-1}\frac{365-k}{365}$$

如果使用阶乘符号，我们得到“在 n 个人中，所有人的生日都互不相同”的概率是

$$\begin{aligned}\prod_{k=0}^{n-1}\frac{365-k}{365}&=\frac{365\cdot 364\cdots\left(365-(n-1)\right)}{365^n}\\&=\frac{365\cdot 364\cdots\left(365-(n-1)\right)}{365^n}\frac{(365-n)!}{(365-n)!}\\&=\frac{365!}{365^n(365-n)!}\end{aligned}$$

所以，有两个以上人的生日相同的概率：$1-\dfrac{365!}{365^n(365-n)!}$

用到两个结论：

1. $\sum_{l=0}^{m} l = m(m+1)/2$

2. 对数函数$\ln(1-x)$当$|x|<1$时的泰勒级数展开：$-\sum_{l=1}^{\infty}\frac{x^l}{l}$，（在 a=0 位置展开）

泰勒级数：$f(x)=f(a)+\frac{f^1(a)}{1!}(x-a)+\frac{f^2(a)}{2!}(x-a)^2+\cdots+\frac{f^n(a)}{n!}(x-a)^n+\cdots$

因此，生日都不相同的概率：$p_n=\prod_{k=0}^{n-1}\frac{365-k}{365}=\prod_{k=0}^{n-1}\left(1-\frac{k}{365}\right)$，两边取对数得

$$\ln p_n=\ln\left(\prod_{k=0}^{n-1}\left(1-\frac{k}{365}\right)\right)=\sum_{k=0}^{n-1}\ln\left(1-\frac{k}{365}\right)$$

应用级数展开，有

$$\ln p_n\approx\sum_{k=0}^{n-1}-\frac{k}{365}$$

再应用前面的求和公式有

$$\ln p_n\approx\sum_{k=0}^{n-1}-\frac{k}{365}=-\frac{1}{365}\frac{(n-1)n}{2}$$

要使概率为 0.5，即令$p_n=0.5$，有

$$\begin{aligned}
&\ln\frac{1}{2}=-\frac{1}{365}\frac{(n-1)n}{2} \qquad \Rightarrow\\
&-\ln 2\approx-\frac{1}{365}\frac{(n-1)n}{2} \qquad \Rightarrow\\
&n(n-1)\approx n^2\approx 365\cdot 2\ln 2 \qquad \Rightarrow\\
&n\approx\sqrt{365\cdot 2\ln 2}
\end{aligned}$$

最后可计算得到 n 大约为 23.

扩展，如果一年的天数为 D，则有$n\approx\sqrt{D\ln 4}$

计算机仿真验证

```
clc;
clear all;
D = 365;    % 一年的天数

N = 80;     % 人数

%%%%%%%%%%%%%%%%% 理论计算 %%%%%%%%%%%%%%%%%

noshare(1) = 1;  % 只有一个人时，生日各不相同的概率为 1
for n = 1:N-1
    noshare(n+1) = noshare(n)*(D-n-1)/D;
end


share = 1-noshare;  % 有两个人以上生日相同的概率为 1-对立事件的概率

plot(share,'-'); hold on;

%%%%%%%%%%%%%% 数值仿真 %%%%%%%%%%%%%

iteration = 1e4;
for n = 1:N
    share = 0;
    for it = 1:iteration
        daylist = randi([1 D],1,n);
        tmp_L = length(unique(daylist));
        if tmp_L<length(daylist)
            share = share + 1;
        end

    end
    share_p(n) = share/iteration;
end
plot(share_p,'rp');
```

