作业 4：无模型方法一——蒙特卡洛

陈达贵　深蓝学院

2018-12-28

1　文字作业

1. (2 分) 课堂上提到了蒙特卡洛方法的估计是无偏估计，但是方差较大。
   - 证明蒙特卡洛方法得到的 $V(s)$ 是 $v_\pi(s)$ 的无偏估计
   - 方差较大，代表每一次更新的置信度较低。我们可以通过增加采样量的方式减小方差，请给出方差和采样样本数量之间的关系
   - 常量步长的增量式蒙特卡洛等价于指数平均。请证明指数平均下的 $V(s)$ 是 $v_\pi(s)$ 的渐进无偏估计
2. (1 分) 课堂上提到重要性采样会显著增加方差，请证明（或举例说明）为什么使用重要性采样会显著增加方差
3. (1 分) 课堂上给出了增量式离策略每次拜访蒙特卡洛策略评价算法，请将每次拜访替换成首次拜访，给出相应的算法

2　编程作业

1. (6 分) 使用第三课的编程作业中的格子迷宫环境。需要实现以下算法。(在更新算法时，必须时无模型的，即不能使用 $\mathcal{P}, \mathcal{R}$), 可以与第三课的作业进行对比

- (在策略) 使用增量式蒙特卡洛的方法，计算随机策略的值函数。
- 使用常量步长，计算随机策略的值函数。对比不同的值函数初始化下 (比如把值函数的初始值设为一个均值很大的正态分布等) 常量步长和一般增量式蒙特卡洛的收敛性。
- 实现 GLIE 蒙特卡洛优化算法，求出最优策略。
- 尝试在蒙特卡洛方法中使用贪婪的策略提升，是否能找出最优策略？跟什么有关?
- (离策略) 选择行为策略 μ 为随机策略，然后使用增量式离策略每次拜访蒙特卡洛优化算法，求出此时的最优策略。

注：由于这里是一个连续性环境，而 MC 方法只能适用于片段性任务，因此我们这里可以设置一个片段的最大步数，不过求出来可能会有一定的误差