

1. Markov Chains

Suppose that $(S_t)_{t \in \mathbb{N}}$ is a Markov Chain with values in some finite set \mathcal{S} on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by P the transition matrix. Assume further that $\mathbb{P}(S_n = s') > 0$ for some $n \in \mathbb{N}$ and $s' \in \mathcal{S}$ and define the probability measure $\tilde{\mathbb{P}} = \mathbb{P}(\cdot \mid S_n = s')$. Prove that the shifted process $(\tilde{S}_t)_{t \in \mathbb{N}} = (S_{t+n})_{t \in \mathbb{N}}$ is again a Markov chain on $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ with transition matrix P .

Solution:

We prove this by simply calculating the path probabilities. Let all $s_i \in \mathcal{S}$. Then with the definition of \tilde{S} and the fact that P is the transition matrix of S .

$$\begin{aligned}
 \tilde{\mathbb{P}}(\tilde{S}_t = s_t, \dots, \tilde{S}_1 = s_1) &= \mathbb{P}(S_{t+n} = s_t, \dots, S_{t+1} = s_1 \mid S_n = s') \\
 &= \frac{\mathbb{P}(S_{t+n} = s_t, \dots, S_{t+1} = s_1, S_n = s')}{\mathbb{P}(S_n = s')} \\
 &= \frac{\sum_{s'_0, \dots, s'_{n-1}} \mathbb{P}(S_{t+n} = s_t, \dots, S_{t+1} = s_1, S_n = s', S_{n-1} = s'_{n-1}, \dots, S_0 = s'_0)}{\sum_{s'_0, \dots, s'_{n-1}} \mathbb{P}(S_n = s', S_{n-1} = s'_{n-1}, \dots, S_0 = s'_0)} \\
 &= \frac{\sum_{s'_0, \dots, s'_{n-1}} \mu(s'_0) \left(\prod_{i=1}^{n-1} p_{s'_{i-1}, s'_i} \right) p_{s'_{n-1}, s'} p_{s', s_1} \left(\prod_{i=2}^t p_{s_{i-1}, s_i} \right)}{\sum_{s'_0, \dots, s'_{n-1}} \mu(s'_0) \left(\prod_{i=1}^{n-1} p_{s'_{i-1}, s'_i} \right) p_{s'_{n-1}, s'}} \\
 &= p_{s', s_1} \prod_{i=2}^t p_{s_{i-1}, s_i}.
 \end{aligned}$$

Assume that $\tilde{\mathbb{P}}(S_0 = s_0, \dots, \tilde{S}_t = s_t) > 0$. We follow the Markov property from the path probabi-

ities as follows

$$\begin{aligned}
& \tilde{\mathbb{P}}(\tilde{S}_{t+1} | \tilde{S}_0 = s_0, \dots, \tilde{S}_t = s_t) \\
&= \frac{\tilde{\mathbb{P}}(\tilde{S}_{t+1}, \tilde{S}_0 = s_0, \dots, \tilde{S}_t = s_t)}{\tilde{\mathbb{P}}(\tilde{S}_0 = s_0, \dots, \tilde{S}_t = s_t)} \\
&= \frac{p_{s', s_1} \prod_{i=2}^{t+1} p_{s_{i-1}, s_i}}{p_{s', s_1} \prod_{i=2}^t p_{s_{i-1}, s_i}} \\
&= p_{s_t, s_{t+1}} \\
&= \frac{\sum_{s_1, \dots, s_{t-1}} p_{s', s_1} \prod_{i=2}^{t+1} p_{s_{i-1}, s_i}}{\sum_{s_1, \dots, s_{t-1}} p_{s', s_1} \prod_{i=2}^t p_{s_{i-1}, s_i}} \\
&= \frac{\sum_{s_1, \dots, s_{t-1}} \tilde{\mathbb{P}}(\tilde{S}_{t+1} = s_{t+1} | \tilde{S}_t = s_t, \dots, \tilde{S}_1 = s_1)}{\sum_{s_1, \dots, s_{t-1}} \tilde{\mathbb{P}}(\tilde{S}_t = s_t, \dots, \tilde{S}_1 = s_1)} \\
&= \frac{\tilde{\mathbb{P}}(\tilde{S}_{t+1} = s_{t+1}, \tilde{S}_t = s_t)}{\tilde{\mathbb{P}}(\tilde{S}_t = s_t)} \\
&= \tilde{\mathbb{P}}(\tilde{S}_{t+1} | \tilde{S}_t = s_t)
\end{aligned}$$

