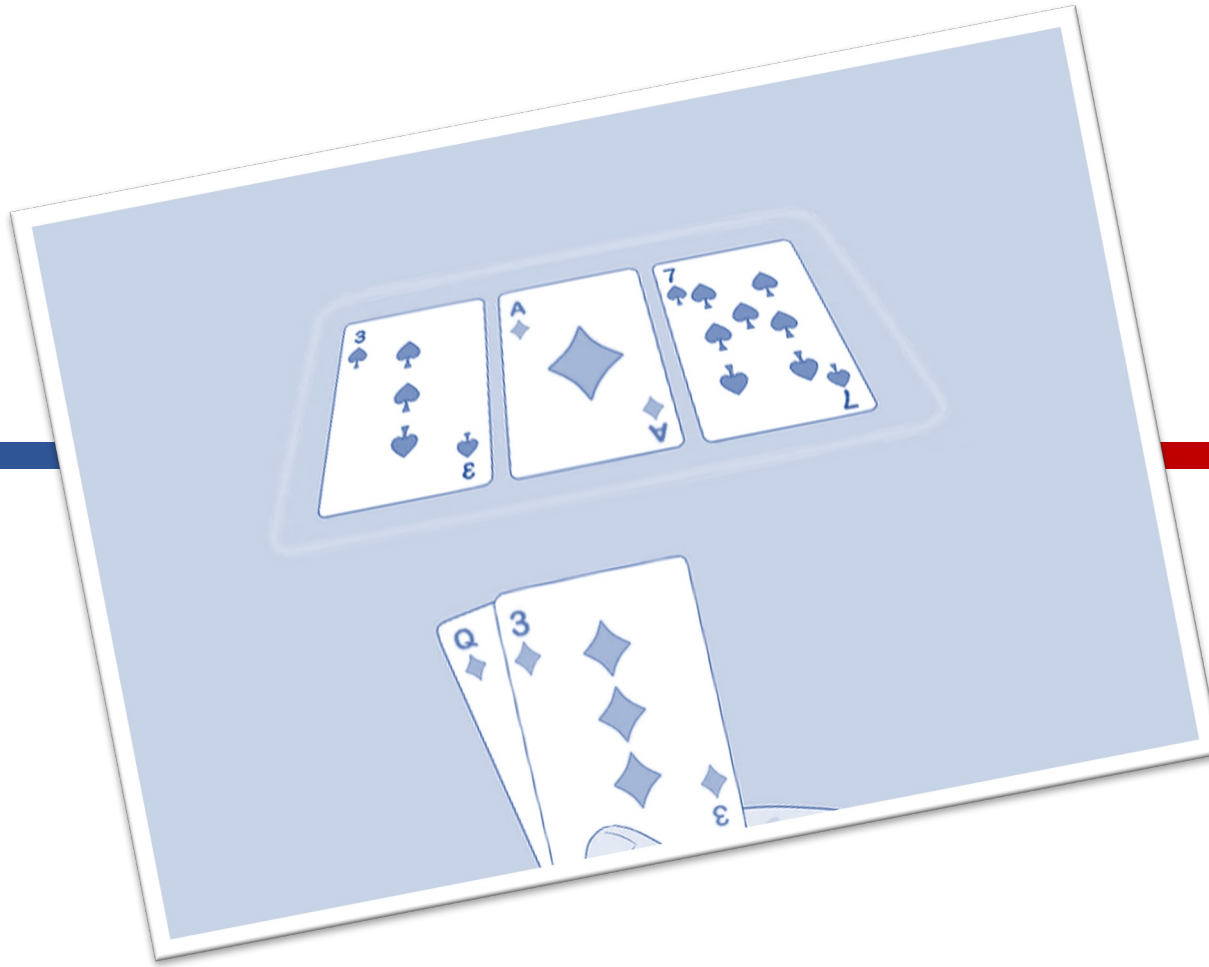
算法博弈论

专题：反事实遗憾最小化



计算机学院
余皓然
2024/5/14

一、棋牌AI设计

棋牌桌游AI设计

扑克类游戏AI设计（德州扑克、斗地主、升级等）

相比于国际象棋、围棋，难点在于对手信息未知（不完美信息博弈）

棋牌桌游AI设计

德州扑克AI设计的研究多次在Science发表，被视作AlphaGo之后证明AI超越人类玩家的又一重要研究

Science

Current Issue First rele

HOME > SCIENCE > VOL. 347, NO. 6218 > HEADS-UP LIMIT HOLD'EM POKER IS SOLVED

RESEARCH ARTICLE

Heads-up limit hold'em poker is solved

MICHAEL BOWLING, NEIL BURCH, MICHAEL JOHANSON, AND OSKARI TAMMELIN Authors Info & Affiliations

SCIENCE · 9 Jan 2015 · Vol 347, Issue 6218 · pp. 145-149 · DOI: 10.1126/science.1259433

Science

Current Issue First releas

HOME > SCIENCE > VOL. 356, NO. 6337 > DEEPSTACK: EXPERT-LEVEL ARTIFICIAL INTELLIGENCE IN HEADS-UP NO-

RESEARCH ARTICLE

DeepStack: Expert-level artificial intelligence in heads-up no-limit poker

MATEJ MORAVČÍK, MARTIN SCHMID, NEIL BURCH, VILIAM LISÝ, DUSTIN MORRILL, MICHAEL BOWLING Authors Info & Affiliations

SCIENCE · 2 Mar 2017 · Vol 356, Issue 6337 · pp. 508-513 · DOI: 10.1126/science.aam6960

Science

Current Issue First release papers Archive About Su

HOME > SCIENCE > VOL. 359, NO. 6374 > SUPERHUMAN AI FOR HEADS-UP NO-LIMIT POKER: LIBRATUS BEATS TOP PROFESSIONALS

RESEARCH ARTICLE

Superhuman AI for heads-up no-limit poker: Libratus beats top professionals

NOAM BROWN AND TUOMAS SANDHOLM Authors Info & Affiliations

SCIENCE · 17 Dec 2017 · Vol 359, Issue 6374 · pp. 418-424 · DOI: 10.1126/science.aao1733

Science

HOME > SCIENCE > VOL. 365, NO. 6456 > SUPERHUMAN AI FOR MULTIPLAYER

RESEARCH ARTICLE

Superhuman AI for multiplayer poker

NOAM BROWN AND TUOMAS SANDHOLM Authors Info & Affiliations

SCIENCE · 11 Jul 2019 · Vol 365, Issue 6456 · pp. 885-890 · DOI: 10.1126/science.aay2400

2019年8月 Science封面

棋牌桌游AI设计

反事实遗憾最小化（counterfactual regret minimization）是解决此类不完美信息博弈的基础算法

Regret minimization in games with incomplete information

M Zinkevich, M Johanson… - Advances in neural …, 2007 - proceedings.neurips.cc

… based on **regret minimization**. In particular, we introduce the notion of **counterfactual regret**,

… We show how **minimizing counterfactual regret minimizes** overall **regret**, and therefore in self…

☆ Save 99 Cite Cited by 943 Related articles All 20 versions »

最早的研究工作发表在 NIPS 2007

Counterfactual Regret Minimization In 2006, the Annual Computer Poker Competition was started (*25*). The competition drove advancements in solving larger and larger games, with multiple techniques and refinements being proposed in the years that followed (*33, 34*). One of the techniques to emerge, and currently the most widely adopted in the competition, is counterfactual regret minimization (CFR) (*35*). CFR is an iterative method for approximating a Nash equilibrium of an extensive-form game through the process of repeated self-play between two regret-minimizing algorithms (*19, 36*). Regret is the loss in utility an algorithm suffers for not having selected the single best deterministic strategy, which can only be known in hindsight. A regret-minimizing algorithm is one that guarantees that its regret grows sub-linearly over time, and so eventually achieves the same utility as the best deterministic strategy. The key insight of CFR is that instead of storing and minimizing regret for the exponential number of deterministic strategies, CFR stores and minimizes a modified regret for each information set and subsequent action, which can be used to form an upper bound on the regret for any deterministic strategy. An approximate Nash equilibrium is retrieved by averaging each player's strategies over all of the iterations, and the approximation improves as the number of iterations increases. The memory needed for the algorithm is linear in the number of information sets, rather than quadratic, which is the case for efficient LP methods (*37*). Because solving large games is usually bounded by available memory, CFR has resulted in an increase in the size of solved games similar to that of Koller et al.'s advance. Since its introduction in 2007, CFR has been used to solve increasingly complex simplifications of HULHE, reaching as many as 3.8×10^{10} information sets in 2012 (*38*).

