1. 背景知识
    a. 关于所讲内容的必要简言
        i. 需要理解的 **关键而基础的概念**
        ii. Policy 和 Policy function（策略 和 策略函数）
            - 策略决定了：Agent *选择* **要执行的 action** *的方式*
            - Policy function 则是给出 **action** 的函数。
        - Value 和 Value function（值 和 值函数）
            - Value function 表示 **Agent** 处于特定状态的好坏。
            - 它取决于 Policy，且通常用 v(s) 表示。
            - 在这里，**value** 等于 **the total expected reward** *received by the agent* **starting from the current state**.
    b. 蒙特卡罗方法的简介。
        i. 也称统计模拟方法，是在1940年代中期，由于科学技术的发展和电子计算机的发明，而提出的一种以概率统计理论为指导的数值计算方法。是指使用随机数来解决很多计算问题的方法。
        ii. 蒙特卡罗方法于美国在第二次世界大战中研制原子弹的“曼哈顿计划”，计划的成员S.M.乌拉姆和J.冯·诺伊曼提出。数学家冯·诺伊曼用驰名世界的赌城—摩纳哥的Monte Carlo—来命名这种方法，为它蒙上了一层神秘色彩。在这之前，蒙特卡罗方法就已经存在。1777年，法国数学家布丰（Georges Louis Leclere de Buffon，1707—1788）提出用投针实验的方法求圆周率π。这被认为是蒙特卡罗方法的起源。
    c. 重要性
        i. Monte Carlo is one of the most popular and most commonly used algorithms in various fields ranging from physics and mechanics to computer science.
    d. 在强化学习中，使用蒙特卡罗方法可以解决的问题
        i. 一般的无模型问题
            1. The model of the environment is not known.
        ii. It is very powerful for finding optimal policies when we don't have enough knowledge of the environment.
2. 基本思想和算法流程

a. 基本思想：
    i. 通过随机抽样求取近似解
    ii. It is a statistical technique to find an approximate answer through sampling.
b. 算法整体流程：
    i. 让 Agent 和 环境交互后得到交互序列
    ii. 通过交互序列计算出每个state的价值
    iii. 将这些价值累积到值函数中进行更新
    iv. 根据更新后的值函数，来更新策略

3. 蒙特卡罗预测
    a. 作用：
        i. 使用蒙特卡罗预测，可以估计 给定策略的值函数
4. 蒙特卡罗控制
    a. 对比 control 与 prediction 的作用
        i. Monte Carlo prediction：对 给定策略的值函数 进行估计。
        ii. Monte Carlo control：对 值函数、 策略不断进行优化，从而使得值函数更加准确。
5. 实例
    a. 玩游戏：21点游戏（也叫Blackjack, 是一种流行的纸牌游戏）。
        i. 题外话：
            1. 更复杂的游戏，蒙特卡罗方法（其变体和核心思想）也能派上用场，比如AlphaZero 中应用到的MCTS。
    b. 用蒙特卡罗估计 pi 的值