2. Proof of Theorem 3.1.3

Complete the Proof of Theorem 3.1.3. from the lecture:

a) Show that defining \mathbb{P}_T on the singletons

$$\begin{aligned}
\mathbb{P}_T(\{(s_0, a_0, r_0, \dots, s_T, a_T, r_T)\}) &:= \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \prod_{i=1}^T p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \\
&\quad \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \cdot p(\mathcal{S} \times \{r_T\}; s_T, a_T).
\end{aligned}$$

yields a probability measure.

Solution:

We denote with Ω_T the space of all trajectories until time point T ,

$$\Omega_T = \{(s_0, a_0, r_0, \dots, s_T, a_T, r_T) : s_i \in \mathcal{S}, a_i \in \mathcal{A}_{s_i}, r_0 \in \mathcal{R}, \forall i \in \{0, \dots, T\}\}.$$

The corresponding σ -algebra is chosen to be the power set.

We will first prove that the total mass of the measure is indeed 1 by induction:

Induction beginning: Assume $T = 1$, then

$$\begin{aligned}
\mathbb{P}_1(\Omega_1) &= \sum_{s_0} \sum_{r_0} \sum_{a_0} \sum_{s_1} \sum_{r_1} \sum_{a_1} \mathbb{P}_T(\{(s_0, a_0, r_0, s_1, a_1, r_1)\}) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \sum_{s_1} \sum_{a_1} \sum_{r_1} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad p(\{(s_1, r_0)\}; s_0, a_0) \cdot \pi_i(\{a_1\}; s_0, a_0, s_1) p(\mathcal{S} \times \{r_1\}; s_1; a_1) \\
&= \sum_{s_0} \mu(\{s_0\}) \cdot \sum_{a_0} \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \sum_{s_1} \sum_{r_0} p(\{(s_1, r_0)\}; s_0, a_0) \cdot \sum_{a_1} \pi_i(\{a_1\}; s_0, a_0, s_1) \underbrace{\sum_{r_1} p(\mathcal{S} \times \{r_1\}; s_1; a_1)}_{=1} \\
&= \sum_{s_0} \mu(\{s_0\}) \cdot \sum_{a_0} \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \sum_{s_1} \sum_{r_0} p(\{(s_1, r_0)\}; s_0, a_0) \cdot \underbrace{\sum_{a_1} \pi_i(\{a_1\}; s_0, a_0, s_1)}_{=1} \\
&= \sum_{s_0} \mu(\{s_0\}) \cdot \sum_{a_0} \pi_0(\{a_0\}; s_0) \cdot \underbrace{\sum_{s_1} \sum_{r_0} p(\{(s_1, r_0)\}; s_0, a_0)}_{=1} \\
&= \sum_{s_0} \mu(\{s_0\}) \cdot \underbrace{\sum_{a_0} \pi_0(\{a_0\}; s_0)}_{=1} \\
&= \underbrace{\sum_{s_0} \mu(\{s_0\})}_{=1} \\
&= 1
\end{aligned}$$

Induction assumption: Assume the claim holds for $T - 1$.