Science 2015论文对算法的介绍

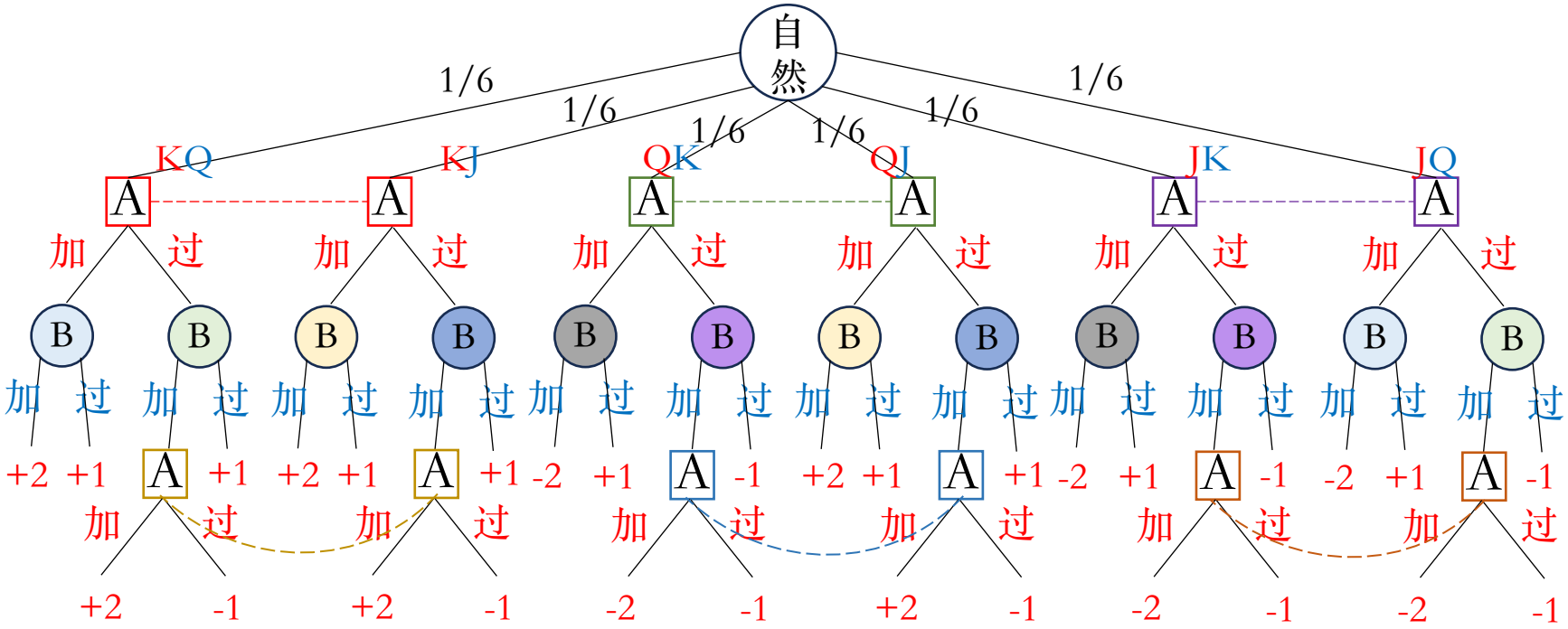
双人库恩扑克

库恩扑克是德州扑克的极简版

- 游戏规则

玩家A	玩家B	玩家A	结果
过牌	过牌	\	牌值大的玩家 +1
加注	加注	\	牌值大的玩家 +2
过牌	加注	过牌	玩家B +1
过牌	加注	加注	牌值大的玩家 +2
加注	过牌	\	玩家A +1

双人库恩扑克

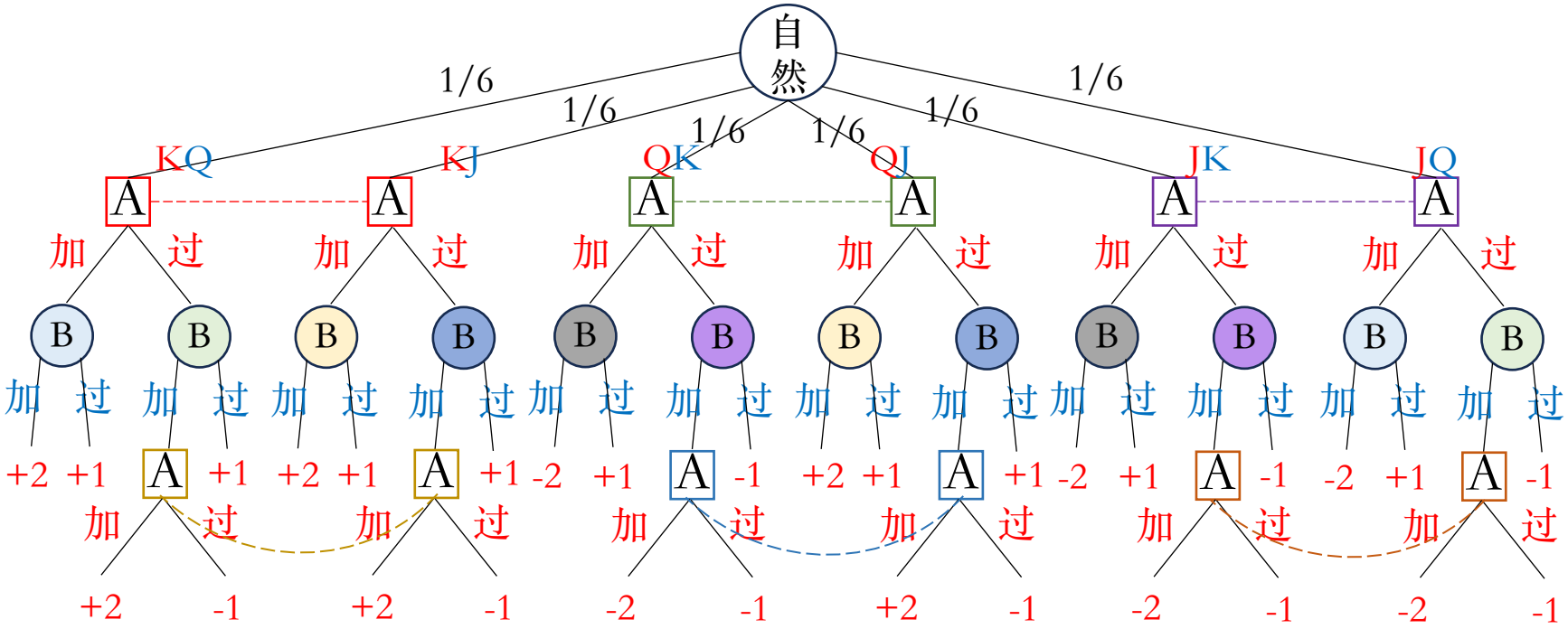


- 游戏规则

玩家A	玩家B	玩家A	结果
过牌	过牌	\	牌值大的玩家 +1
加注	加注	\	牌值大的玩家 +2
过牌	加注	过牌	玩家B +1
过牌	加注	加注	牌值大的玩家 +2
加注	过牌	\	玩家A +1

图中仅显示玩家A的收益，玩家B收益是相应负值
红色对应玩家A的牌、行为及收益；蓝色对应玩家B的牌、行为

双人库恩扑克



- 游戏规则

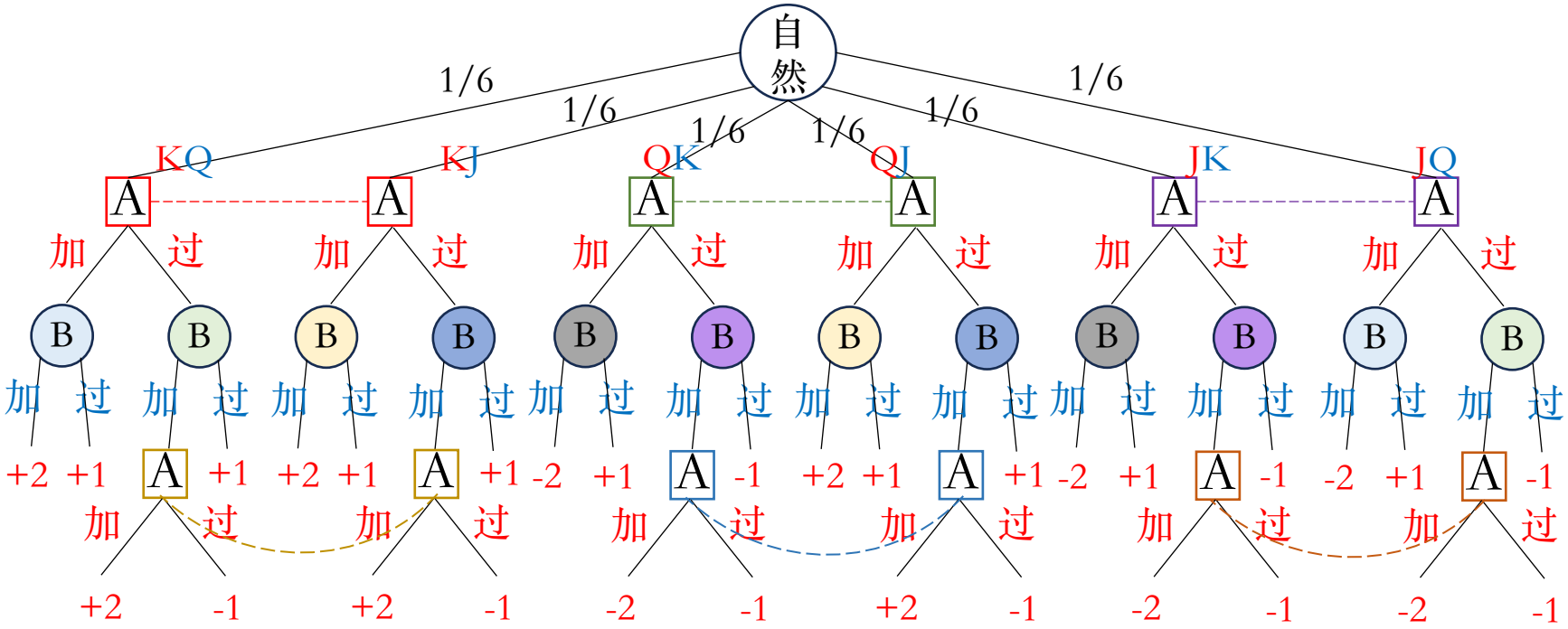
玩家A	玩家B	玩家A	结果
过牌	过牌	\	牌值大的玩家 +1
加注	加注	\	牌值大的玩家 +2
过牌	加注	过牌	玩家B +1
过牌	加注	加注	牌值大的玩家 +2
加注	过牌	\	玩家A +1

图中仅显示玩家A的收益，玩家B收益是相应负值
红色对应玩家A的牌、行为及收益；蓝色对应玩家B的牌、行为

Prof. Bryce Wiedenbeck

上图及本讲大部分内容修改自Bryce Wiedenbeck教授在Algorithmic Game Theory课程中的Fictitious Play and Regret Matching及Counterfactual Regret Minimization两讲，特此致谢！

双人库恩扑克



- 游戏规则

玩家A	玩家B	玩家A	结果
过牌	过牌	\	牌值大的玩家 +1
加注	加注	\	牌值大的玩家 +2
过牌	加注	过牌	玩家B +1
过牌	加注	加注	牌值大的玩家 +2
加注	过牌	\	玩家A +1

