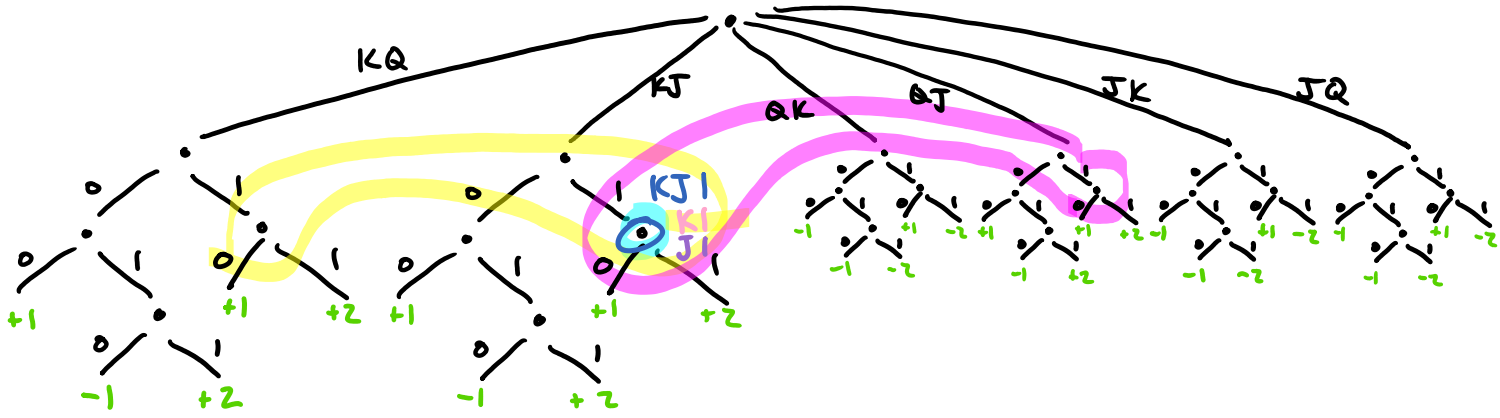
Extensive Form



history : sequence of actions leading to a state

information set : set of states consistent w/ private history

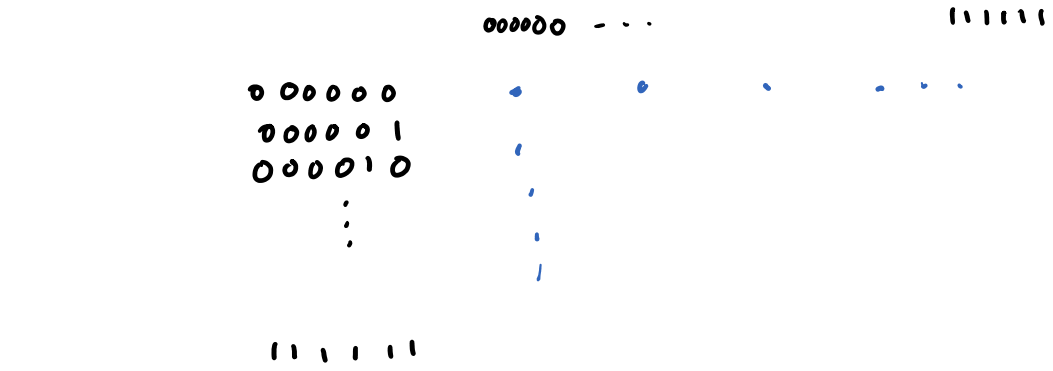
strategy : function from info sets to prob dist over actions possible at info sets

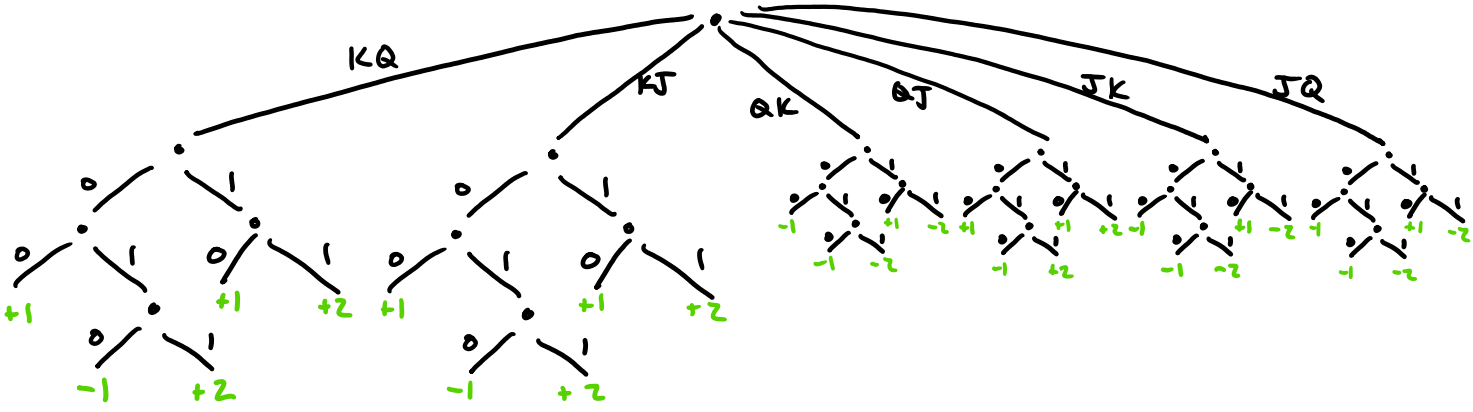
private history: for each p, what p has observed