Induction conclusion: Then for T ,

$$\begin{aligned}
\mathbb{P}_T(\Omega_T) &= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_T} \sum_{a_T} \sum_{r_T} \mathbb{P}_T(\{(s_0, a_0, r_0, \dots, s_T, a_T, r_T)\}) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_T} \sum_{a_T} \sum_{r_T} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^T p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \cdot p(\mathcal{S} \times \{r_T\}; s_T, a_T) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{T-1}} \sum_{a_{T-1}} \sum_{r_{T-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{T-1} p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \cdot \sum_{s_T} \sum_{r_T} p(\{(s_T, r_{T-1})\}; s_{T-1}, a_{T-1}) \cdot \underbrace{\sum_{a_T} \pi_T(\{a_T\}; s_0, a_0, \dots, a_{T-1}, s_T)}_{=1} \\
&\quad \cdot \underbrace{\sum_{r_T} p(\mathcal{S} \times \{r_T\}; s_T, a_T)}_{=1} \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{T-1}} \sum_{a_{T-1}} \sum_{r_{T-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{T-1} p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \sum_{s_T} \sum_{r_T} p(\{(s_T, r_{T-1})\}; s_{T-1}, a_{T-1}) \cdot \underbrace{\sum_{a_T} \pi_T(\{a_T\}; s_0, a_0, \dots, a_{T-1}, s_T)}_{=1} \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{T-1}} \sum_{a_{T-1}} \sum_{r_{T-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{T-1} p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \sum_{s_T} \sum_{r_T} p(\{(s_T, r_{T-1})\}; s_{T-1}, a_{T-1}) \\
&\quad \underbrace{\hspace{10em}}_{=1} \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{T-1}} \sum_{a_{T-1}} \sum_{r_{T-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{T-1} p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&= \mathbb{P}_{T-1}(\Omega_{T-1}) \\
&= 1
\end{aligned}$$

This proves that the total mass of the measure is 1. Next, we show that the σ -additivity. Therefore we will assume a sequence of disjoint sets $E_n \in \mathcal{F}_T$, $n \in \mathbb{N}$, and denote by

$\tau = (s_0, a_0, r_0, \dots, s_T, a_t, r_t)$ the elements of Ω_T . Moreover, we write

$$p(\tau) = \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \prod_{i=1}^T p(\{(s_i, r_{i-1})\}; s_{i-1}, a_{i-1}) \\ \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \cdot p(\mathcal{S} \times \{r_T\}; s_T, a_T)$$

to save notation. Now it holds that

$$\mathbb{P}_T(\bigsqcup_{n \in \mathbb{N}} E_n) = \sum_{\tau \in \bigsqcup_{n \in \mathbb{N}} E_n} p(\tau) \\ = \sum_{n \in \mathbb{N}} \sum_{\tau \in E_n} p(\tau) \\ = \sum_{n \in \mathbb{N}} \mathbb{P}_T(E_n).$$

This completes the proof.

b) Check the claimed conditional probability identity

$$\mathbb{P}_\mu^\pi(S_{t+1} = s_{t+1}, R_t = r_t | S_t = s, A_t = a) = p(s_{t+1}, r_t; s, a).$$

where we assume that $\mathbb{P}_\mu^\pi(S_t = s, A_t = a) > 0$.

Solution:

