Mod 3. HW 4 Margarita Prikhodko

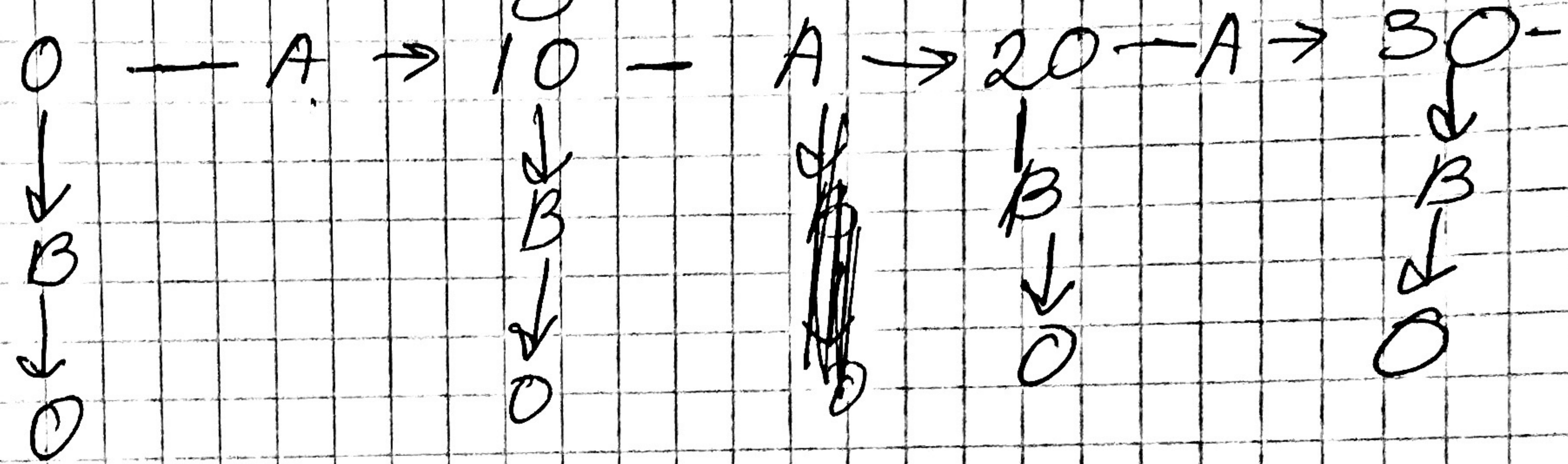
1\. a) Expect. $A = P_A \cdot$ Amount − Cost to Play

Expect $A = 0.1 \cdot 20 - 10 = -8$

Expect. $B = P_B \cdot$ Amount − Cost to Play $=$

$= 0.4 \cdot 30 - 20 = 12 - 20 = -8$

b.) Action Play: MDP diagram



c) All possible policies:

$\beta \in [0, 1]$ where β is probabity to play on A if you have 20\$

You have 20\$ → 2 states

play A or play B

To play Probability for A $= \beta$

To play Probility for B $= 1 - \beta$

For policy π_β

$\pi(a = A \mid s = \$20) = \beta$

$\pi(a = B \mid s = \$20) = 1 - \beta$

Mod3. HW4 Margarita Prikhodko

1. a) Expect A = $P_A \cdot$ Amount − Cost to Play

Expect A = $0.1 \cdot 20 - 10 = -8$

Expect. B = $P_B \cdot$ Amount − Cost to Play =

$= 0.4 \cdot 30 - 20 = 12 - 20 = -8$

b.) Action Play : MDP diagram

$$\begin{tikzcd} 0 \arrow[r, "A"] \arrow[d] & 10 \arrow[r, "A"] \arrow[d] & 20 \arrow[r, "A"] \arrow[d] & 30 \arrow[d] \\ B \arrow[d] & B \arrow[d] & B \arrow[d] & B \arrow[d] \\ 0 & 0 & 0 & 0 \end{tikzcd}$$

c) All possible policies :

$\beta \in [0, 1]$ where β is probabity to play on A if you have 20\$

You have 20\$ → 2 states

play A or play B

To play Probability for A = β

To play Probility for B = $1 - \beta$

For policy π_β

$\pi(a = A \mid s = \$20) = \beta$

$\pi(a = B \mid s = \$20) = 1 - \beta$

In order to compute

$v^{\pi_\beta}(s)$ for all non-terminal States Bellman equation:

