4.3

Problem

3．序贯投资问题。设投资者有 D 元资金，有 N 个投资周期。每个投资周期中出现投资机会的概率是 p ，它与历史无关。如果投资机会出现，则投资者必须决定是否投资，以及在他的剩余资金中拿出多少来投资。如果投资 y ，则他将在周期 N 末时获得收益 $R(y)$ 。试写出此投资问题的MDP模型及最优方程。进而，若 $R(y)=(1+r)y, r>0$ ，尝试求解 $N=2$ 时的最优投资策略。

Solution

定义 $V_i(x)$ 为在周期 i 开始时，拥有资金 x 的情况下，从周期 i 到周期 N 的最大期望总收益。
－对于 $i=1,2,\ldots,N-1$ ：
$V_i(x)=p\cdot\max_{0\leq y\leq x}\{R(y)+V_{i+1}(x-y)\}+(1-p)\cdot V_{i+1}(x)$
－对于 $i=N$ ：
$V_N(x)=p\cdot\max_{0\leq y\leq x}R(y)+(1-p)\cdot 0$

其中，当没有投资机会时，收益为 0 。

当 $R(y)=(1+r)y, r>0$ 时，求解 $N=2$ 的最优投资策略计算 $V_2(x)$

$V_2(x)=p\cdot\max_{0\leq y\leq x}(1+r)y=p(1+r)x$

因为 $(1+r)y$ 是 y 的线性增函数，所以最优解为 $y=x$.

计算 $V_1(x)$

$V_1(x)=p\cdot\max_{0\leq y\leq x}\{(1+r)y+V_2(x-y)\}+(1-p)\cdot V_2(x)$

代入 $V_2(x-y)=p(1+r)(x-y)$:

$V_1(x)=p\cdot\max_{0\leq y\leq x}\{(1+r)y+p(1+r)(x-y)\}+(1-p)\cdot p(1+r)x$

有:

$\max_{0\leq y\leq x}(1+r)[px+(1-p)y]=(1+r)x$

因此：

$$V_1(x) = p \cdot (1+r)x + (1-p) \cdot p(1+r)x = p(1+r)x[1+(1-p)] = p(1+r)x(2-p)$$

可得结果：

在周期 1，如果投资机会出现，则最优投资额为 $y = x$ ，即投资所有资金。
在周期 2 ，如果投资机会出现，则最优投资额为 $y = x$ ，即投资所有剩余资金。

4.4

Problem

4． 52 张扑克牌一张一张地翻过来，在翻过来之前，你有机会猜测这张牌是否是梅花A。假定只能猜一次，目标是使得猜中的概率达到最大。
（1）建立此问题的动态规划模型；
（2）最优策略是什么：
（3）如果可以猜 n 次，试求此时的最优策略：
（4）如果是猜梅花牌（不仅仅是梅花A)，则最优策略又是怎样的?

Solution (1)

最优方程如下：

- 当 $k=1$ 时，$V(1)=1$ 。
- 当 $k \geq 2$ 时，$V(k) = \max\left\{\frac{1}{k}, \frac{k-1}{k}V(k-1)\right\}$ 。

其中：

- $\frac{1}{k}$ 表示猜测当前牌是梅花A的期望收益（猜中概率为 $\frac{1}{k}$ ）。
- $\frac{k-1}{k}V(k-1)$ 表示不猜测当前牌时的期望收益（翻牌后如果不是梅花 A ，则进入状态 $k-1$ ）。

Solution (2)

由动态规划模型计算可得 $V(k) = \frac{1}{k}$ 对于所有 k 。因此，对于初始状态 $k=52$ ，猜中梅花A的概率为 $\frac{1}{52}$。最优策略是：在任意一张牌之前猜测它是梅花A，猜中的概率均为 $\frac{1}{52}$ 。策略选择不影响概率，因此任何在某一时刻猜测的策略都是最优的。

Solution (3)

定义值函数 $V(k,m)$ 表示剩余 k 张牌且有 m 次猜测机会时，猜中梅花A的最大概率。其中 $m \geq 1$ 。最优方程如下：

- 如果 $m=0$ ，则 $V(k,0)=0$（无猜测机会，赢概率为 0 ）。
- 如果 $k=0$ ，则 $V(0,m)=0$（无牌可翻，赢概率为 0 ）。
- 对于 $k \geq 1$ 且 $m \geq 1$ ：
  $V(k,m)=\max\left\{\frac{1}{k}+\frac{k-1}{k}V(k-1,m-1), \frac{k-1}{k}V(k-1,m)\right\}$

其中：
$-\frac{1}{k}+\frac{k-1}{k}V(k-1,m-1)$ 表示猜测当前牌的期望收益（猜中则赢，猜错则进入状态 $(k-1,m-1)$)。
。 $\frac{k-1}{k}V(k-1,\ m)$ 表示不猜测当前牌的期望收益（翻牌后如果不是梅花A，则进入状态 $(k-1,\ m)$ ）。

