

考虑闰年条件的生日问题

吕振华*, 刘 丹

沈阳航空航天大学理学院, 辽宁 沈阳

收稿日期: 2025年5月23日; 录用日期: 2025年6月26日; 发布日期: 2025年7月15日

摘 要

很多概率论教材中有关计算存在两人相同生日概率的问题, 也把解决这样的问题作为训练巩固学生理解等可能概率的重要手段。文章讨论了含有闰年情况的相同(不同)生日问题, 通过分组方式给出了不同情况下生日不同的概率, 并提出了一种非等可能放球的新的盒子模型, 丰富了盒子模型理论。

关键词

概率论, 古典概型, 等可能概型, 生日问题

Birthday Problem Considering Leap Year Condition

Zhenhua Lyu*, Dan Liu

School of Science, Shenyang Aerospace University, Shenyang Liaoning

Received: May 23rd, 2025; accepted: Jun. 26th, 2025; published: Jul. 15th, 2025

Abstract

Many probability theory textbooks have the problem of calculating the probability that two people have the same birthday, and they also take solving such problems as an important means of training and consolidating students' understanding of possible probabilities. This paper discusses the problem of the same (different) birthdays with leap years and presents a new box model with unequal possibilities of putting the ball, which enriches the box model theory.

Keywords

Probability Theory, Classical Probability, Equal Probability Model, Birthday Problem

*通讯作者。



1. 引言

在概率论中, 古典概型(等可能概型)是一个重要的内容。等可能概型是概率论最早研究的情形, 它与组合数学紧密结合, 是概率的最简单分配形式。古典概型的基本思想为([1], p. 15): 1) 所涉及的随机现象只有有限个样本点, 记为 n ; 2) 每个样本点发生的概率(可能性)相同; 3) 若事件 A 包含 k 个样本点, 则事件 A 发生的概率为:

$$P(A) = \frac{\text{事件}A\text{中的样本点个数}}{\text{所有样本点的个数}} = \frac{k}{n} \quad (1)$$

因此计算事件的概率转化为了计算样本点的个数。

等可能概型中有一个经典的生日问题([2], p. 11):

问题 1.1 n 个人的生日全不相同的概率是多少?

“生日问题”在密码学中有着广泛应用。它为密码学中的生日攻击方法提供了理论依据[3]。很多的教材都把这个当成“盒子模型”或“放球模型”的一个经典案例。

定理 1.2 (盒子模型) 设有 n 个球, 每个球都等可能地被放到 m 个盒子中的任一个, 每个盒子所放球数不限。如果 $n \leq m$, 把 n 个球放到不同的盒子中的概率是 A_m^n / m^n 。

用上述的盒子模型解决生日问题, 我们需要一个前提条件: 假设每人的生日在一年 365 天中的任一天是等可能的。这样很容易得到以下结论。

定理 1.3 假设每人的生日在一年 365 天中的任一天是等可能的。令 $n \leq 365$, 则 n 个人的生日全不相同的概率是 $\frac{A_{365}^n}{365^n}$ 。

定理 1.3 的结果有趣且有悖于人们的直觉[4][5]。该结果告诉我们只要人数为 57, 则存在两人生日相同的概率高达 0.9901。当然如果要存在两人生日相同的概率等于 1, 则需要 366 (或 367) 个人, 这是经典的鸽笼原理(抽屉原理)告诉我们的。

跳出“等可能”的束缚, 这个问题本身也是有趣的。那么考虑实际的情况, 即某些年份中一年有 366 天, 那么 n 个人的生日全不相同的概率是多少? 关于非等可能的生日问题, 目前研究的成果较少。只有文献[6]中讨论了每个人在 2 月 29 日出生的概率是其他日子出生概率 $1/4$ 的情况。文献[4]初步讨论了生日概率不均衡(非等可能)的情况。本文研究考虑闰年情况的生日问题, 本质上是研究非等可能概型的概率问题。我们提出分组的概念, 可以在等可能概型的基础上解决一部分非等可能概率问题。这样的思路在教学上可以拓展学生的思路, 锻炼学生的逻辑思维能力, 为后续概率模型的学习做好准备和铺垫。

下面介绍闰年的概念: 人们在使用历法时发现, 随着时间的积累, 历法所体现的日历日期与实际季节之间的总会渐行渐远, 产生较大误差。这是因为地球公转的时间并不是地球自转时间的正整数倍, 即地球公转的时间并不等于 365 天或者是 366 天, 而是 365 天 5 小时 48 分 46 秒(合 365.242 19 天)。因此, 为了确保日期与季节之间的对应关系不断保持。人们通过引入闰年的方式, 调整日历, 使其与地球公转周期保持一致。在现行公历中, 闰年为 366 天, 比常规年份多一天, 即 2 月 29 日。在中国农历中, 闰年为包含闰月的年份, 比常规年份多一个月。