$v^{\pi_\beta}(s) = \sum_{a \in Action} \pi_\beta(a|s) \cdot$
$\cdot \left(Reward(s,a) + \sum_{s'} p(s'|s,a) \, v^{\pi_\beta}(s')\right)$

$v^{\pi_\beta}(s)$ is ~~v~~ value of state

$\pi_\beta(a|s)$ is the probability of taking action a in state s under policy π_β

Reward (s, a) is immediate reward of taking action a in state s

$p(s'|s,a)$ is the probability of transitioning to state s' from state s undre action a

1. For state s = \$20 : A state

Reward (s = 20, a = A) = ~~\$~~20 or 0
win losses

$\pi_\beta(a=A \mid s=20) =$ ~~beta~~ $= \beta$

Transition $S=0$ / $S=30 \Rightarrow$ probality $=1$

$$v_A^{\pi_\beta} = \beta \cdot (20 + \text{Reward}(s=0) \cdot 1) + (1-\beta) \cdot (\text{Reward}(S=0) \cdot 1))$$

2. For Action playing: B

$$\pi_\beta (a=B \mid s=\$20) = 1-\beta$$

$$\text{Reward}(S=\$20, a=B) = \underset{(\text{won } B)}{\$30} \text{ or } \underset{(\text{loses } B)}{\$0}$$

$$v_B^{\pi_\beta} = (1-\beta)(30 + \text{Reward}(s=0) \cdot 1) + \beta \cdot (\text{Reward}(S=0) \cdot 1)$$

For state \$0, \$10, \$30 similar

State 0 $\quad v_A^{\pi_\beta} = \beta(0 + \text{Reward} \cdot 1) + (1-\beta) \cdot (\text{Reward} \cdot 1) = 0$

$$v_B^{\pi_\beta} = (1-\beta)(\text{Reward} \cdot 1) + (\cancel{1-}\beta) \cdot \text{Reward} = 0$$

For state \$10

Finally
State $20

$v_{\pi_\beta}(20) = \pi_\beta(A|20) \cdot$
$\cdot (0.1(20-10) + 0.9 \cdot 0) + \pi_\beta(B|20) \cdot$
$\cdot (0.4 \cdot (30-20) + 0.6) =$

$\pi_\beta(A|20) = \beta \qquad \pi_\beta(B|20) = (1-\beta)$

$v_{\pi_\beta} = \beta \cdot 0.1 \cdot \$10 + (1-\beta) \cdot 0.4 \cdot$
$\cdot (30-20) = 0.1\beta + 0.4(1-\beta) =$
$= 0.4 + (-\beta) \cdot 0.3 = 0.4 - 0.3 \cdot \beta$

For state $10

$v_{\pi_\beta}(\$10) = \pi_\beta(A|\$10) \cdot (0.1 \cdot (\$20 - \$10)$
$+ 0.9 \cdot 0) + \pi_\beta(B|\$10) \cdot (0.4 \cdot (30-20) +$
$+ 0.6 \cdot 0) = 0.1 \cdot \beta + 0.4(1-\beta)$

For state $0

$v_A^{\pi_\beta} = \beta \cdot (0 + \text{Reward}.\ 1) + (1-\beta) \cdot 0 =$

$v_B^{\pi_\beta} = 0$ – Terminal state.

d.) Expected value foo machine B

$= p \cdot \$30 - \$20 = \$30 \cdot p - \$20$

$\$30p$

Expected foo value machine A

$E(A) = (1-p) \cdot 20 - 10$

$E(A) = E(B)$

$\$30 \cdot p - \$20 = (1-p) \cdot 20 - 10$

$50p = 20 - 10 + 20 = 30$

$p = \frac{30}{50} = 0.6$

1) Expected value for machine B

$= p \cdot \$30 - \$20 = \$30 \cdot p - \$20$

$\$30p$

Expected for value machine A

$E(A) = (1-p) \cdot 20 - 10$

$E(A) = E(B)$

$\$30 \cdot p - \$20 = (1-p) \cdot 20 - 10$

$50p = 20 - 10 + 20 = 30$

$p = \frac{30}{50} = 0.6$