Last Class

Stochastic optimization. Newsvendor.

Info is available. State. Policies

$$\min_{\pi: \mathcal{S} \to \mathcal{A}} \mathbb{E}[C(S, \pi(S), W)]$$

$$\equiv \pi(s) = \arg\min_{a \in \mathcal{A}} \mathbb{E}[C(s, a, W) | S = s], \quad \forall s \in \mathcal{S}$$

Today

- Sufficient / Irrelevant info
- Formulations / Results - Dynamic Programming + MDPs
- Example: Inventory management

IRRELEVANT INFO

$c : \mathcal{S} \times \mathcal{A} \times \mathcal{W} \to \mathbb{R}_{\geq 0}$

$(\Omega, \mathcal{F}, P)$ Prob Space.

(S, Y, W) RVs on Prob Space.

DM obs. (S, Y) and choose A.

$A = \pi(S, Y)$.

When can the DM ignore Y.

Blackwell's Principle of Irrelevant Info (B2/64)

If $Y \perp\!\!\!\perp W \mid S$ $\Rightarrow$ $P(W \mid S, Y) = P(W \mid S)$

Y & W are cond. indep given S

then without loss of opt.

$$\pi^\circ(s) = \arg\min_{a \in A} \mathbb{E}[\, c(s, a, W) \mid S = s \,] \quad \forall s \in \mathcal{S}$$

is opt

---

cf

$$\pi^*(s, y) = \arg\min_{a \in A} \mathbb{E}[\, c(s, a, W) \mid S = s, Y = y \,] \quad \forall s, y.$$

$$\mathbb{E}[\, c \mid s = s, Y = y \,] = \sum_{w} P(W = w \mid S = s, Y = y)\, c(s, a, w)$$

$$= \sum_{\omega} P(W=\omega \mid S=s)\, c(s,a,\omega).$$

$$= \mathbb{E}\left[c(s,a,W) \mid S=s \right]$$

$$\mathbb{E}\left[c(s, \pi^*(s), W) \mid S=s \right] \le \mathbb{E}\left[c(s,a,W) \mid S=s \right] \quad \forall a \in \mathcal{A}$$

for any $\pi : \mathcal{S} \times \mathcal{Y} \to \mathcal{A}$

$$\le \mathbb{E}\left[c(s, \pi(s,y), W) \mid S=s \right] \quad \forall y \in \mathcal{Y}.$$

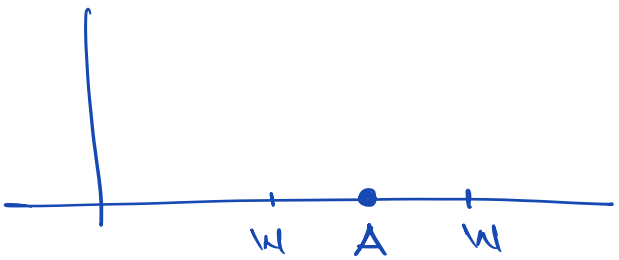
$$= \mathbb{E}\left[c(s, \pi(s,y), W) \mid \begin{matrix} S=s \\ Y=y \end{matrix} \right]$$

MARKOV DECISION PROCESSES

Newsvendor $S_1 = 0$

Purchase A_1 at price per unit p.

