马尔科夫属性

定义如下随机过程：

$$P_r\{S_{t+1}=s', R_{t+1}=r \mid S_0, A_0, R_1, \cdots, S_{t-1}, A_{t-1}, R_t, S_t, A_t\}$$

也就是说在过去所有的历史信息的情况下，Agent 在状态 S_t，采取动作 A_t，进入新状态

$(S_0, A_0, R_1, \cdots, S_{t-1}, A_{t-1}, R_t)$

$S_{t+1}=s'$ 且 Reward $=r$ 的联合概率分布。

如果系统满足马尔科夫属性，则所有历史信息都压缩在当前状态 S_t，在给定 S_t 的条件下与过去的历史无关，我们有：

$$p(s', r \mid s, a) \equiv P_r\{S_{t+1}=s', R_{t+1}=r \mid S_t=s, A_t=a\} \quad ①$$

马尔科夫决策过程 (Markov Decision Processes)

这个过程即定义①描述的过程。

1. 计算在状态 $S_t=s, A_t=a$ 时的期望回报

$$r(s,a) \equiv E[R_{t+1} \mid S_t=s, A_t=a] = \sum_{r\in R} r\, P_r(R_{t+1}=r \mid S_t=s, A_t=a)$$

Note：边缘分布是联合分布的求和或积分 $= \sum_{r\in R} r \sum_{s'\in S} p(s', r \mid s, a)$

2. 计算状态转移概率

$$p(s' \mid s, a) \equiv P_r(S_{t+1}=s' \mid S_t=s, A_t=a) = \sum_{r\in R} p(s', r \mid s, a)$$