图中仅显示玩家A的收益，玩家B收益是相应负值
红色对应玩家A的牌、行为及收益；蓝色对应玩家B的牌、行为

本质上是不完美信息 双人 零和 拓展式博弈

如何求解玩家A和玩家B的均衡策略：反事实遗憾最小化

二、预备知识

预备知识一：混合均衡策略

考虑右图所示矩阵博弈，玩家A应采用什么策略确保不被克制

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	1, -1
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

预备知识一：混合均衡策略

考虑右图所示矩阵博弈，玩家A应采用什么策略确保不被克制

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	1, -1
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

玩家A应采用混合策略，记为$\sigma_A = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

该矩阵博弈的混合策略均衡为$\sigma_A^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \sigma_B^* = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

玩家A的期望收益为$u_A(\sigma_A^*, \sigma_B^*) = 0$

预备知识一：混合均衡策略

考虑右图所示矩阵博弈，玩家A应采用什么策略确保不被克制

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

该矩阵博弈的混合策略均衡为$\sigma_A^* = \left(\frac{1}{5}, \frac{7}{15}, \frac{1}{3}\right), \sigma_B^* = \left(\frac{1}{3}, \frac{7}{15}, \frac{1}{5}\right)$

玩家A的期望收益为$u_A(\sigma_A^*, \sigma_B^*) = \frac{2}{15}$

如果玩家A改为其它策略，可能被玩家B克制

例如，$\sigma_A = \left(\frac{1}{2}, \frac{1}{2}, 0\right)$，玩家B选择$\sigma_B = (0,1,0)$时，玩家A收益为$-\frac{1}{2}$

如何求该矩阵博弈的混合策略均衡？

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1)

玩家B各行为累加遗憾值 初始化为 (1,1,1)

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

在regret matching的第一个迭代回合，令玩家A选择策略σ_1、玩家B选择策略σ_2

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

在regret matching的第一个迭代回合，令玩家A选择策略σ_1、玩家B选择策略σ_2

分析玩家A应该如何更新累加遗憾值、从而更新σ_1：

$$u_A(\sigma_1,\sigma_2) = \frac{1}{9}(0-1+3+1+0-1-1+1+0) = \frac{2}{9}$$

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

在regret matching的第一个迭代回合，令玩家A选择策略σ_1、玩家B选择策略σ_2

分析玩家A应该如何更新累加遗憾值、从而更新σ_1：

$$u_A(\sigma_1,\sigma_2) = \frac{1}{9}(0-1+3+1+0-1-1+1+0) = \frac{2}{9}$$

若玩家A改σ_1为石头：$u_A(\text{石头},\sigma_2) = \frac{2}{3}$　多赚$\frac{2}{3}-\frac{2}{9}=\frac{4}{9}$

若玩家A改σ_1为布：$u_A(\text{布},\sigma_2) = 0$　无法多赚

若玩家A改σ_1为剪刀：$u_A(\text{剪刀},\sigma_2) = 0$　无法多赚

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

在regret matching的第一个迭代回合，令玩家A选择策略σ_1、玩家B选择策略σ_2

分析玩家B应该如何更新累加遗憾值、从而更新σ_2：

$$u_B(\sigma_1,\sigma_2) = \frac{1}{9}(0+1-3-1+0+1+1-1+0) = -\frac{2}{9}$$

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

在regret matching的第一个迭代回合，令玩家A选择策略σ_1、玩家B选择策略σ_2

分析玩家B应该如何更新累加遗憾值、从而更新σ_2：

$$u_B(\sigma_1,\sigma_2) = \frac{1}{9}(0+1-3-1+0+1+1-1+0) = -\frac{2}{9}$$

若玩家B改σ_2为石头：$u_B(\sigma_1,\text{石头}) = 0$　　赚$-\left(-\frac{2}{9}\right) = \frac{2}{9}$

若玩家B改σ_2为布：$u_B(\sigma_1,\text{布}) = 0$　　赚$-\left(-\frac{2}{9}\right) = \frac{2}{9}$

若玩家B改σ_2为剪刀：$u_B(\sigma_1,\text{剪刀}) = -\frac{2}{3}$　　无法多赚

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

玩家A各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_1 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

玩家B各行为累加遗憾值 初始化为 (1,1,1) 归一化$\sigma_2 = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

regret matching的第一个迭代回合结束：

若玩家A改σ_1为石头，赚$\frac{4}{9}$

若玩家A改σ_1为布，无法多赚

若玩家A改σ_1为剪刀，无法多赚

遗憾更新 $\left(1+\frac{4}{9},1,1\right)$

σ_1更新为 $\left(\frac{13}{31},\frac{9}{31},\frac{9}{31}\right)$

若玩家B改σ_2为石头，赚$\frac{2}{9}$

若玩家B改σ_2为布，赚$\frac{2}{9}$

若玩家B改σ_2为剪刀，无法多赚

遗憾更新 $\left(1+\frac{2}{9},1+\frac{2}{9},1\right)$

σ_2更新为 $\left(\frac{11}{31},\frac{11}{31},\frac{9}{31}\right)$

预备知识二：遗憾值匹配

遗憾值匹配（regret matching）是求解矩阵博弈混合策略均衡的方法之一（不能确保对所有矩阵博弈都收敛到均衡）

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

玩家A各行为累加遗憾值 为$\left(1+\frac{4}{9},1,1\right)$　　归一化$\sigma_1=\left(\frac{13}{31},\frac{9}{31},\frac{9}{31}\right)$

玩家B各行为累加遗憾值 为$\left(1+\frac{2}{9},1+\frac{2}{9},1\right)$ 归一化$\sigma_2=\left(\frac{11}{31},\frac{11}{31},\frac{9}{31}\right)$

根据σ_1,σ_2执行regret matching的第二个迭代回合...

通过不断迭代，预期(σ_1,σ_2)可收敛到混合策略均衡$(\sigma_A^*,\sigma_B^*)=\left(\left(\frac{1}{5},\frac{7}{15},\frac{1}{3}\right),\left(\frac{1}{3},\frac{7}{15},\frac{1}{5}\right)\right)$

预备知识

- 混合均衡策略
- 遗憾值匹配

A \ B	石头	布	剪刀
石头	0, 0	-1, 1	3, -3
布	1, -1	0, 0	-1, 1
剪刀	-1, 1	1, -1	0, 0

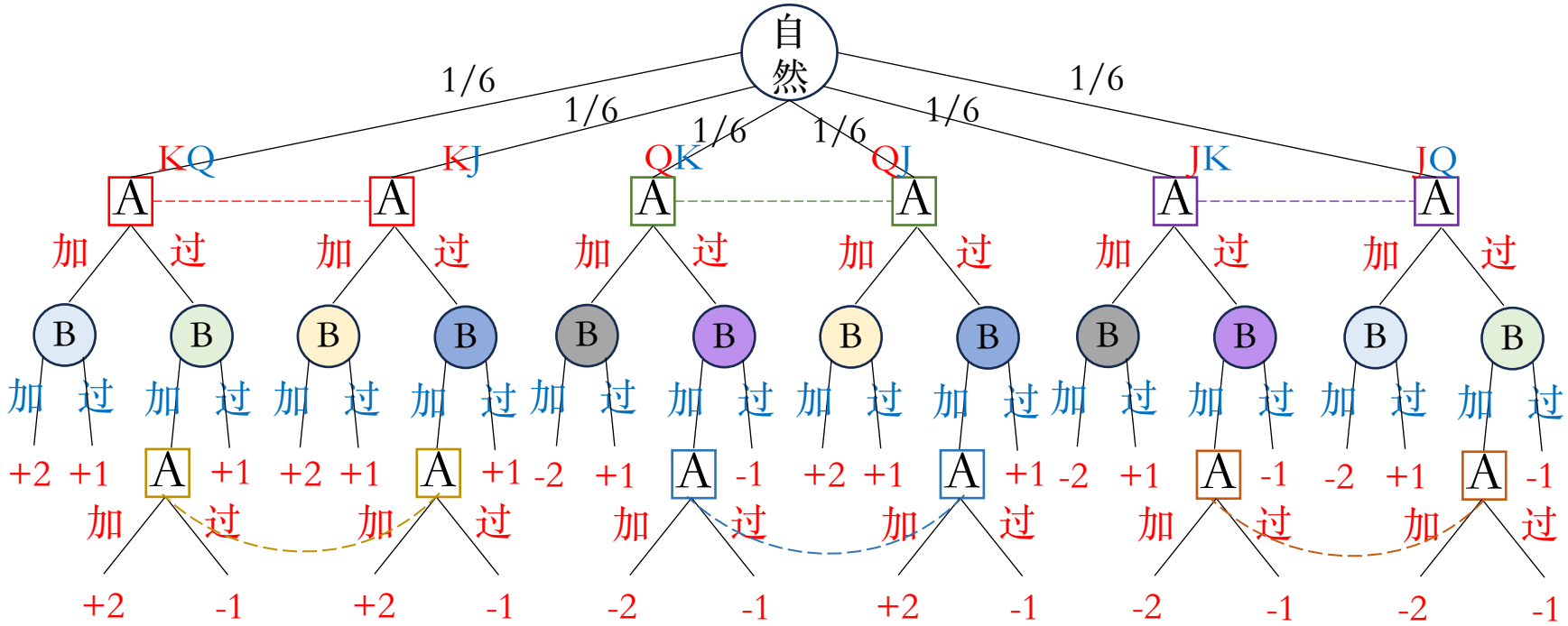
对双人库恩扑克（不完美信息 双人 零和 拓展式博弈）是否也可采用遗憾值匹配求解混合均衡策略?

