Class 5

Shikhar Saxena

January 17, 2023

Contents

Markov Chain as recursions 1
Markov Reward Process 2
Standard Optimization 3
Stochastic Optimization 3

Markov Chain as recursions

Theorem 1. *Consider recursion* $X_{n+1} = f(X_n, U_n),\ n \geq 0$ *where* $f : \mathcal{S} \times [0,1] \to \mathcal{S}$ *and* $\{U_n, n \geq 0\}$ *is an iid sequence of random variables. Then* $\{X_n, n \geq 0\}$ *defines a Markov chain with tpm*

$$P_{ij} = P(f(i, U) = j)$$

Conversely, every Markov Chain can be represented by such a recursion for some f *and* $\{U_n, n \geq 0\}$

U_n adds randomness to the deterministic Markov chain $X_{n+1} = f(X_n)$ for example.

Proof. The forward part is trivial. Converse follows from inverse transform method (simulation).

For the converse, we have CDF over the state space. So for $i \in \mathcal{S}$, we have $F_i(x)$: CDF of the i^{th} row in the tpm P.

Now, we simulate X_{n+1} using inverse transform method (given $X_n = i$):

$$X_{n+1} = f(i, U_n) = F_i^{-1}(U_n)$$

When F_i is discrete we have $F_i^{-1}(y) := \min\{x : F_i(x) \geq y\}$

$$X_{n+1} = f(X_n, U_n) = F_{X_n}^{-1}(U_n)$$

□

Remark. *U can be transformed to any distribution G. So this theorem applies for any general distribution.*

Markov Reward Process

Consider a Markov Chain $\{X_n, n \geq 0\}$ on $\mathcal{X}$ with $|\mathcal{X}| = M$.

- $r(X_t)$:= Reward obtained when in state X_t at time t
- $\beta \in (0, 1)$:= discount factor
- cumulative expected discounted reward (conditioned on starting in state x):

$$V(x) = \mathbb{E}_x \left[\sum_{t=0}^{\infty} \beta^t r(X_t) \right]$$

- $\mathbb{E}_x$: Conditional Expectation of starting in x

Lemma 1. *$V(x)$ is a unique solution to:*

$$V(x) = \beta(PV)(x) + r(x)$$

for $x \in \mathcal{X}$.

Proof is as follows:[1]

Proof.

$$\begin{aligned}
V(x) &= \mathbb{E}_x \left[\sum_{t=0}^{\infty} \beta^t r(X_t) \right] \\
&= r(x) + \beta \mathbb{E}_x \left[\sum_{t=1}^{\infty} \beta^{t-1} r(X_t) \right] \\
&= r(x) + \beta \mathbb{E}_x \mathbb{E} \left[\sum_{t=1}^{\infty} \beta^{t-1} r(X_t) \middle| X_1 \right] \\
&= r(x) + \beta \mathbb{E}_x \mathbb{E}_{X_1} \left[\sum_{t=1}^{\infty} \beta^{t-1} r(X_t) \right] \\
&= r(x) + \beta \mathbb{E}_x V(X_1) \\
&= r(x) + \beta \sum_{x_1} P_{x x_1} V(X_1) \\
&= r(x) + \beta(PV)(x)
\end{aligned}$$

□

Therefore, we obtain $V = (I - \beta P)^{-1} r$ but this might only help when n is very small and finite since inverse is difficult to compute. This is $O(M^3)$ operation.

- What if inverse doesn't exist or continous time? Death.

[1] Refer Neil Walton's notes for proof of uniqueness.

Standard Optimization

- Optimization Problem $\min_{a \in \mathcal{A}} c(a)$
- $a^* = arg \min_{a \in \mathcal{A}} c(a)$
- When $c(\cdot)$ is convex, then $a^* = \left\{ a : \frac{dc(a)}{da} = 0 \right\}$

Stochastic Optimization

- Consider objective function of form $c(a, W)$ where W is a random variable (typically noise).
- But $c(a, W)$ itself is a random variable.
- So objective function:

$$\min_{a \in \mathcal{A}} E[c(a, W)]$$

  - ⋆ But this is still deterministic where $V(a) = E[c(a, W)]$.

Now, we consider $E[c(S, a, W)]$ where S is the state observed before choosing the action a. This provokes the need for policy.

$\pi : S \to \mathcal{A}$: Decision Rule or ***Policy***

∴ Optimization Problem is now defined as:

$$\min_{\pi} E[c(S, \pi(S), W)] \tag{1}$$

Remark. *Here expectation is over W or S (over the sources of randomness). Note, policy is fixed.*

Another way to view this,

Define $Q(s, a) := E[c(s, a, W)|S = s]$. Then minimize $Q(s, a)$ for every s and store these actions in policy π^*.

$$\min_{a \in \mathcal{A}} Q(s, a) \text{ and } \pi^*(s) = arg \min_{a \in \mathcal{A}} Q(s, a) \tag{2}$$

Problem 1 is functional optimization while 2 is parameter optimization (which are generally easier to solve).