First we have,

$$\begin{aligned}
& \mathbb{P}_\mu^\pi(S_{t+1} = s_{t+1}, R_t = r_t, S_t = s, A_t = a) \\
&= \mathbb{P}_{t+1}(S_0 \in \mathcal{S}, A_0 \in \mathcal{A}, R_0 \in \mathcal{R}, \dots, S_t = s, A_t = a, R_t = r_t, S_{t+1} = s_{t+1}, A_{t+1} \in \mathcal{A}, R_{t+1} \in \mathcal{R}) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \sum_{a_{t+1}} \sum_{r_{t+1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
& \quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
& \quad \cdot p(\{s, r_{t-1}\}; s_{t-1}, a_{t-1}) \cdot \pi_i(\{a\}; s_0, a_0, \dots, a_{t-1}, s) \\
& \quad \cdot p(\{s_{t+1}, r_t\}; s, a) \cdot \pi_i(\{a_{t+1}\}; s_0, a_0, \dots, s, a, s_{t+1}) \\
& \quad \cdot p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1}) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
& \quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
& \quad \cdot p(\{s, r_{t-1}\}; s_{t-1}, a_{t-1}) \cdot \pi_i(\{a\}; s_0, a_0, \dots, a_{t-1}, s) \\
& \quad \cdot p(\{s_{t+1}, r_t\}; s, a) \cdot \underbrace{\sum_{a_{t+1}} \pi_i(\{a_{t+1}\}; s_0, a_0, \dots, s, a, s_{t+1})}_{=1} \\
& \quad \cdot \underbrace{\sum_{r_{t+1}} p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1})}_{=1} \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
& \quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
& \quad \cdot p(\{s, r_{t-1}\}; s_{t-1}, a_{t-1}) \cdot \pi_i(\{a\}; s_0, a_0, \dots, a_{t-1}, s) \cdot p(\{s_{t+1}, r_t\}; s, a) \\
&= p(\{s_{t+1}, r_t\}; s, a) \\
& \quad \cdot \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
& \quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
& \quad \cdot p(\{s, r_{t-1}\}; s_{t-1}, a_{t-1}) \cdot \pi_i(\{a\}; s_0, a_0, \dots, a_{t-1}, s)
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{P}_\mu^\pi(S_t = s, A_t = a) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \sum_{r_t} \mathbb{P}_t(S_0 = s_0, A_0 = a_0, R_0 = r_0 \dots, A_{t-1} = a_{t-1}, S_t = s, A_t = a, R_t = r_t) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \sum_{r_t} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \times p(\{(s, r_{t-1})\}; s_{t-1}, a_{t-1}) \cdot \pi_t(\{a\}; s_0, a_0, \dots, s_{t-1}, a_{t-1}, s) \cdot p(\mathcal{S} \times \{r_t\} | s, a) \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \times p(\{(s, r_{t-1})\}; s_{t-1}, a_{t-1}) \cdot \pi_t(\{a\}; s_0, a_0, \dots, s_{t-1}, a_{t-1}, s) \cdot \underbrace{\sum_{r_t} p(\mathcal{S} \times \{r_t\} | s, a)}_{=1} \\
&= \sum_{s_0} \sum_{a_0} \sum_{r_0} \cdots \sum_{s_{t-1}} \sum_{a_{t-1}} \sum_{r_{t-1}} \mu(\{s_0\}) \cdot \pi_0(\{a_0\}; s_0) \cdot \\
&\quad \prod_{i=1}^{t-1} p(\{s_i, r_{i-1}\}; s_{i-1}, a_{i-1}) \cdot \pi_i(\{a_i\}; s_0, a_0, \dots, a_{i-1}, s_i) \\
&\quad \times p(\{(s, r_{t-1})\}; s_{t-1}, a_{t-1}) \cdot \pi_t(\{a\}; s_0, a_0, \dots, s_{t-1}, a_{t-1}, s).
\end{aligned}$$

Hence, using the definition of conditional probability the fraction cancels to give

$$\mathbb{P}_\mu^\pi(S_{t+1} = s_{t+1}, R_t = r_t | S_t = s, A_t = a) = p(s_{t+1}, r_t; s, a).$$

3. Probabilistic interpretation of MDPs

Use the formal definition of the stochastic process (S, A, R) to check that

$$p(s'; s, a) = \mathbb{P}(S_{t+1} = s' | S_t = s, A_t = a),$$

$$p(r; s, a) = \mathbb{P}(R_t = r | S_t = s, A_t = a),$$

$$r(s, a) = \mathbb{E}[R_t | S_t = s, A_t = a],$$

$$r(s, a, s') = \mathbb{E}[R_t | S_t = s, A_t = a, S_{t+1} = s']$$

holds if the events in the condition have positive probability, for any $t \in \mathbb{N}_0$.