反事实遗憾最小化算法

三、库恩扑克的混合均衡策略

双人库恩扑克

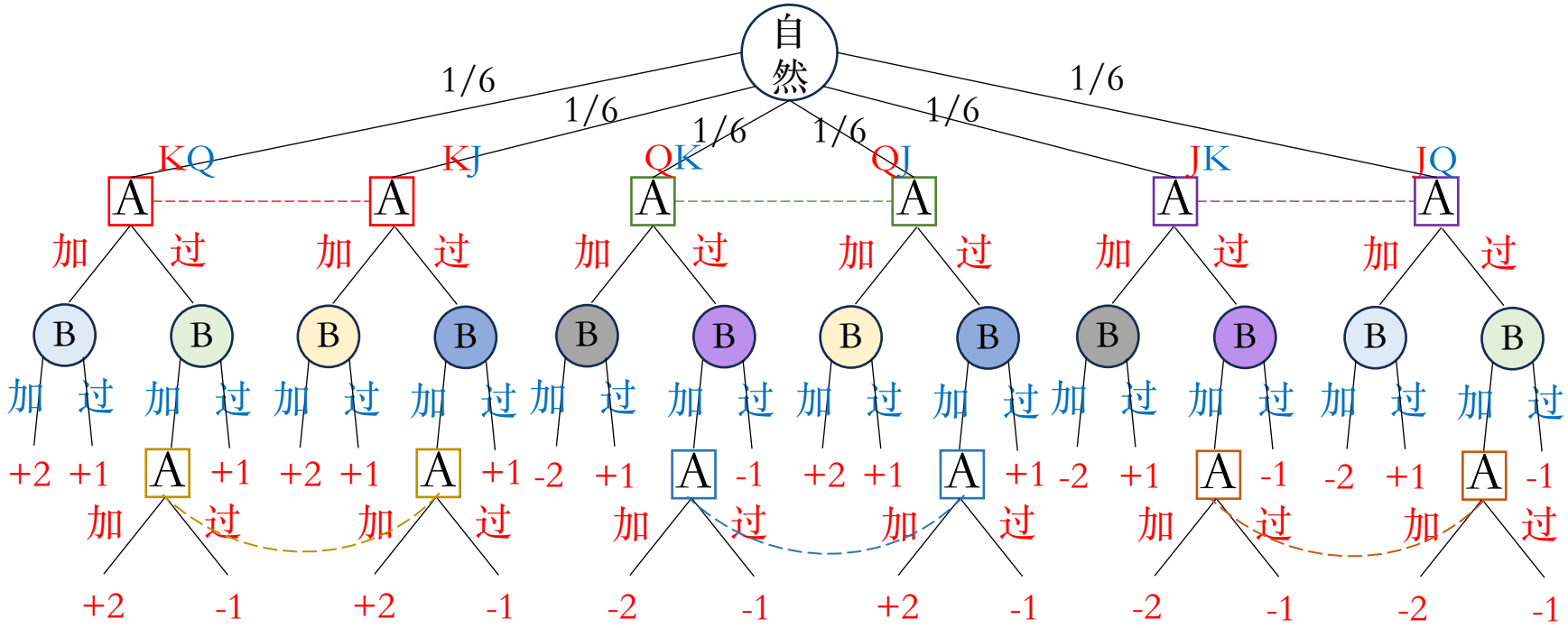
步骤一 定义混合策略



双人库恩扑克

步骤一 定义混合策略

为每个信息集定义对应的策略



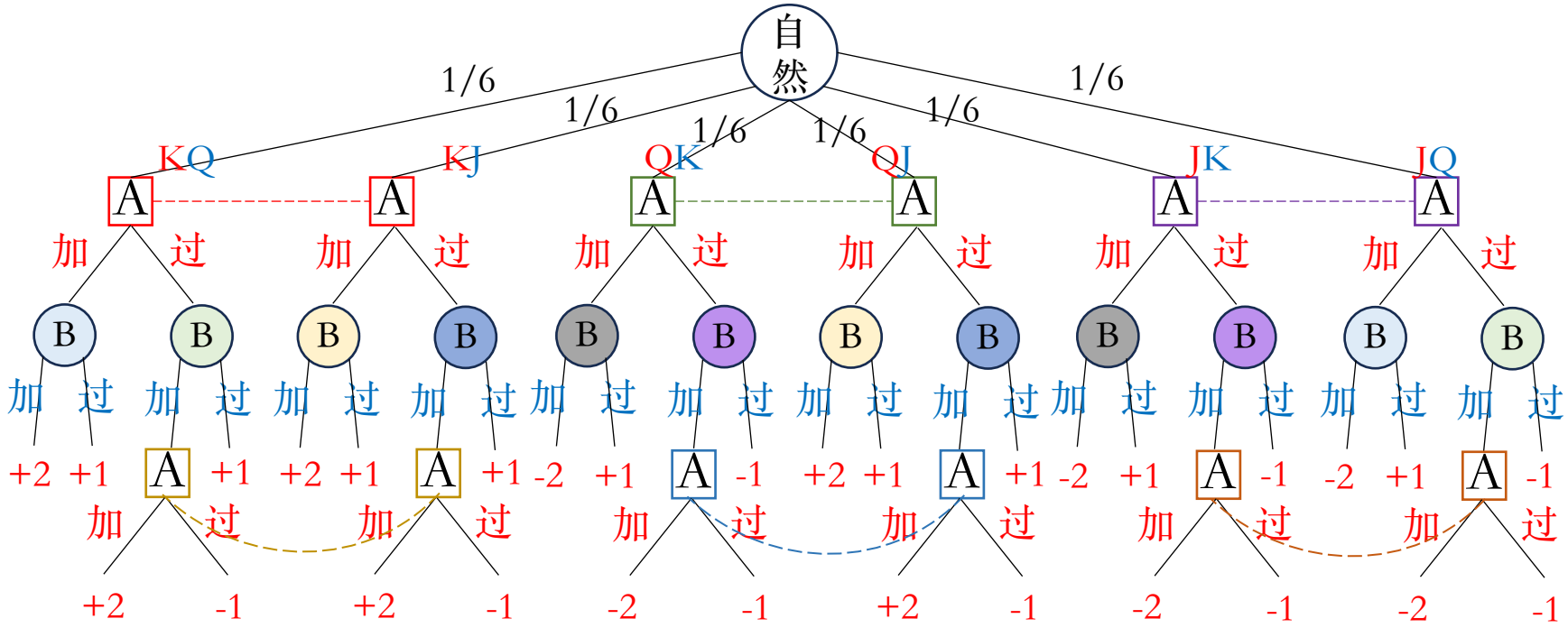
玩家A

	加注	过牌
K		
Q		
J		
K/过/加		
Q/过/加		
J/过/加		

玩家B

	加注	过牌
K/加		
K/过		
Q/加		
Q/过		
J/加		
J/过		

双人库恩扑克



步骤一 定义混合策略

为每个信息集定义对应的策略

玩家A　　随机初始化

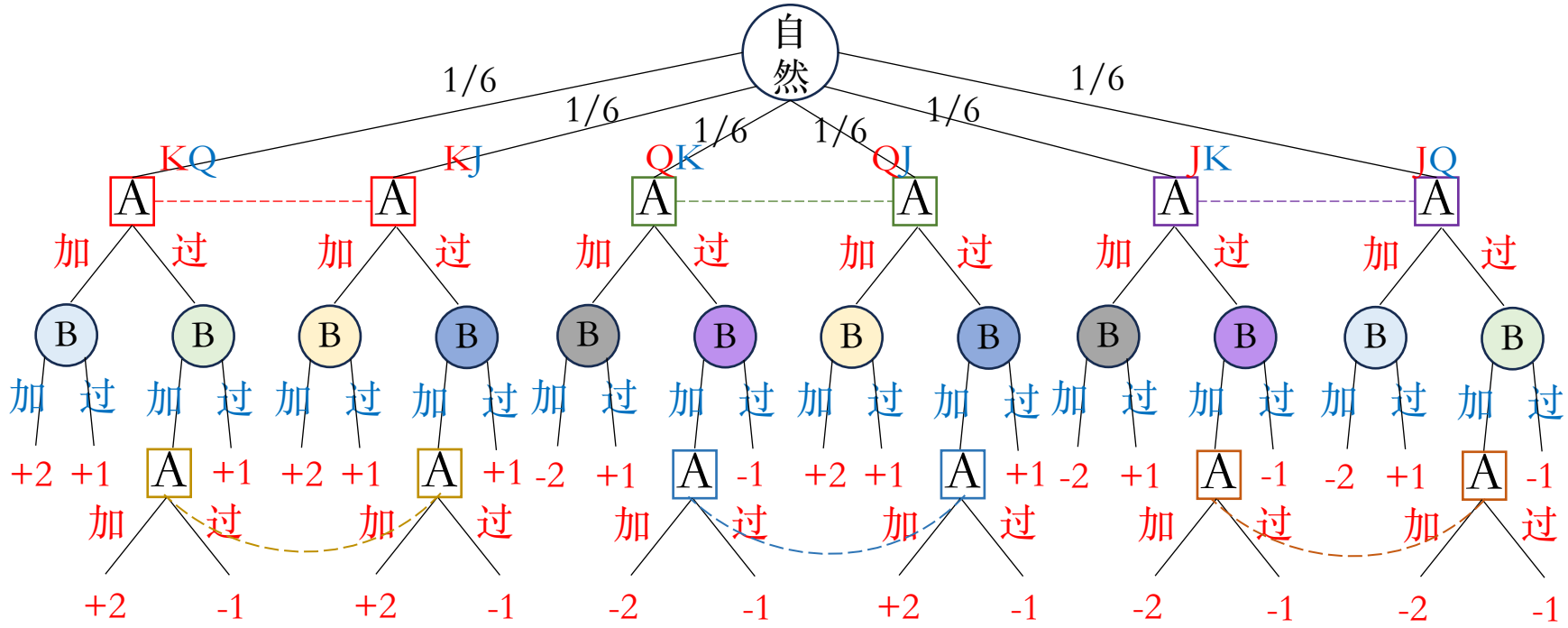
	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

