概率论与数理统计课程小论文

题目：“生日悖论”的概率论解释

班　　号	
学　　号	
姓　　名	
日　　期	
得　　分	9

摘　要

本文以“生日悖论”为研究对象，运用古典概率、排列组合等概率论基础知识，推导并分析了“n 个人中至少两人同天生日”的概率公式。通过数值计算展示了不同人数下的概率结果，解释了“仅需 23 人就有 50% 以上概率存在同天生日”这一反直觉现象的数学原理。此外，本文结合美国 CDC 发布的真实人口出生数据，分析了非均匀分布对概率的影响，证明了理论模型实际上是真实世界碰撞概率的下界。

关键词：生日悖论；古典概率；排列组合；概率计算；非均匀分布

1 研究背景与意义

在概率论的诸多趣闻中，美国数学家维弗（Warren Weaver）的餐桌赌局常被人津津乐道。据说，他曾在与朋友聚餐时提出一个令人难以置信的说法：在约四十人的场合中，至少会有两个人生日相同，而且他愿意因此下注。朋友们直觉上认为这几乎不可能发生，但维弗却凭借坚定的数学信心参与了赌局。结果不出所料，当在场朋友都报出自己的生日后，他成功赢得了这场看似大胆的赌注。

我们可以从这场赌局中看出，在 40 人的群体中出现同生日的概率非常之高，这绝非巧合，而是数学概率的必然结果。正是这样的故事，让"生日悖论"成为概率论中最具反直觉魅力的经典例子，引发人们对日常生活中概率现象的重新思考。

在日常生活中，"生日相同"往往被视为小概率事件。但"生日悖论"却揭示了一个反直觉的结论：仅需 23 人，就有超过 50% 的概率至少存在两人同天生日；50 人时这个概率将超过 97%。这一现象看似"悖论"，实则是古典概率在实际场景中的典型应用。

本研究旨在通过概率论基础知识，严谨推导生日悖论的概率公式，分析数值规律，并引入真实数据探讨非均匀分布下的概率修正，从而加深对古典概率、排列组合等知识点的理解。

2 生日悖论的概率推导

要计算"n 个人中至少两人同天生日"的概率，直接分析所有可能的情况非常复杂。因此，我们采用更简洁的逆向思维，即计算其**对立事件**——"n 个人的生日全不相同"的概率，再用 1 减去该概率。

假设一年有 365 天（为简化模型，忽略闰年）。对于 n 个人的群体，每个人的生日都可以是 365 天中的任意一天。因此，样本空间大小为：

$$\text{总基本事件数} = 365^n$$

对立事件"n 个人生日全不相同"等价于从 365 天中无放回地抽取 n 个不同的日期。其排列数为：

$$A_{365}^n = 365 \times 364 \times \cdots \times (365 - n + 1) \tag{1}$$

根据古典概率定义,“n 个人生日全不相同”的概率为:

$$P(\text{全不同}) = \frac{A_{365}^{n}}{365^{n}} = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^{n}} \tag{2}$$

因此,“n 个人中至少两人同天生日”的概率为:

$$P(\text{至少两人同天}) = 1 - P(\text{全不同}) = 1 - \frac{A_{365}^{n}}{365^{n}} \tag{3}$$

3 数值分析与结果讨论

我们通过计算不同 n 值对应的概率，来直观感受生日悖论的“反直觉性”。

表 1: 不同人数下“至少两人同天生日”的概率

人数 n	概率 P(至少两人同天)
10	11.7%
20	41.1%
23	**50.7%**
30	70.6%
40	89.1%
50	**97.0%**

从表1可见，当 $n = 23$ 时，概率已超过 50%。这一结果与直觉的偏差源于大脑不擅长处理指数增长。我们通常直觉地计算“找到一个与**自己**生日相同”的概率，而本问题讨论的是“任意两人”匹配的概率。

在 n 个人的群体中，总配对数为组合数 $C_n^2 = \frac{n(n-1)}{2}$。当 $n = 23$ 时，总配对数达到 253 对。正是这大量的潜在配对，使得“至少有一对生日相同”的概率迅速累积。

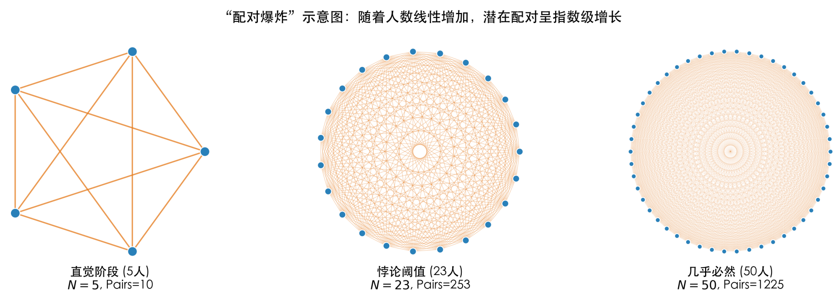

图 1: 配对爆炸示意图：随着人数 N 线性增加，潜在配对数 C_n^2 迅速“爆炸”，导致碰撞概率急剧上升。

4　拓展与思考

4.1　同月生日问题

若将问题简化为“n 个人中至少两人同月生日”，概率公式为：

$$P(\text{至少两人同月}) = 1 - \frac{12 \times 11 \times \cdots \times (12 - n + 1)}{12^n} \tag{4}$$