Solution:

For the first claim we have by Exercise 2 b) that

$$\begin{aligned}
p(s'; s, a) &:= p(\{s'\} \times \mathcal{R}; s, a) \\
&= \mathbb{P}_\mu^\pi(S_{t+1} \in \{s'\}, R_t \in \mathcal{R} | S_t = s, A_t = a) \\
&= \mathbb{P}_\mu^\pi(S_{t+1} = s' | S_t = s, A_t = a).
\end{aligned}$$

For the second Claim we have similarly

$$\begin{aligned}
p(r; s, a) &:= p(\mathcal{S} \times \{r\}; s, a) \\
&= \mathbb{P}_\mu^\pi(S_{t+1} \in \mathcal{S}, R_t = r \mid S_t = s, A_t = a) \\
&= \mathbb{P}_\mu^\pi(R_t = r \mid S_t = s, A_t = a).
\end{aligned}$$

For the next equation we have similar to above

$$\begin{aligned}
r(s, a) &:= \sum_r r p(\mathcal{S} \times \{r\}; s, a) \\
&= \sum_r r \mathbb{P}_\mu^\pi(S_{t+1} \in \mathcal{S}, R_t \in \{r\} \mid S_t = s, A_t = a) \\
&= \sum_r r \mathbb{P}_\mu^\pi(R_t = r \mid S_t = s, A_t = a) \\
&= \mathbb{E}_\mu^\pi[R_t \mid S_t = s, A_t = a],
\end{aligned}$$

where we have used the definition of expectation for discrete random variables in the last equation. The last claim can be proven as follows

$$\begin{aligned}
r(s, a, s') &:= \sum_r r \frac{p(s', r; s, a)}{p(s'; s, a)} \\
&= \sum_r r \frac{\mathbb{P}_\mu^\pi(S_{t+1} = s', R_t = r \mid S_t = s, A_t = a)}{\mathbb{P}_\mu^\pi(S_{t+1} = s' \mid S_t = s, A_t = a)} \\
&= \sum_r r \frac{\frac{\mathbb{P}_\mu^\pi(S_{t+1}=s', R_t=r, S_t=s, A_t=a)}{\mathbb{P}_\mu^\pi(S_t=s, A_t=a)}}{\frac{\mathbb{P}_\mu^\pi(S_{t+1}=s', S_t=s, A_t=a)}{\mathbb{P}_\mu^\pi(S_t=s, A_t=a)}} \\
&= \sum_r r \frac{\mathbb{P}_\mu^\pi(S_{t+1} = s', R_t = r, S_t = s, A_t = a)}{\mathbb{P}_\mu^\pi(S_{t+1} = s', S_t = s, A_t = a)} \\
&= \sum_r r \mathbb{P}_\mu^\pi(R_t = r \mid S_{t+1} = s', S_t = s, A_t = a) \\
&= \mathbb{E}_\mu^\pi[R_t \mid S_t = s, A_t = a, S_{t+1} = s'].
\end{aligned}$$

4. Proof of Proposition 3.1.13

Prove that if $\pi \in \Pi_S$ then $(S_t, A_t, R_t)_{t \in \mathbb{N}}$ satisfies the Markov reward process property

$$\begin{aligned}
\mathbb{P}((S_{t+1}, A_{t+1}, R_{t+1}) = (s_{t+1}, a_{t+1}, r_{t+1}) \mid (S_t, A_t) = (s_t, a_t), \dots, (S_0, A_0) = (s_0, a_0)) \\
= \mathbb{P}((S_{t+1}, A_{t+1}, R_{t+1}) = (s_{t+1}, a_{t+1}, r_{t+1}) \mid (S_t, A_t) = (s_t, a_t))
\end{aligned}$$

with time-homogeneous state/reward transition probabilities

$$p_{(s,a),(s',a',r')} = p(s', r; s, a)\pi(a'; s')$$

Solution:

We use the calculations for the path probabilities from the lecture and the definition of a stationary policy. Further we denote by $*$ the sum over all r_t and s_i, a_i, r_i for i from 0 to $t-1$, i.e.