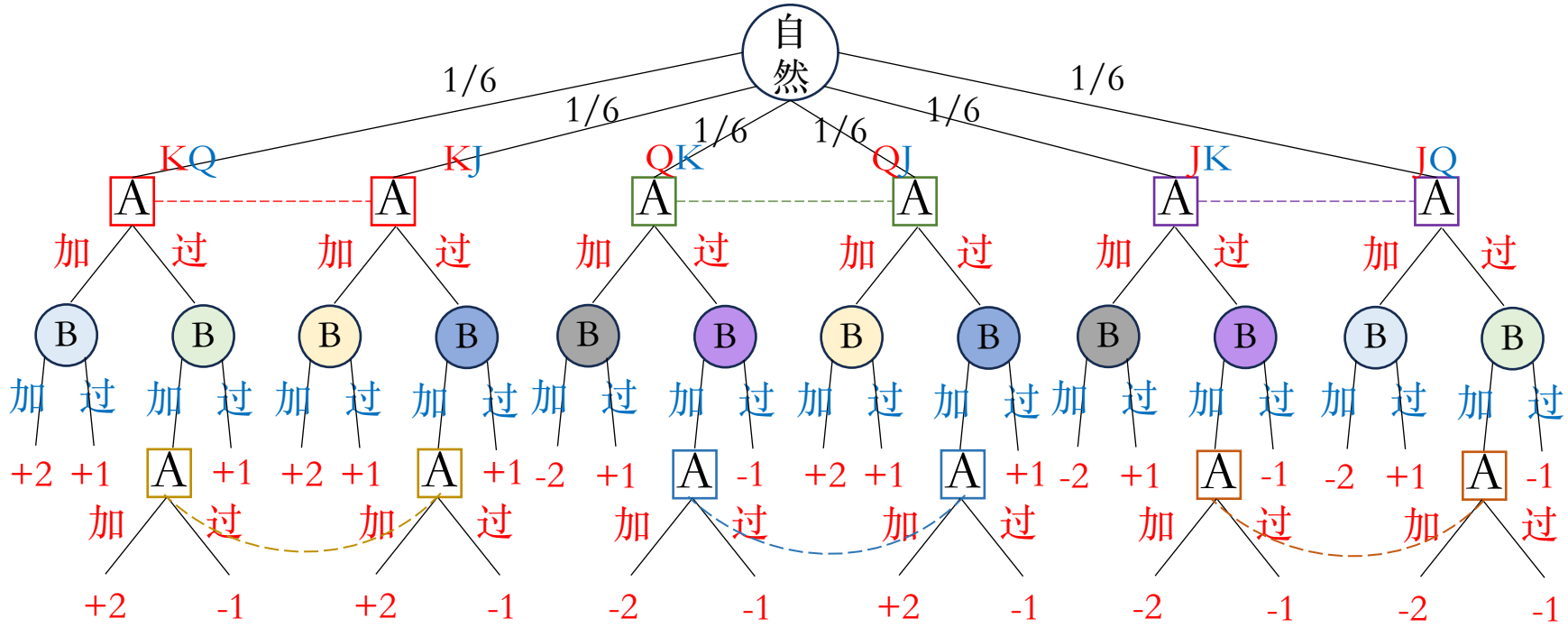
KQ？

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2

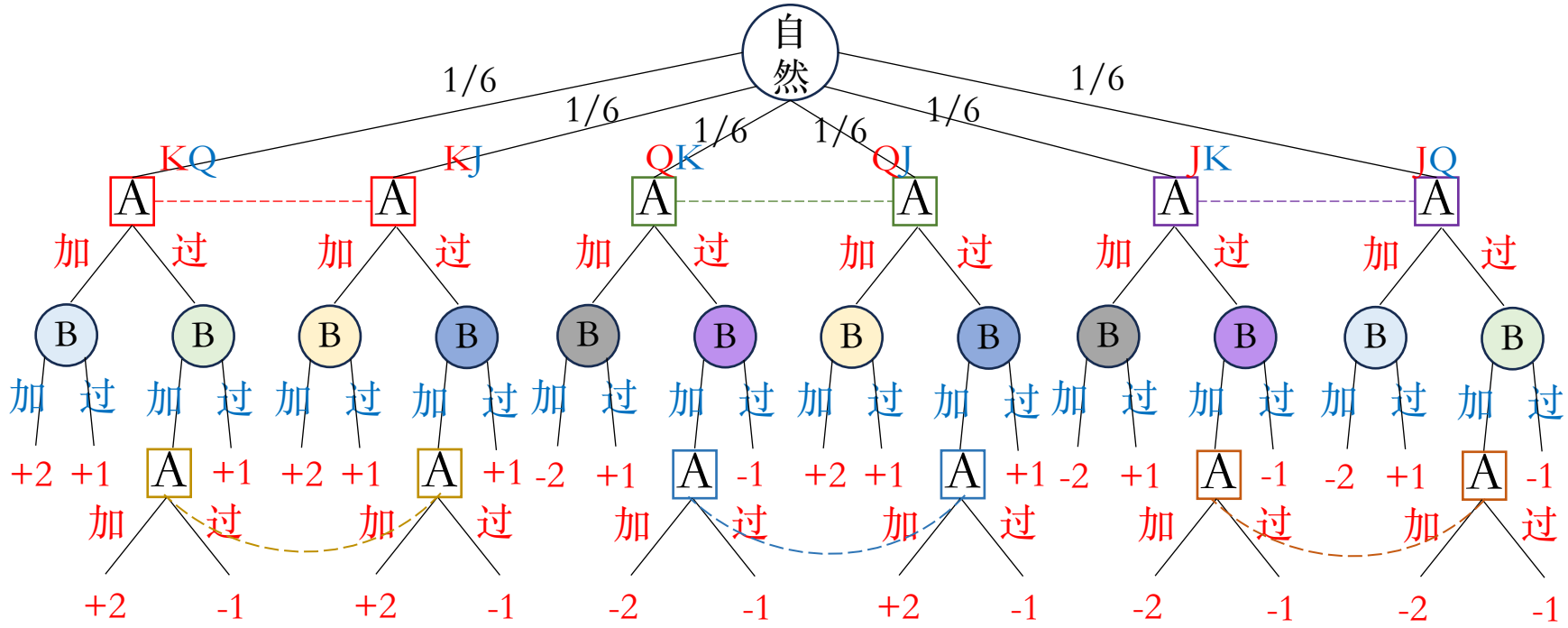
$$p(KQ|K)=\frac{p(KQ)}{p(K)}=\frac{p(KQ)}{p(KQ)+p(KJ)}=\frac{1/6}{1/3}=\frac{1}{2}$$

