Markov Decision Process II

2023年10月5日 星期四　　下午9:31

定理：V 是 Bellman 方程的唯一的解. 更进一步的. 如果 $I^*(x) \in \underset{a\in A}{\operatorname{Argmax}} \{R(x,a) + \sum_{z\in S} p_a(x,z) V(z)\}$

则有 $\alpha^*(n,x) = I^*(x)$ 是一个最优控制.

证明（待上次）：$U(x,\alpha) = E_x\left[\sum_{n\geq 0} \lambda^n R(X_n^\alpha, \alpha(n, X_n^\alpha))\right]$

$$E_x \sum_{n\geq 0} \lambda^n V(X_n^{\alpha^*}) = E_x \sum_{n\geq 0} \lambda^n \left[R(X_n^{\alpha^*}, \alpha^*(X_n^{\alpha^*})) + \lambda \sum_{z\in S} p_{\alpha^*}(X_n^{\alpha^*}, z) V(z)\right].$$

$$= U(x,\alpha^*) + E_x \sum_{n\geq 0} \lambda^{n+1} E_x\left[V(X_{n+1}^{\alpha^*}) \mid X_n^{\alpha^*}\right]$$

$$= U(x,\alpha^*) + \sum_{n\geq 0} \lambda^{n+1} E_x\left[V(X_{n+1}^{\alpha^*})\right].$$

$$\Rightarrow V(x) = U(x,\alpha^*).$$

例子：假设代理人有随机的资金 X_n. 在每个时间节点消耗了一些，把消耗量视为控制，R 是效用函数，代理人的消耗资金的功效　代理人希望最大化 $E_x\left[\sum_{n=0} \frac{1}{(1+r)^n} R(X_n^\alpha, \alpha(n, X_n^\alpha))\right]$，$r$ 是利率.

Bellman 方程：$V(x) = \sup_a \{R(x,a) + \sum_{z\in S} p(x,a,z) V(z)\}$