1) Players reveal strategies
2) Play game according to strategies

P0 : $\overline{K}$ $\overline{Q}$ $\overline{J}$ $\overline{K01}$ $\overline{Q01}$ $\overline{J01}$

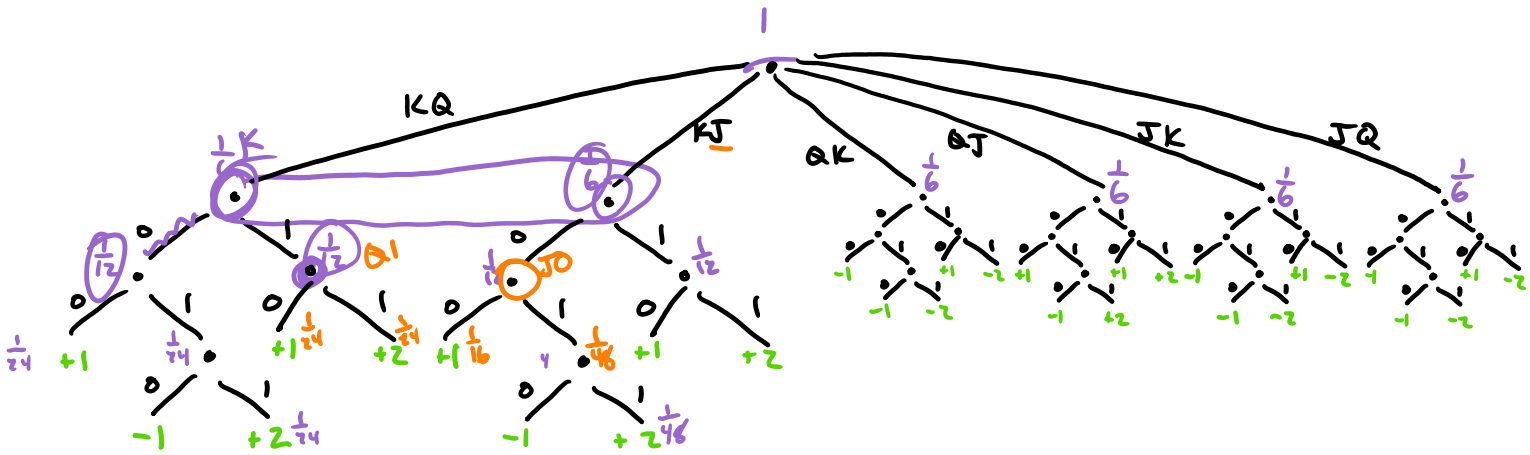
P1 : $\overline{K0}$ $\overline{Q0}$ $\overline{J0}$ $\overline{K1}$ $\overline{Q1}$ $\overline{J1}$