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

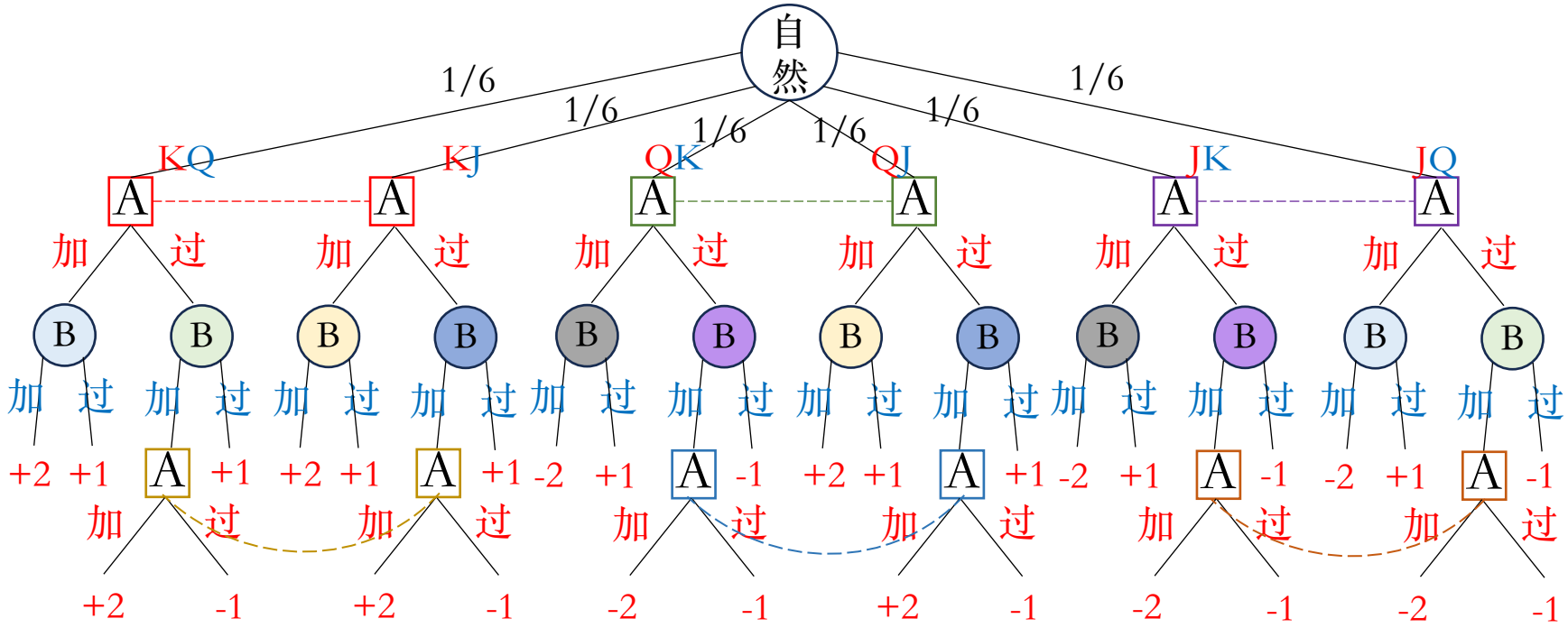
JK 1/2　JQ 1/2

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3

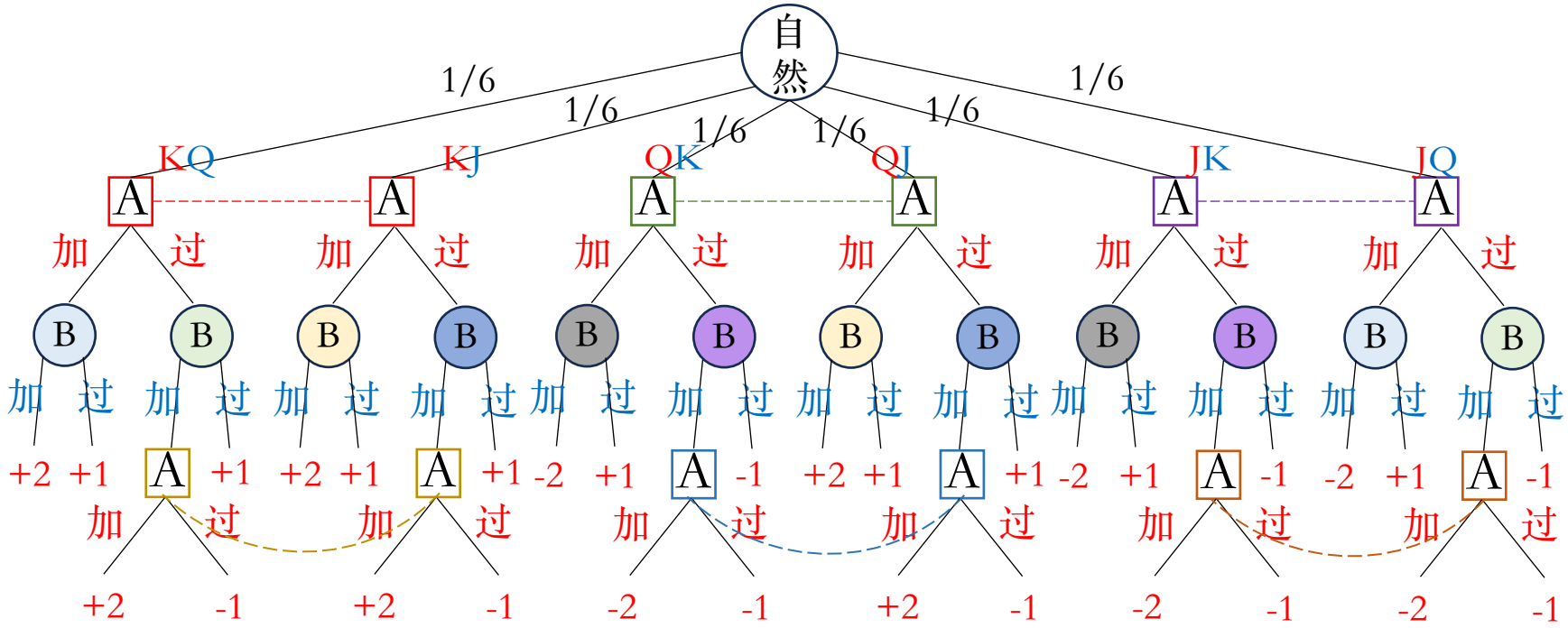
$$p(KQ过加|K过加) = \frac{p(KQ过加)}{p(K过加)} = \frac{p(KQ过加)}{p(KQ过加)+p(KJ过加)} = \frac{\frac{1}{6}*\frac{1}{3}*\frac{2}{3}}{\frac{1}{6}*\frac{1}{3}*\frac{2}{3}+\frac{1}{6}*\frac{1}{3}*\frac{1}{3}}$$

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3

QK/过/加 3/4

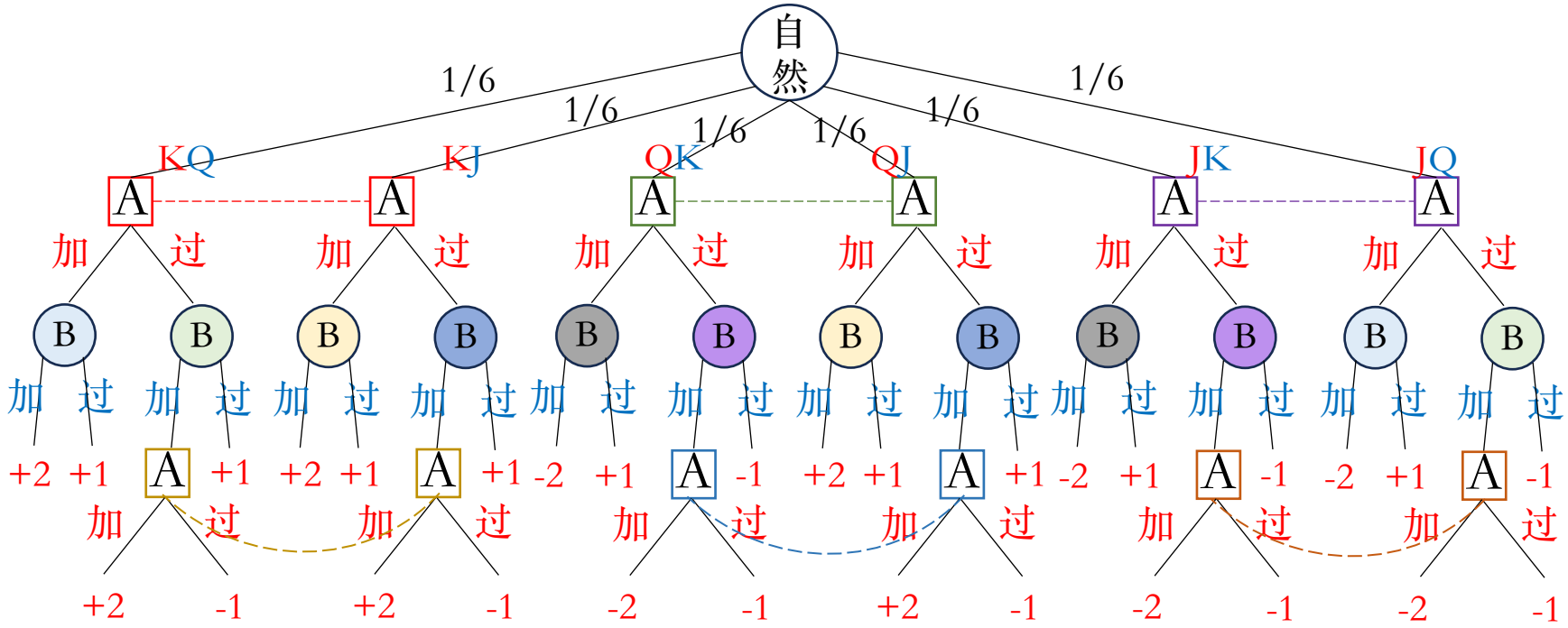
$$p(QK过加|Q过加) = \frac{p(QK过加)}{p(Q过加)} = \frac{p(QK过加)}{p(QK过加)+p(QJ过加)} = \frac{\frac{1}{6}*\frac{1}{2}*1}{\frac{1}{6}*\frac{1}{2}*1+\frac{1}{6}*\frac{1}{2}*\frac{1}{3}}$$