Demand W_1



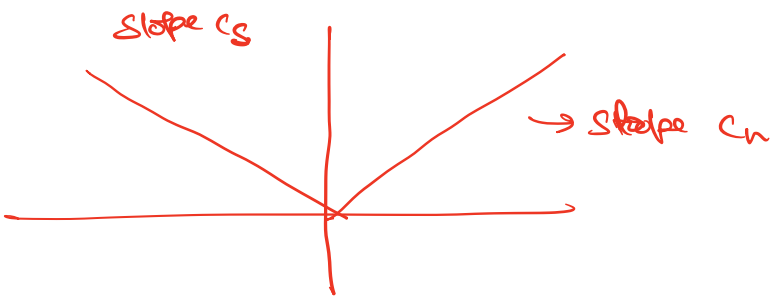
$$S_2 = S_1 + A_1 - W_1$$

Holding

Cost

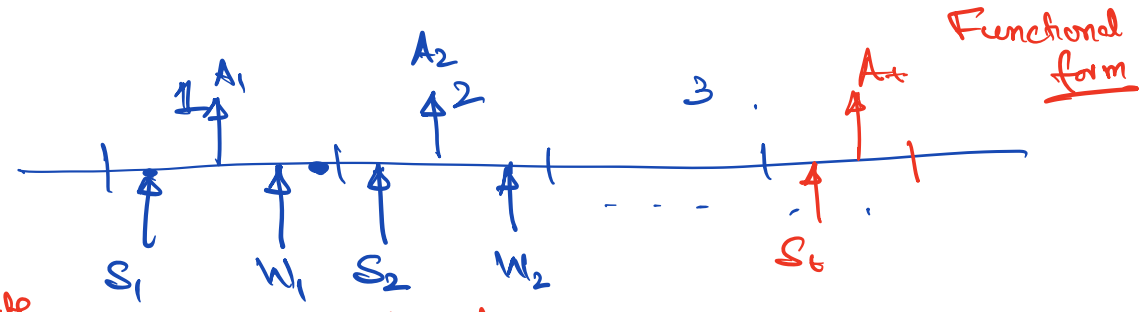
$$c(s) = \begin{cases} c_h s & s \geq 0 \\ -c_s s & s < 0 \end{cases}$$

Shortage



$S_1 = 0$

$$\pi_1(s) = \arg\min_{a \in A} \mathbb{E}_{W_1}\left[c(s_1 + a - W_1) \right]$$



State

System dynamics

$$S_2 = f_1(S_1, A_1, W_1) = S_1 + A_1 - W_1$$

Shocks = Noise / Disturbance / Process Noise

Primitive Random Variable $(S_1, W_1, \ldots W_T)$

Horizon

all defined on a common prob space.

Known Dist

$$A_1 = \pi_1(S_1) \qquad |S|^{|A|}$$

$$A_2 = \pi_2(S_1, S_2, A_1) \qquad (|S^2|^2|A|)^{|A|}$$

$$A_3 = \pi_3(S_1, S_2, S_3, A_1, A_2) \qquad (|S|^3|A^2|)^{|A|}$$

$$\vdots$$

$$A_t = \pi_t(S_{1:t}, A_{1:t-1})$$

$$\pi = (\pi_1, \pi_2, \ldots \pi_T)$$ Policy

Performance.

$$J(\pi) = \mathbb{E}\left[\sum_{t=1}^{T} \left[pA_t + c(S_{t+1})\right]\right]$$

Per-step cost $\rightarrow c(S_t, A_t)$ or $c(S_t, A_t, S_{t+1})$

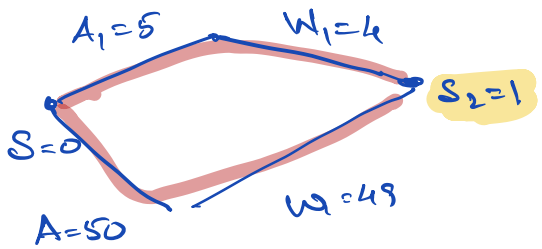
$$\min_{\pi} J(\pi)$$

Assump: All primitive random variables are indep

1 Optimality of Markov policies

WLO,

$$A_t = \pi_t(S_t)$$



Controlled Markov Prop

Matrix form

$$P(S_{t+1} \mid S_{1:t}, A_{1:t}, A_t) = P(S_{t+1} \mid S_t, A_t)$$

$$A_1 = \pi_1(S_1) \quad |S|^{|A|}$$
$$A_2 = \pi_2(S_2) \quad |S|^{|A|}$$
$$A_3 = \pi_3(S_3) \quad |S|^{|A|}$$

$$(\pi_1 \dots \pi_T) : = (|S|^{|A|})^T$$

2. Dynamic Prog. Decomp.

1.

$V_t : \mathcal{S} \to \mathbb{R}$.

$Q_t : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$. $\pi_t : \mathcal{S} \to \mathcal{A}$.

Init $V_{T+1}(s) = 0 \quad \forall s \in \mathcal{S}$.

Recursive for $t \in \{T, T-1, \dots 1\}$

$$Q_t(s,a) = c_t(s,a) + \mathbb{E}\left[V_{t+1}\left(\underbrace{f_t(s,a,W_t)}_{S_{t+1}} \right) \right]$$

$$V_t(s) = \min_{a \in \mathcal{A}} Q_t(s,a)$$

$$\pi_t(s) \in \arg\min_{a \in \mathcal{A}} Q_t(s,a)$$

Optimal Value Fn

$T\left(|\mathcal{S}||\mathcal{A}||\mathcal{W}| + |\mathcal{S}||\mathcal{A}|\right)$ $O(T|\mathcal{S}||\mathcal{A}|)$

$$Q_t(s,a) = c_t(s,a) + \sum_{\omega} P_{W_t}(\omega)\, V_{t+1}(f_t(s,a,\omega))$$

1) How to compute efficiently

2) Can we do something beyond comp