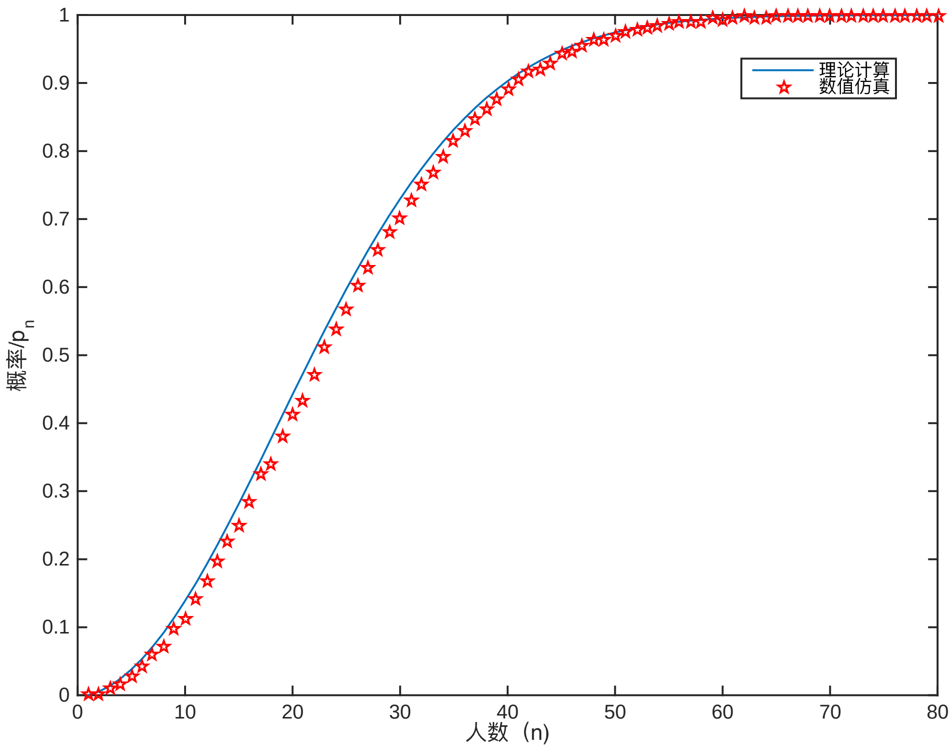

图 1. 理论与数值仿真比较：一年天数为 365

在一年有 D 天的前提下 (每一天都等可能地成为生日，所有人的出生日期都相互独立)，保证“至少两个人的生日在同一天”的概率为 50%所需要的最少人数的理论曲线

```
clc;
clear all;
Day = [100 200 365 730 1000 2000 4000 5000 8000 10000
20000 40000 60000 1e5 2e5 4e5 6e5 8e5 1e6];

n = sqrt(Day.*log(4));

plot(Day,n);
xlabel('Days');
ylabel('n');
```

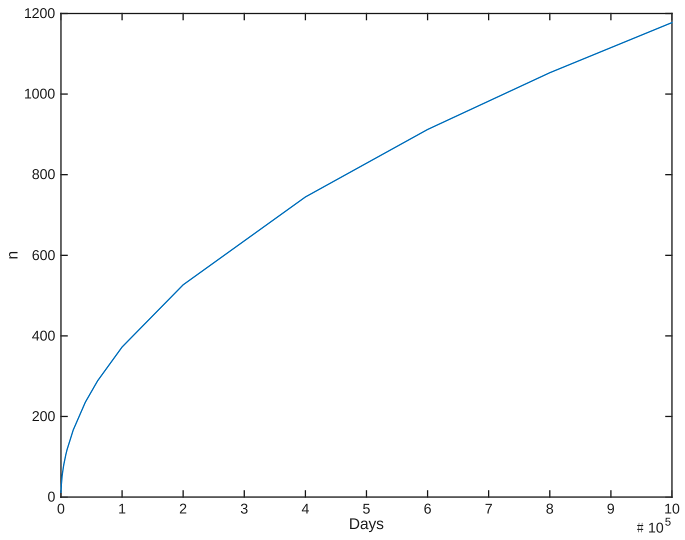

图 2. 在一年有 D 天的前提下 (每一天都等可能地成为生日, 所有人的出生日期都相互独立), 保证"至少两个人的生日在同一天"的概率为 50% 所需要的最少人数