$\sum_* = \sum_{s_0, a_0, r_0} \cdots \sum_{s_{t-1}, a_{t-1}, r_{t-1}} \sum_{r_t}$. Then,

$$\begin{aligned}
& \mathbb{P}((S_{t+1}, A_{t+1}, R_{t+1}) = (s_{t+1}, a_{t+1}, r_{t+1}) | (S_t, A_t) = (s_t, a_t)) \\
&= \frac{\mathbb{P}(S_{t+1} = s_{t+1}, A_{t+1} = a_{t+1}, R_{t+1} = r_{t+1}, S_t = s_t, A_t = a_t)}{\mathbb{P}(S_t = s_t, A_t = a_t)} \\
&= \frac{\sum_* \mathbb{P}(S_0 = s_0, A_0 = a_0, R_0 = r_0, \dots, S_{t+1} = s_{t+1}, A_{t+1} = a_{t+1}, R_{t+1} = r_{t+1})}{\sum_* \mathbb{P}(S_0 = s_0, A_0 = a_0, R_0 = r_0, \dots, S_t = s_t, A_t = a_t, R_t = r_t)} \\
&= \frac{\sum_* \mu(s_0) \pi(a_0; s_0) \prod_{i=1}^{t+1} p(s_i, r_{i-1}; s_{i-1}, a_{i-1}) \pi(a_i; s_i) p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1})}{\sum_* \mu(s_0) \pi(a_0; s_0) \prod_{i=1}^t p(s_i, r_{i-1}; s_{i-1}, a_{i-1}) \pi(a_i; s_i) p(\mathcal{S} \times \{r_t\}; s_t, a_t)} \\
&= \frac{\sum_{r_t} p(s_{t+1}, r_t; s_t, a_t) \pi(a_{t+1}; s_{t+1}) p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1})}{\sum_{r_t} p(\mathcal{S} \times \{r_t\}; s_t, a_t)} \\
&= p(s_{t+1}; s_t, a_t) \pi(a_{t+1}; s_{t+1}) p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1}) \\
&= \frac{\sum_{r_0, \dots, r_t} \mu(s_0) \pi(a_0; s_0) \prod_{i=1}^{t+1} p(s_i; s_{i-1}, a_{i-1}) \pi(a_i; s_i) p(\mathcal{S} \times \{r_{t+1}\}; s_{t+1}, a_{t+1})}{\sum_{r_0, \dots, r_t} \mu(s_0) \pi(a_0; s_0) \prod_{i=1}^t p(s_i, r_{i-1}; s_{i-1}, a_{i-1}) \pi(a_i; s_i)} p(\mathcal{S} \times \{r_t\}; s_t, a_t) \\
&= \frac{\sum_{r_0, \dots, r_t} \mathbb{P}(S_{t+1} = s_{t+1}, A_{t+1} = a_{t+1}, R_{t+1} = r_{t+1}, S_t = s_t, A_t = a_t, R_t = r_t, \dots, S_0 = s_0, A_0 = a_0, R_0 = r_0)}{\sum_{r_0, \dots, r_t} \mathbb{P}(S_t = s_t, A_t = a_t, R_t = r_t, \dots, S_0 = s_0, A_0 = a_0, R_0 = r_0)} \\
&= \frac{\mathbb{P}(S_{t+1} = s_{t+1}, A_{t+1} = a_{t+1}, R_{t+1} = r_{t+1}, S_t = s_t, A_t = a_t, \dots, S_0 = s_0, A_0 = a_0)}{\mathbb{P}(S_t = s_t, A_t = a_t, \dots, S_0 = s_0, A_0 = a_0)} \\
&= \mathbb{P}((S_{t+1}, A_{t+1}, R_{t+1}) = (s_{t+1}, a_{t+1}, r_{t+1}) | (S_t, A_t) = (s_t, a_t), \dots, (S_0, A_0) = (s_0, a_0))
\end{aligned}$$