公元年数能被 4 整除的年为闰年, 其余年为平年。但对整世纪年只有世纪数也能被 4 整除的才是闰

年, 其余的虽然年数都能被 4 整除, 但仍为平年, 比如 2000 年是闰年, 而 1900 年不是。每到闰年, 在这一年的 2 月多加一天, 即 2 月 29 日, 全年共 366 天。即: “四年一闰, 百年不闰, 四百年再闰”。

本文计算以下两种情况的概率值: 1) 每个人出生在其他日子的概率是出生在 2 月 29 日概率的 4 倍; 2) 每个人出生在其他日子的概率是出生在 2 月 29 日概率的 $\frac{400}{97}$ 倍。并给出一般模型可以计算出其他日子的概率是出生在 2 月 29 日概率的任意倍数(有理数)的情况。最后提供新的非等可能放球的盒子模型, 该模型丰富了盒子模型理论。

2. 主要结果

为了说清楚问题 1.1, 我们从简单的设定开始。假设每四年出现一次 2 月 29 日。每个人出生在其他日子的概率是出生在 2 月 29 日概率的 4 倍。2009 年, 谢彬[6]研究了有闰年的生日问题, 并给出这种假设下的准确概率。不过在证明过程中用到了 $1/4$ 个盒子这样不准确的表达。下面我们通过分组的方式解决这种非等可能生日问题。

定理 2.1 假设每个人出生在其他日子的概率是出生在 2 月 29 日概率的 4 倍。令 $n \leq 365$, 则 n 个人的生日完全不相同的概率是 $\frac{A_{365}^{n-1} \times 4^{n-1} \times (1464 - 3n)}{1461^n}$ 。

证明 设有 $365 \times 4 + 1 = 1461$ 个盒子, 分成 366 组, 其中一组只有 1 个盒子(称为独立组), 其他每组都有 4 个盒子。我们把 n 个球放到这些盒子中, 假设每个球放到每个盒子的概率相同。 n 个人的生日全不相同的概率等于把 n 个球放到不同的组中的概率, 记为 P_n 。分情况讨论: 1) 没有球放到独立组中, 把 n 个球放到不同的组中有 A_{365}^n 种放法。考虑到在每个组中有 4 种选择, 所以一共有 $A_{365}^n \times 4^n$ 种放球的方法; 2) 有 1 个球放到独立组的那个盒子中, 从 n 个球中取 1 个放到独立组中, 有 n 种情况, 接着把 $n-1$ 个球放到不同的 365 组中的各个盒子里, 有 $A_{365}^{n-1} \times 4^{n-1}$ 种放法, 因此这种情况共有 $A_{365}^{n-1} \times 4^{n-1} \times n$ 种放球的方法。把 n 个球放到这些盒子中, 一共有 1461^n 种情况。根据等可能概率模型公式可得:

$$P_n = \frac{n \times A_{365}^{n-1} \times 4^{n-1} + A_{365}^n \times 4^n}{1461^n} = \frac{A_{365}^{n-1} \times 4^{n-1} \times (1464 - 3n)}{1461^n} \quad (2)$$

□

根据定理 2.1 中的思路, 我们可以得到以下结论。

定理 2.2 (分组盒子模型) 设有 m 组盒子, 每组有 k_i 个盒子, 其中 $i \in \{1, 2, \dots, m\}$ 。假设每个球放到每个盒子的概率相同, 把 $n (\leq m)$ 个球放到不同组中的概率是:

$$P = \frac{\sum_{i_1, i_2, \dots, i_n \in \{1, 2, \dots, m\} \text{ 且各不相同}} \prod_{j=1}^n k_{i_j}}{\left(\sum_{i=1}^m k_i \right)^n} \quad (3)$$

证明 根据(1), 只须计算总的样本点的个数(把 n 个球等可能放入 $\sum_{i=1}^m k_i$ 个盒子有多少种放法), 以及事件包含的样本数(把 n 个球等可能放入不同组有多少种放法)。应用组合排列知识, 把 n 个球放入 $\sum_{i=1}^m k_i$ 个盒子有 $\left(\sum_{i=1}^m k_i \right)^n$ 种方法。把 n 个球放入不同的组有 A_m^n 种放法。注意这里的每种放法(放入不同组)对应不同个的盒子数量。设第一个球放在 i_1 组中, 第二个球放入 i_2 组中, 以此类推, 第 n 个球放在 i_n 组中。显然 $i_1, i_2, \dots, i_n$ 各不相同。第一个球有 k_{i_1} 个盒子可以等可能放入, 因此有 k_{i_1} 放法。以此类推, 第二个球有 k_{i_2} 放法……第

n 个球有 k_i 放法。因此对应一种放入不同组的情况, n 个球有 $\prod_{j=1}^n k_{i_j}$ 种放入盒子的方法。因此把所有放入不同组的放法对应的数量求和, 可得到把 n 个球等可能放入不同组的各个盒子有

$$\sum_{i_1, i_2, \dots, i_n \in \{1, 2, \dots, m\} \text{ 且各不相同}} \prod_{j=1}^n k_{i_j} \text{ 种方法。应用(1.1)可得(2.2)。} \quad \square$$