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤二 分析实际所处状态



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

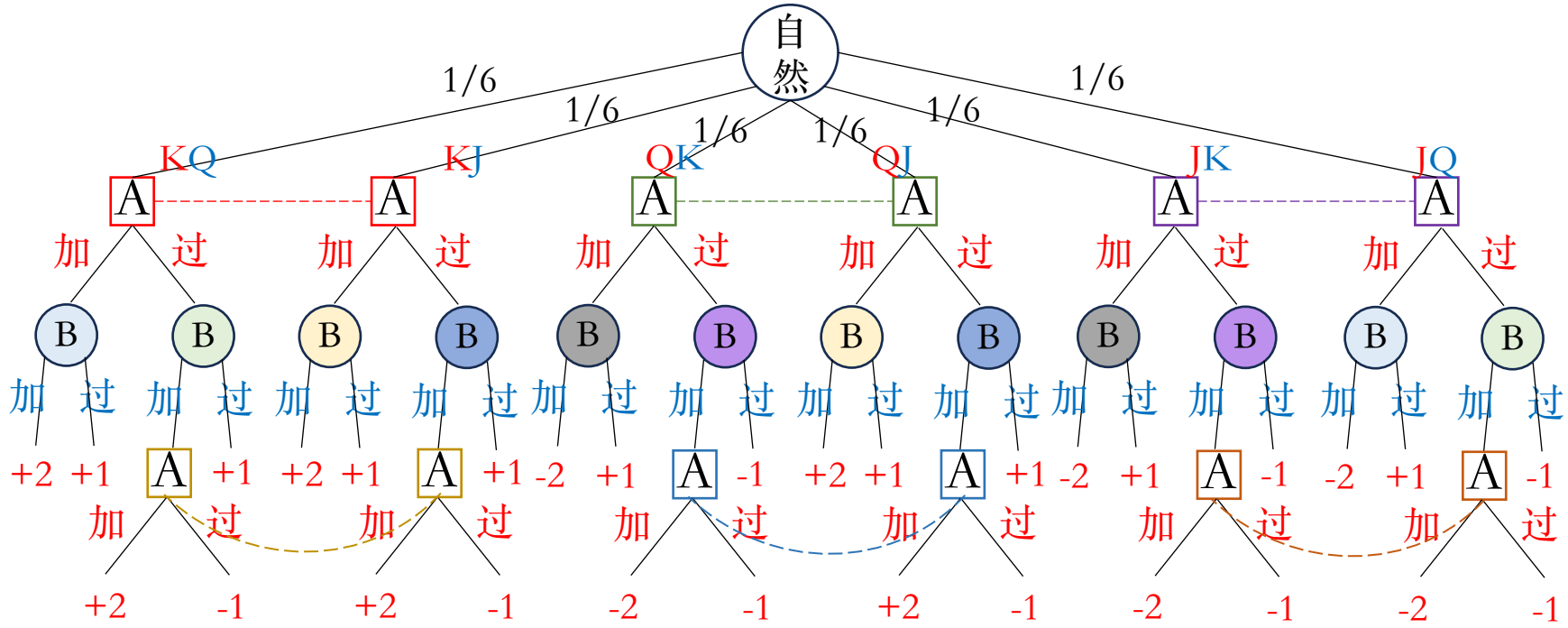
JK/过/加 3/5　JQ/过/加 2/5

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

计算行为收益

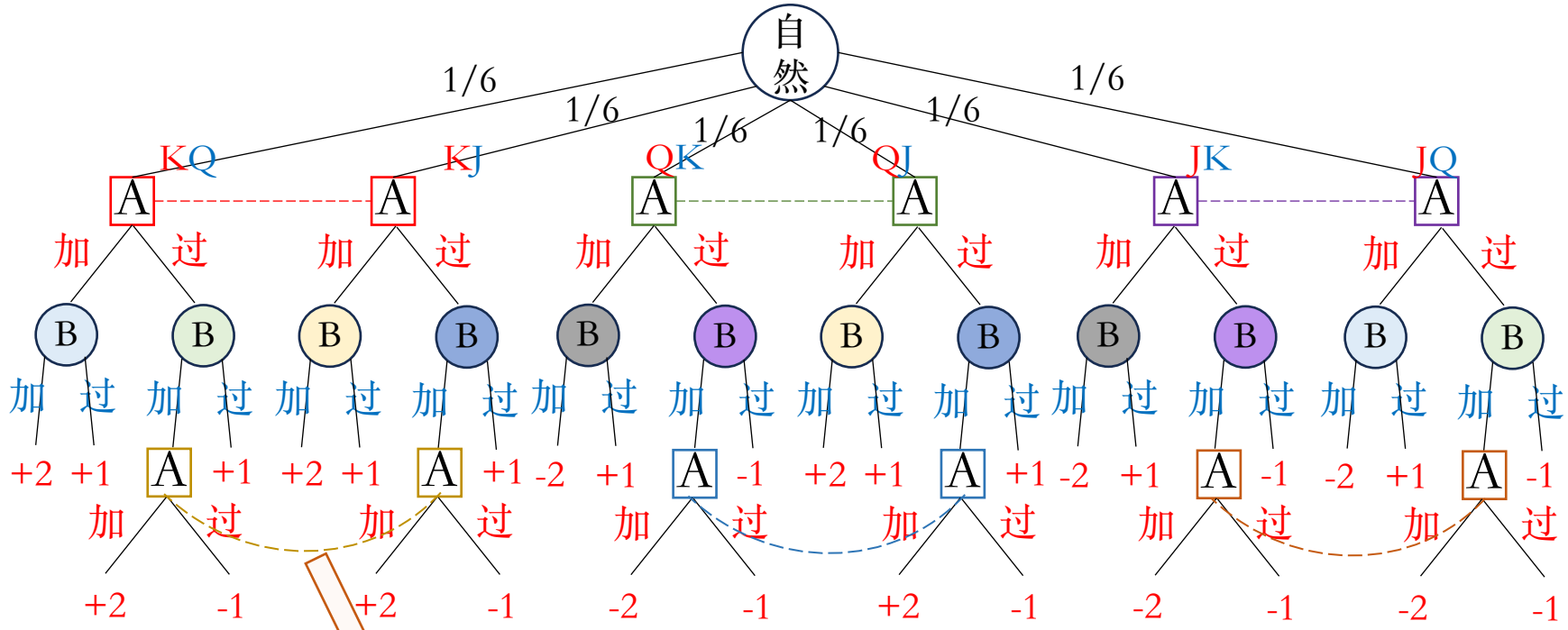
加注	过牌

玩家B　　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

计算行为收益

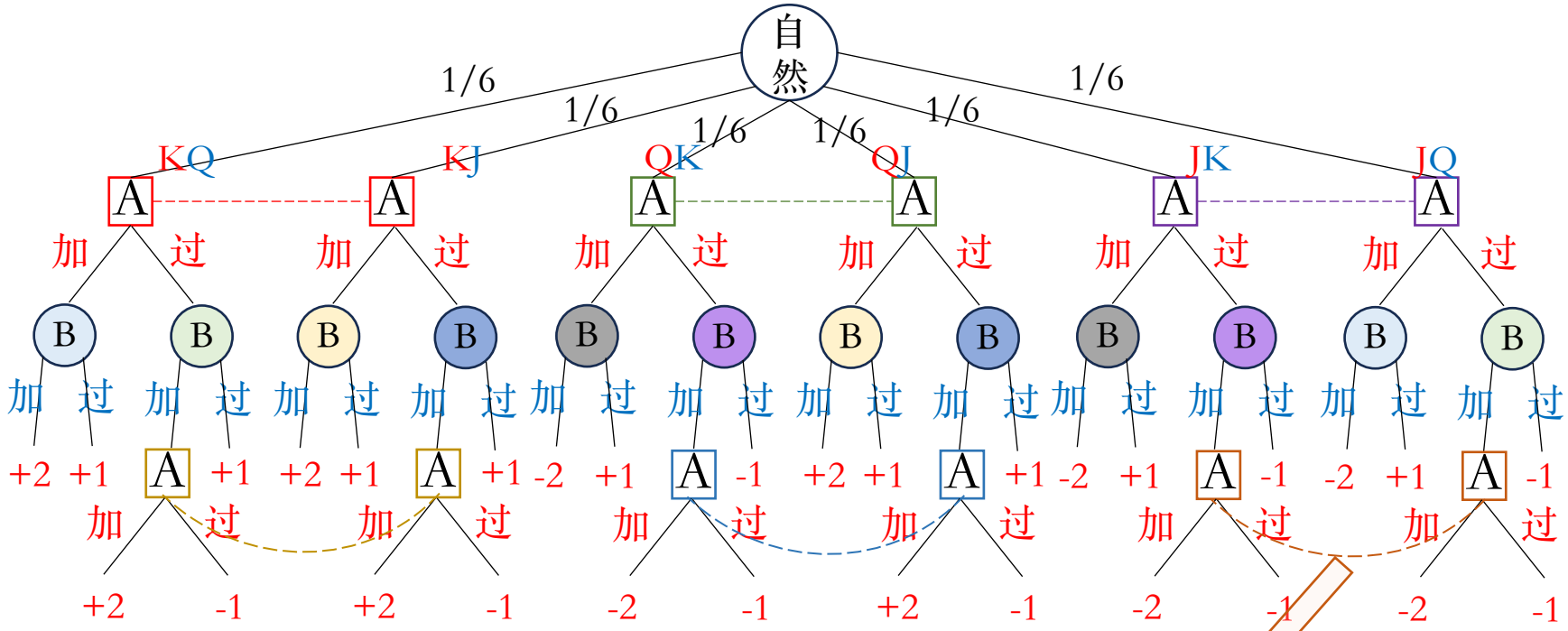
加注	过牌
2	-1

玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

计算行为收益

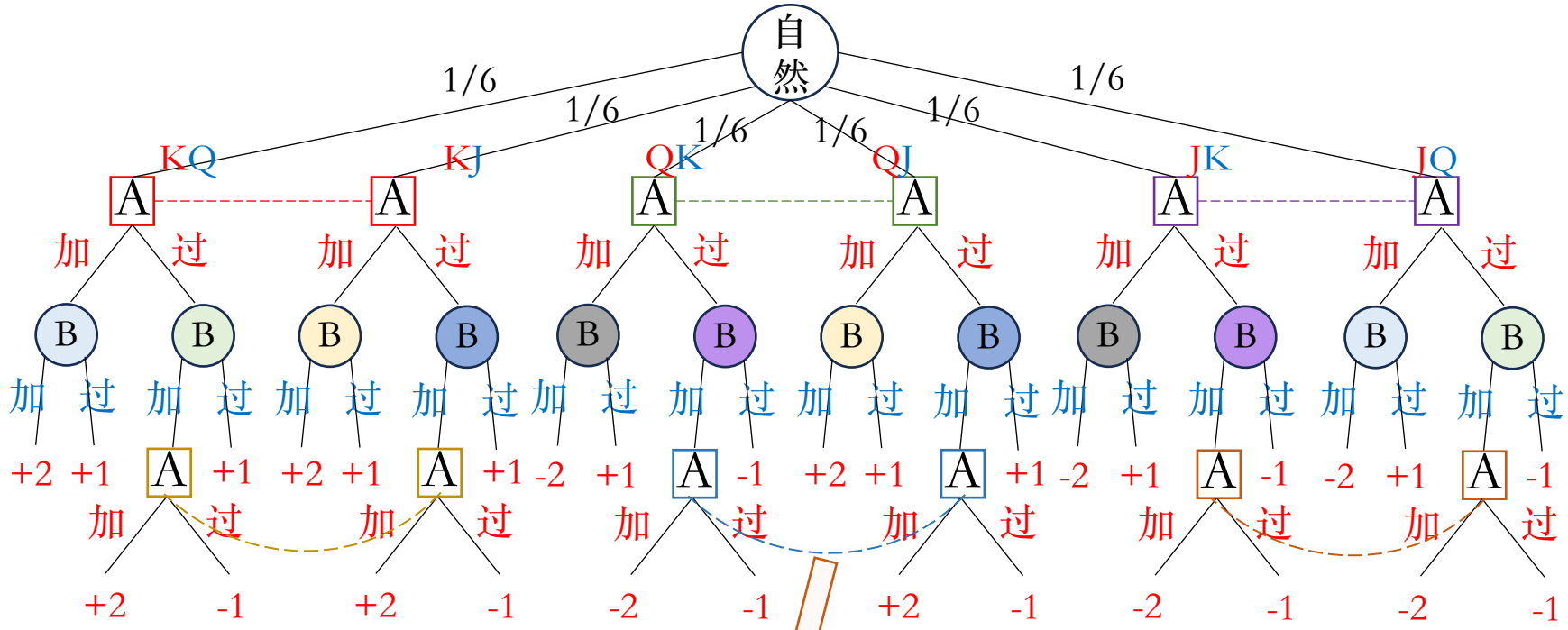
加注	过牌
2	-1
-2	-1

玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为收益

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

加注	过牌
2	-1
-2	-1

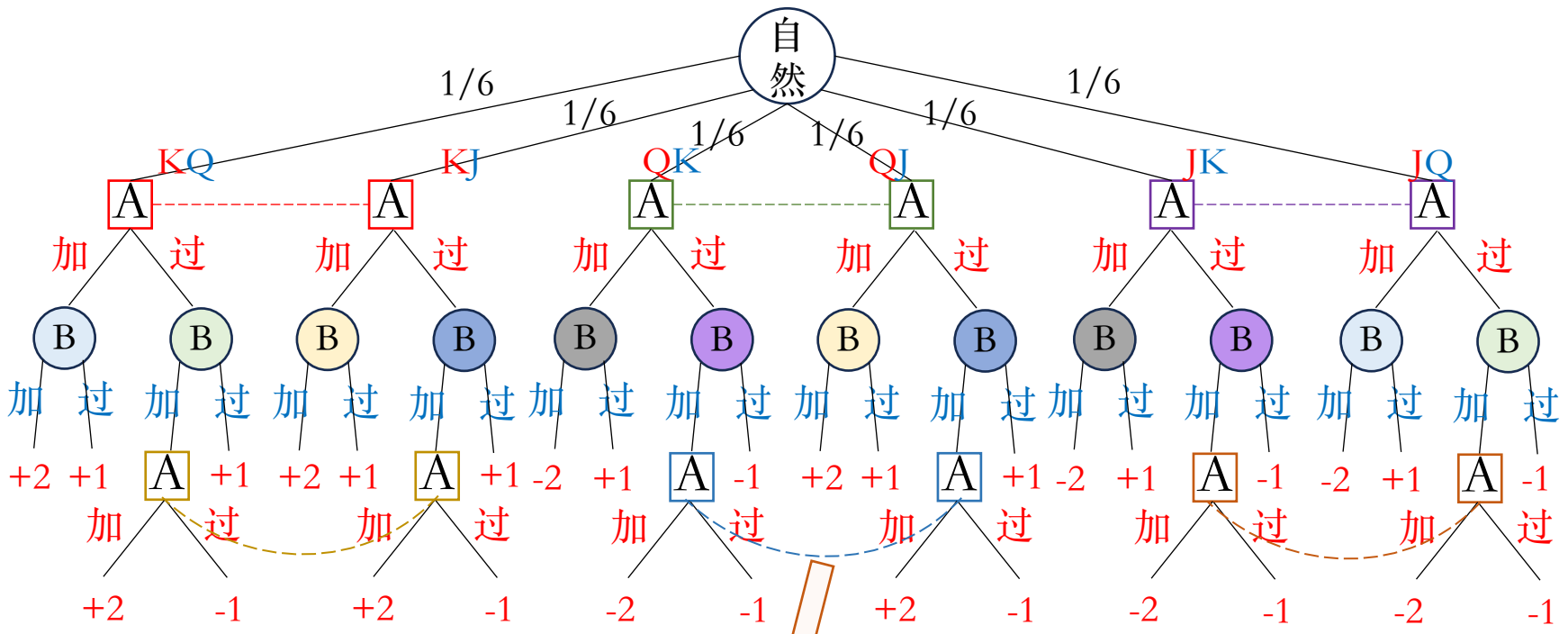
玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