表 2: 不同人数下“至少两人同月生日”的概率

人数 n	概率 P	人数 n	概率 P
3	23.6%	6	77.7%
4	42.7%	8	95.4%
5	**61.7%**	**10**	**99.6%**

计算可得（见表2），仅需 5 人，概率即超过 50%。

4.2　基于真实数据的非均匀分布修正

前文假设了出生概率均匀分布（$p = 1/12$）。为了验证非均匀分布的影响，我们查阅了美国国家卫生统计中心（NCHS/CDC）2019 年的出生统计数据 [1]。数据表明，7 月和 8 月为出生高峰。

真实数据分布如下表所示：

表 3: 2019 年美国出生人口月度分布（数据源自 CDC Table I-2 [1]）

月份	出生人数	比例 p_i	月份	出生人数	比例 p_i
1 月	310,872	8.30%	7 月	333,646	8.90%
2 月	279,963	7.47%	**8 月**	341,685	**9.12%**
3 月	304,237	8.12%	9 月	325,781	8.69%
4 月	298,947	7.98%	10 月	325,043	8.67%
5 月	316,386	8.44%	11 月	298,086	7.95%
6 月	304,092	8.11%	12 月	308,802	8.24%

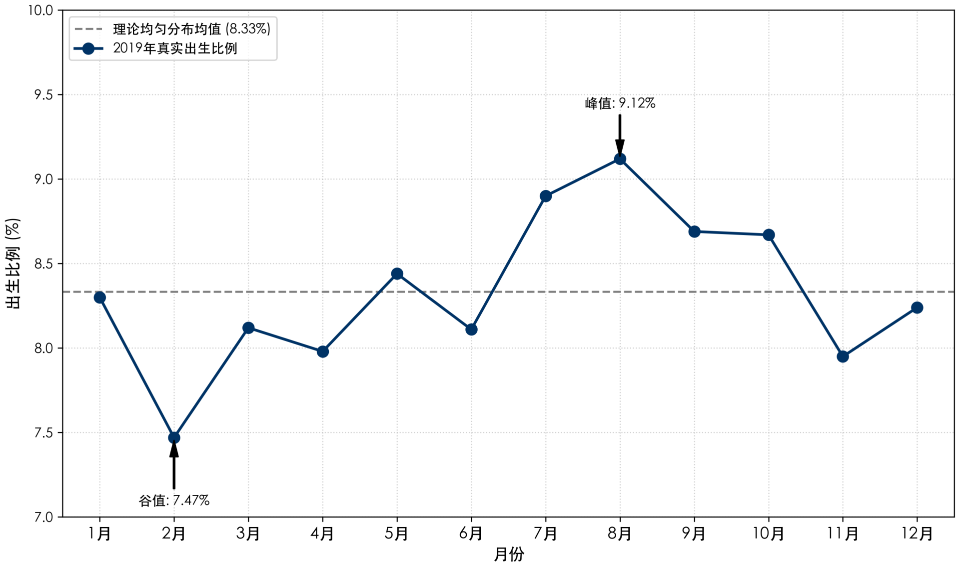

图 2: 2019 年美国出生人口比率与理论均值对比分析

在非均匀分布下，任意两人同月出生的概率为各月概率的平方和：

$$P_{\text{real}} = \sum_{i=1}^{12} p_i^2 \approx 0.0841 \tag{5}$$

作为对比，理想均匀分布下的概率为：

$$P_{\text{ideal}} = \sum_{i=1}^{12} \left(\frac{1}{12}\right)^2 \approx 0.0833 \tag{6}$$

5 总结

通过上述推导与计算，我们得出以下结论：

首先，**直觉往往会低估群体中“巧合”发生的概率**。无论是生日问题（23 人）还是同月问题（5 人），数学计算的结果都远超直觉预期。这是因为随着人数增加，可能的配对数量呈二次多项式增长。

其次，数据对比显示 $P_{\text{real}}(8.41\%) > P_{\text{ideal}}(8.33\%)$。这验证了一个重要的数学规律：分布越不均匀，碰撞概率越大。

最后，这一差值虽然微小，但它从理论上证明了基于均匀分布推导出的“生日悖论”概率公式，实际上是真实世界中碰撞概率的下界。在现实世界中，由于存在“出生高峰”等社会文化因素，遇到同月或同天生日的人的概率比理论预测的更高。这充分体现了概率论在解释和修正日常认知中的重要价值。

参考文献

[1] Martin, J. A., Hamilton, B. E., Osterman, M. J. K., & Driscoll, A. K. (2021). *Births: Final Data for 2019*. National Vital Statistics Reports, 70(2). National Center for Health Statistics. `https://www.cdc.gov/nchs/data/nvsr/nvsr70/nvsr70-02-508.pdf`