通过计算可得 $V(k,m)=\min\left(1,\frac{m}{k}\right)$ 。对于初始状态 $k=52$ 和 $m=n$ ，猜中概率为 $\frac{n}{52}$ 。

最优策略是：

- 当剩余牌数 k 大于剩余猜测次数 m 时，不猜测。
- 当剩余牌数 k 小于或等于剩余猜测次数 m 时，猜测当前牌。此策略确保猜中概率为 $\frac{n}{52}$ 。

Solution (4)

定义状态为剩余牌数 k 和剩余梅花牌数 s（初始 $k=52, s=13$ ）。值函数 $W(k,s)$ 表示在状态（ k,s ）时猜中梅花牌的最大概率。

最优方程如下：

- 如果 $s=0$ ，则 $W(k,0)=0$（无梅花牌，赢概率为 0 ）。
- 如果 $k=0$ ，则 $W(0,s)=0$（无牌可翻，赢概率为 0 ）。
- 对于 $k \geq 1$ 且 $s \geq 1$ ：
  $W(k,s)=\max\left\{\frac{s}{k}, \frac{k-s}{k}W(k-1,s)\right\}$

其中：

- $\frac{s}{k}$ 表示猜测当前牌是梅花牌的期望收益。
- $\frac{k-s}{k}W(k-1,s)$ 表示不猜测当前牌时的期望收益（翻牌后如果不是梅花牌，则进入状态 $(k-1,s)$ ）。

通过计算可得 $W(k,s)=\frac{s}{k}$ 。对于初始状态 $k=52, s=13$ ，猜中梅花牌的概率为 $\frac{13}{52}=\frac{1}{4}$ 。最优策略是：在第一次机会时猜测当前牌是梅花牌。此策略使猜中概率达到最大，且任何其他策略均不能超过此概率。

习题

Problem

总结第4.2-4.3节中建模的步骤、证明最优方程的步骤。

Solution

一、建模步骤（多阶段动态决策与马尔可夫决策过程）

1．定义阶段

- 将决策过程划分为若干个阶段（或周期），记为 $n=0,1,\ldots,N$ 。
- 初始阶段为 $n=0$ ，终止阶段为 $n=N$ 。

2．定义状态集

- 每个阶段 n 有一个状态集 S_n ，表示系统在该阶段所有可能的状态。
- 状态应包含影响决策和未来收益的所有信息。

3．定义决策集

- 在每个状态 $i \in S_n$ 下，有一个可用的决策集 $A_n(i)$ 。
- 决策是决策者在该状态下可采取的行动。

4．定义报醵函数

- 在阶段 n 、状态 i 、决策 a 下，系统获得报酵 $r_n(i,a)$ 。
- 报酬可以是确定性的，也可以是期望值。

5．定义状态转移机制

- 确定性转移：$T_n(i,a)$ 表示在阶段 n 、状态 i 、决策 a 下，系统在下一阶段的状态。
- 随机性转移（MDP）：引入状态转移概率 $p_n(j \mid i,a)$ ，表示在阶段 n 、状态 i 、决策 a 下，系统转移到状态 j 的概率。

6．定义目标函数

－目标是最大化从阶段 n 到 N 的总报酸（或期望总报酬）：

$V_n(\pi, i)=\sum_{k=n}^N r_k\left(i_k, a_k\right)$

或带折扣因子的形式。

二、证明最优方程的步聚

1．引入子问题概念

－定义从阶段 n 开始的子问题，其最优值函数为 $V_n(i)$ 。

2．建立目标函数关系

－策略 π 下的总报酸满足：

$V_n(\pi, i)=r_n(i, a)+V_{n+1}\left(\pi, T_n(i, a)\right)$

3．推导最优方程

－最优值函数应满足：

$V_n(i)=\sup _{a \in A_n(i)}\left\{r_n(i, a)+V_{n+1}\left(T_n(i, a)\right)\right\}$

对于随机情形（MDP）为：

$V_n(i)=\sup _{a \in A_n(i)}\left\{r_n(i, a)+\sum_j p_n(j \mid i, a) V_{n+1}(j)\right\}$

4．给出边界条件

－在终止阶段 $N+1$ 设定：

$V_{N+1}(i)=0$

5．严格证明最优方程

－使用上确界定义和数学归纳法，证明：

$-V_n(i) \leq \sup _a\left\{r_n(i, a)+V_{n+1}\left(T_n(i, a)\right)\right\}$

$-V_n(i) \geq \sup _a\left\{r_n(i, a)+V_{n+1}\left(T_n(i, a)\right)\right\}$

－从而得出等式成立。

6．构造最优策略

－若在每个阶段 n 和状态 i 中，决策 $f_n^*(i)$ 都能取到最优方程中的上确界，则策略 $\pi^*=\left(f_0^*, f_1^*, \ldots, f_N^*\right)$ 是最优策路。