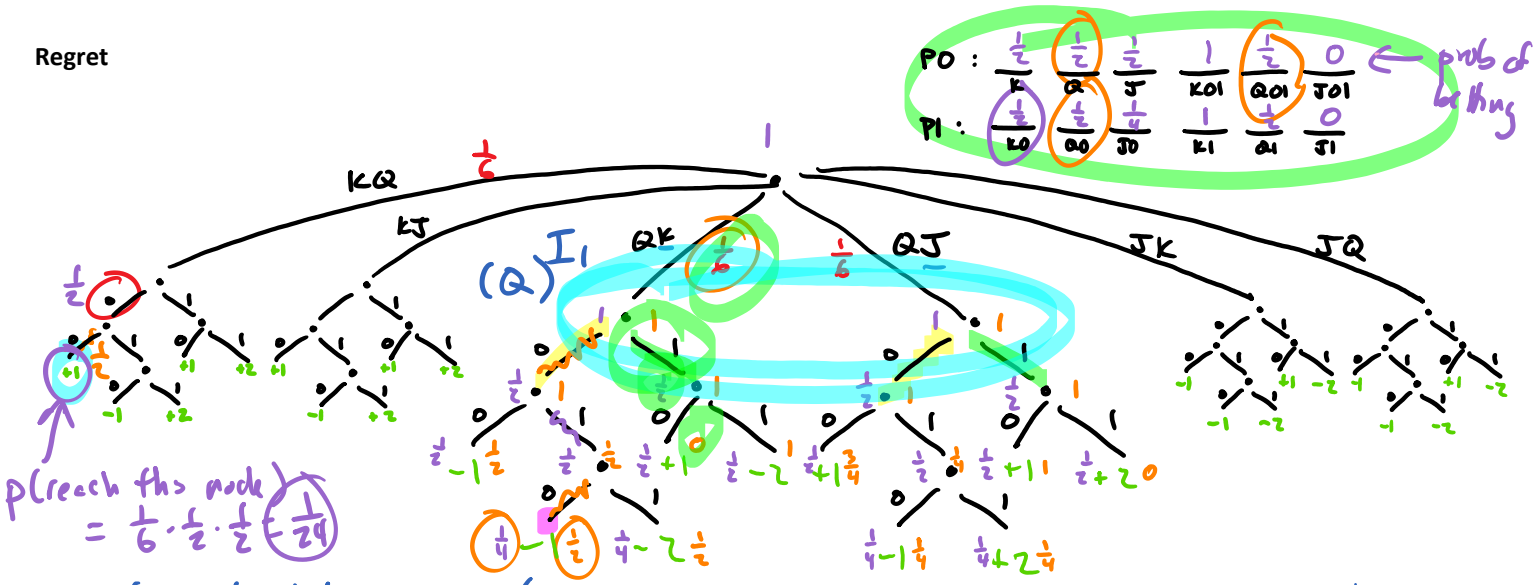


P0 : $\frac{\frac{1}{2}}{K}$ $\frac{\frac{1}{2}}{Q}$ $\frac{\frac{1}{2}}{J}$ $\frac{1}{K01}$ $\frac{\frac{1}{2}}{Q01}$ $\frac{0}{J01}$ ← P(bet)

P1 : $\frac{\frac{1}{2}}{K0}$ $\frac{\frac{1}{2}}{Q0}$ $\frac{\frac{1}{4}}{J0}$ $\frac{1}{K1}$ $\frac{\frac{1}{2}}{Q1}$ $\frac{0}{J1}$


1
KQ
KJ
QK
QJ
JK
JQ

Regret



value of σ contributed by I_1 : $\frac{1}{6}\left(\frac{1}{4}\cdot -1 + \frac{1}{8}\cdot -1 + \frac{1}{8}\cdot -2 + \frac{1}{2}\cdot -2\right) + \frac{1}{6}\left(\frac{3}{8}\cdot 1 + \frac{1}{16}\cdot -1 + \frac{1}{16}\cdot 2 + \frac{1}{2}\cdot 1\right) = \frac{-11}{96}$

value of checking at I_1 :

value of betting at I_1 :

Regret for action = value of action − value of strategy

positive for good actions
negative for bad actions

Counterfactual Regret Minimization

players' strategies $[\pi_0, \pi_1, \pi_c]$

```
CFR(history, σ, π)
    if history is terminal
        return value for current player
    else
        if chance node
            for each chance outcome a w/ prob p_a
                compute p_a · CFR(history+a, σ, [π_0, π_1, π_c · p_a])
            sum those up, return result
        else
            info ← info set for current player
            σ_curr ← strategy proportional to positive observed regret
            compute value of info under each action
            compute value of info under σ_curr
                for each action a with σ_curr(a) > 0
                    compute σ_curr(a) · cfr(hist+a, σ, [π_1, π_0 · σ_curr(a), π_c])
            update regret and average σ
```

CHANCE SAMPLING

OR sample 1 outcome according to prob. dist. of outcomes, let p_a = prob of sampled a

return CFR(....)

-19 (a0) 10 (a1) 9 (a2)

keeps track of cumulative regret for each action

$\sigma_{curr} = 0 \quad \frac{10}{19} \quad \frac{9}{19}$

↳ converges to Nash eq.

↳ weighted by $\pi_1 \cdot \pi_c$

weighted by π_0

```
train(n, players)
    players ← array of players w/ empty stats
    for i = 1 to n
        cfr(root, players, (1.0, 1.0, ...))
    return avg strategy for each player/info set
```

initial state of game

initial probs 1.0 for each player