

信息学中的概率统计：作业一

截止日期：2025 年 9 月 26 日（周五）下课前。请务必通过教学网提交电子版。可下课前同时提交纸质版。

与本次作业相关的事宜，请发邮件给我(ruosongwang@pku.edu.cn), 抄送研究生助教叶昊洋(yhyfhgs@gmail.com), 以及负责本次作业的本科生助教张致铨 (zzq12345@stu.pku.edu.cn)。

第一题

对于 n 个事件 A_1, A_2, \dots, A_n ，从概率的公理化定义和条件概率的定义出发证明下述结论。

(1) 一般加法公式：

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n)。$$

(2) 一般 Union Bound：

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)。$$

(3) 一般乘法公式：若 $P(A_1 A_2 \dots A_n) > 0$ ，有

$$P(A_1 A_2 \dots A_n) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 A_2) \cdot \dots \cdot P(A_n | A_1 A_2 \dots A_{n-1})。$$

第二题

对于三个事件 A , B 和 C ，若 $P(C) > 0$ ，我们称事件 A 和 B 在事件 C 发生时是条件独立的，当且仅当

$$P(AB | C) = P(A | C)P(B | C)。$$

对于下述命题，从概率的公理化定义和条件概率的定义出发给出证明，或给出反例。

- (1) 事件 A 和 B 在事件 C 发生时是条件独立的，且有 $0 < P(C) < 1$ ，则事件 A 和 B 在事件 \bar{C} 发生时条件独立。这里，事件 \bar{C} 是事件 C 的对立事件。
- (2) 事件 A 和 B 相互独立，则对于任意事件 C ，若 $P(C) > 0$ ，事件 A 和 B 在事件 C 发生时是条件独立的。
- (3) 事件 A 和 B 相互独立，则事件 A 和事件 \bar{B} 相互独立。这里，事件 \bar{B} 是事件 B 的对立事件。

第三题

在课上，我们考虑了如下球与桶模型：有 $n \geq 1$ 个球，每个球都等可能被放到 $m \geq 1$ 个桶中的任一个。用 $P_{n,m}$ 表示每个桶中至多有一个球的概率。在课上，我们已经证明了，

$$P_{n,m} \leq e^{-\frac{n(n-1)}{2m}}。$$

现在, 请证明

$$P_{n,m} \geq e^{-\frac{n(n-1)}{2m}} \cdot \left(1 - O\left(\frac{n^3}{m^2}\right)\right).$$

提示: 如果上面的大 O 记号对你来说很难理解, 你也可以选择证明

$$P_{n,m} \geq e^{-\frac{n(n-1)}{2m}} \cdot \left(1 - \frac{10086n^3}{m^2}\right).$$

第四题

共有 n 个玩家, 其中正整数 $n \geq 2$ 为常量。令 X 等概率取 $\{1, 2, \dots, n-1\}$ 中的值, 并从 n 个玩家中随机选取 X 个玩家组成队伍 A , 每个大小为 X 的玩家子集被选为队伍 A 的概率均相同。令 B 为不在队伍 A 中的全部 $n-X$ 个玩家。

- (1) 对于任意 $1 \leq k < n$, 计算第一个玩家所在队伍大小为 k 的概率。
- (2) 每个队伍从其成员中随机选出一个队长。给定第一个玩家为所在队伍的队长, 计算第一个玩家所在队伍大小为 k 的概率。

第五题

疾病在人群中的患病率为 p 。现有两种快速检测:

- 检测 1: 对于患者, 检测为阳性的概率为 p_1 ; 对于未患病人群, 检测为阳性的概率为 q_1 ;
- 检测 2: 对于患者, 检测为阳性的概率为 p_2 ; 对于未患病人群, 检测为阳性的概率为 q_2 。

对同一个受检者同时做两种检测, 假设在“患病/不患”的条件下两次检测结果相互独立。若观察到至少有一个检测呈阳性, 求该受检者实际患病的概率。

第六题

考虑一个有 n 名选手的锦标赛, 选手编号为 $1, 2, \dots, n$ 。这是一个循环赛, 意味着每对不同的选手之间都进行一场比赛。因此, 总共进行了 $\binom{n}{2} = \frac{n(n-1)}{2}$ 场比赛。比赛没有平局, 每场比赛都必须分出胜负。

我们假设比赛结果是随机的。对于任意一对选手 a 和 b , a 战胜 b 的概率为 $1/2$, 同样, b 战胜 a 的概率也为 $1/2$ 。所有比赛的结果都是相互独立的。

- (1) 设整数 k 满足 $1 \leq k < n$, 取一个大小为 k 的选手子集 $V \subseteq \{1, 2, \dots, n\}$, 并任取一个不在 V 中的选手 v 。定义事件 $A_{V,v}$ 为: 选手 v 战胜了集合 V 中的全部选手。计算事件 $A_{V,v}$ 的概率 $P(A_{V,v})$ 。
- (2) 设整数 k 满足 $1 \leq k < n$, 取一个大小为 k 的选手子集 $V \subseteq \{1, 2, \dots, n\}$ 。定义事件 B_V 为: 存在至少一名选手 $v \in \{1, 2, \dots, n\} \setminus V$, 该选手战胜了集合 V 中的所有选手。也即, $B_V = \bigcup_{v \notin V} A_{V,v}$ 。计算事件 B_V 的概率 $P(B_V)$ 。
- (3) 设整数 k 满足 $1 \leq k < n$, 定义事件 C 为: 对于任意一个大小为 k 的选手子集 V , 都存在一名不属于 V 的选手 v 战胜了 V 中的所有选手。也即, $C = \bigcap_{V: |V|=k} B_V$ 。证明 $P(C) \geq 1 - \binom{n}{k}(1-2^{-k})^{n-k}$ 。
- (4) 对于任意正整数 k , 证明存在一个正整数 $n = O(k^2 \cdot 2^k)$ 和 n 个选手的比赛结果, 且对于该比赛结果, 对于任意 k 名选手的组合 $\{v_1, v_2, \dots, v_k\}$, 都存在另一名选手 v 同时战胜了这 k 名选手。

提示: 如果上面的大 O 记号对你来说很难理解, 你也可以选择证明, 对于任意正整数 k , 存在正整数 n 满足 $n \leq 12345 \cdot k^2 \cdot 2^k$ 且上述条件成立。另外, 请回顾课上/作业零中的不等式。