根据第一章中对于闰年的描述, 如果考虑无尽的漫长岁月, 我们只须考虑 400 年内闰年出现的频率即可。按照现有规律 400 年(0 年~399 年)内会出现 97 次闰年。也就是 2 月 29 日出现了 97 次, 而其他的所有日期比如 3 月 5 日出现了 400 次。因此假设每个人出生在其他日子的概率是出生在 2 月 29 日概率的 $\frac{400}{97}$ 倍。应用定理 2.1 中的思路或者直接应用定理 2.2, 可得以下推论:

推论 2.3 假设每个人出生在其他日子的概率是出生在 2 月 29 日概率的 $\frac{400}{97}$ 倍。令 $n \leq 365$, 则 n 个人的生日全不相同的概率是 $\frac{A_{365}^{n-1} \times 400^{n-1} \times (146400 - 303n)}{146097^n}$ 。

下面我们将定理 1.3 (没有 2 月 29 日)、定理 2.1 (其他日期出生概率是 2 月 29 日出生的 4 倍)、推论 2.3 (其他日期出生概率是 2 月 29 日出生的 $\frac{400}{97}$ 倍)的结果做如下比较。可以发现, 定理 2.1 和推论 2.3 的结果已经非常接近, 差距在十万分之一以内。(表 1)

Table 1. Probability of the number of people on the same birthday as at least two people
表 1. 人数与至少两人同一天生日的概率

人数	不考虑闰年	定理 2.1	推论 2.3
10	0.116 949	0.116 815	0.116 818
20	0.411 438	0.411 054	0.411 063
30	0.706 316	0.705 865	0.705 877
40	0.891 232	0.890 924	0.890 932
50	0.970 374	0.970 238	0.970 242
60	0.994 123	0.994 083	0.994 084
70	0.999 160	0.999 152	0.999 152

注记 在定理 2.1 以及推论 2.3 中, 规定每个人出生在 2 月 29 日与出生在其他日子的概率之比并非真正的概率比例, 而是根据闰年出现的规则得到的相应比例。从理论的角度来说, 这样的假定已经足够。如果想计算准确值, 那么可以用统计的放法找出每个人出生在不同日期的概率, 然后应用定理 2.2 得到具体概率值。

3. 结论

本文从经典的等可能球放盒子模型出发, 得到了不同闰年频率下的相同(不同)生日概率。并进一步给出了一种分组的盒子模型, 该模型丰富了盒子模型理论, 可以解决一些非等可能的样本点的概率问题。

问题 3.1 随机实验有 m 个结果 $A_1, A_2, \dots, A_m$, 每种结果对应的概率为 $p_1, p_2, \dots, p_m$ 。独立进行 n 次随机实验, 计算至少两次结果相同的概率。假设 $n \leq m$ 。

思路 想要利用分组的盒子模型解决问题, 难点在于如何去构造分组中盒子的数量。不妨设 $p_1, p_2, \dots, p_m$ 都是有理数(如果存在无理数可以取近似值化为有理数), 不失一般性。假设对于 $i = 1, 2, \dots, m$ 有 $p_i = \frac{k_i}{N}$, 其中 $N = \sum_{i=1}^m k_i$ 且 k_i 为正整数, 构造如下模型: 有 n 个球, 放到 m 组盒子中, 每组中盒子的数

量为 k_i , $i=1,2,\dots,m$ 。显然问题 3.1 的结果等于 n 个球放到不同组的概率, 根据定理 2.2, 结果为

$$1 - \frac{\sum_{i_1, i_2, \dots, i_n \in \{1, 2, \dots, m\} \text{ 且各不相同}} \prod_{j=1}^n k_{i_j}}{N^n}。$$

本文研究了非可能概率问题, 把给定的不同概率转化为不同分组对应的盒子数, 把一些概率问题转化为分组的盒子模型。既推广了盒子模型理论, 也为一些非等可能概率问题找到了解决办法。

基金项目

本文受辽宁省教育厅面上项目(JYTMS20230281)、广西省自然科学基金面上项目(2023GXNSFAA026514)资助。

参考文献

- [1] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程(第三版) [M]. 北京: 高等教育出版社, 2019: 15.
- [2] 盛骤, 谢式千, 潘承毅. 概率论与数理统计(第四版) [M]. 北京: 高等教育出版社, 2008.
- [3] 喻秋叶. 生日悖论在密码学中的应用[D]: [硕士学位论文]. 武汉: 华中科技大学, 2013.
- [4] 徐付霞, 常虹. 概率统计实验实践教学案例——生日问题[J]. 高等教育研究, 2017, 20(1): 121-123+125.
- [5] 刘天昕, 耿凤杰, 赵俊芳. 非均匀分布下生日问题的模拟研究[J]. 计算机应用, 2015, 35(S2): 38-40.
- [6] 谢彬. 生日概率的准确计算[J]. 内江科技, 2009, 30(8): 133-133.