计算过程

加注的收益：

$$\frac{3}{4}*(-2)+\frac{1}{4}*(+2)=-1$$



玩家A

随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为收益

KQ 1/2 KJ 1/2

QK 1/2 QJ 1/2

JK 1/2 JQ 1/2

KQ/过/加 2/3 KJ/过/加 1/3

QK/过/加 3/4 QJ/过/加 1/4

JK/过/加 3/5 JQ/过/加 2/5

加注	过牌
2	-1
-1	-1
-2	-1

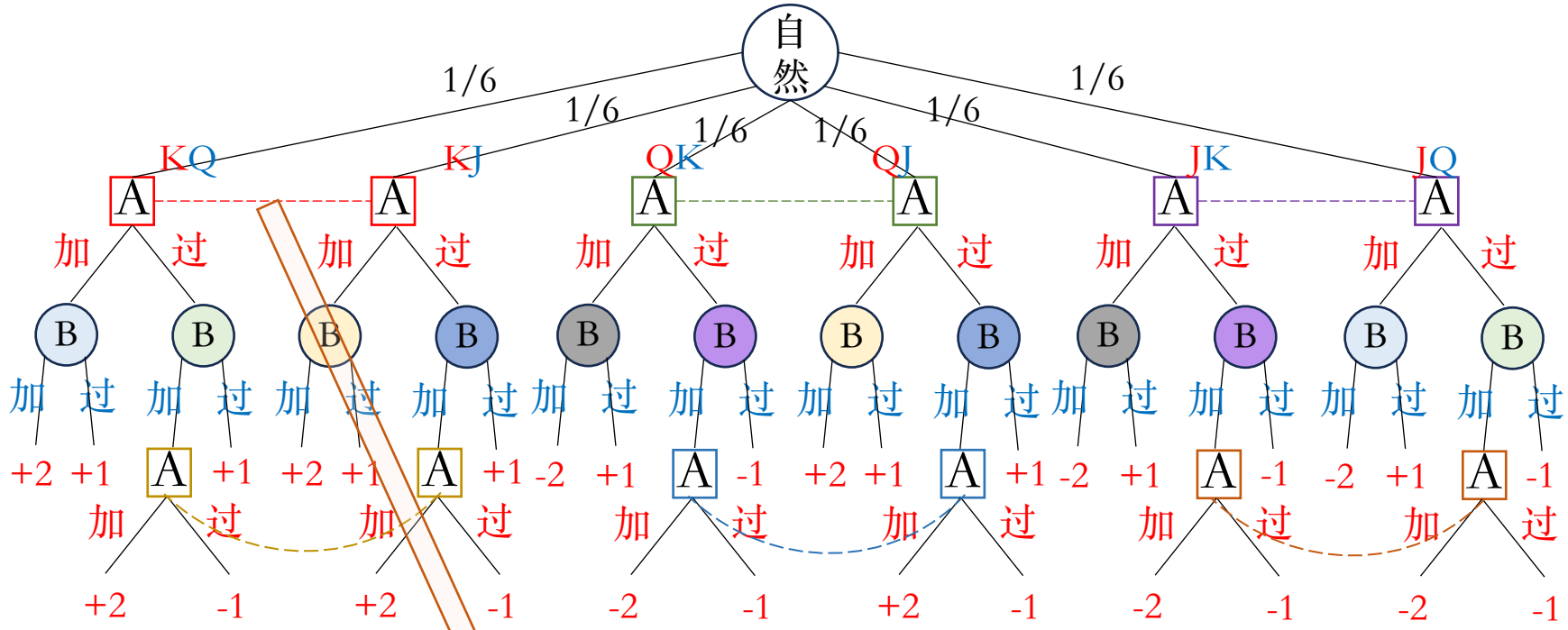
玩家B

随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

计算行为收益

加注	过牌
2	-1
-1	-1
-2	-1

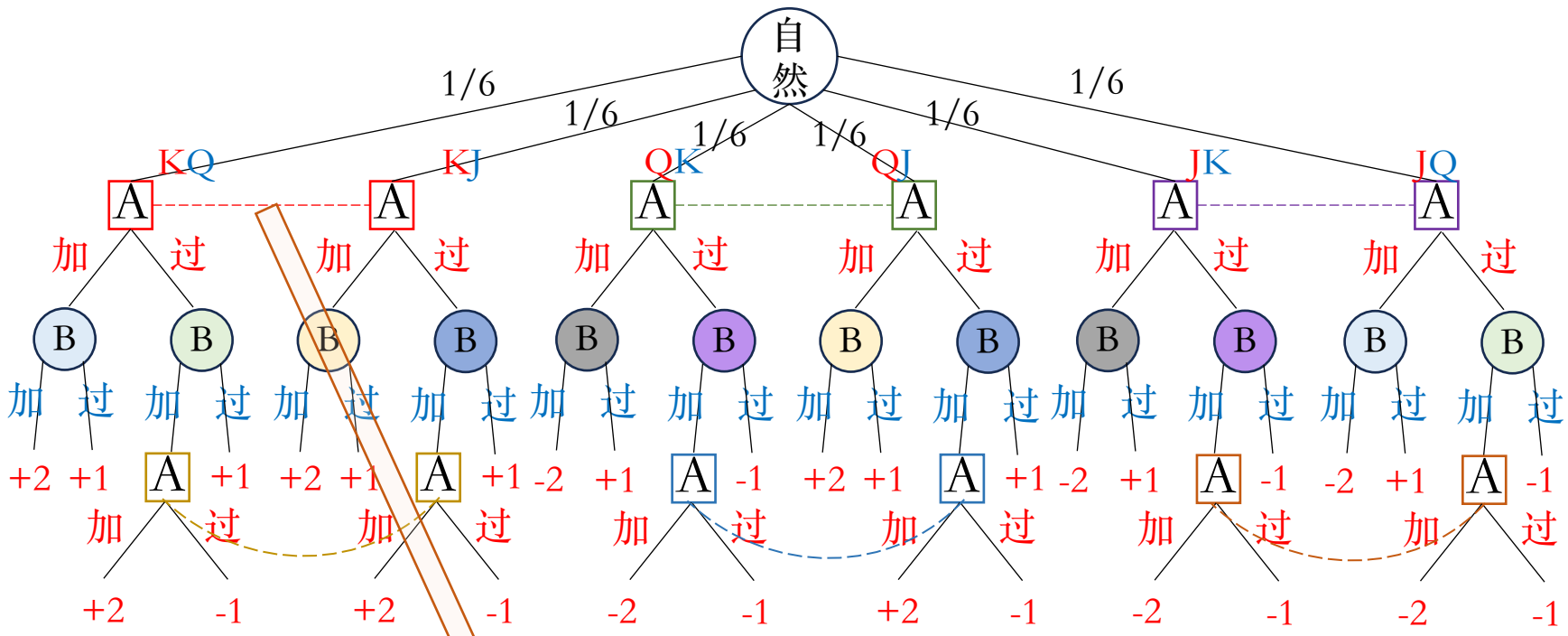
玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

计算过程

加注的收益：

$$\frac{1}{2}*\left(\frac{1}{2}*2+\frac{1}{2}*1\right)+\frac{1}{2}*(0*2+1*1)=\frac{5}{4}$$



玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为收益

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

加注	过牌
5/4	
2	-1
-1	-1
-2	-1

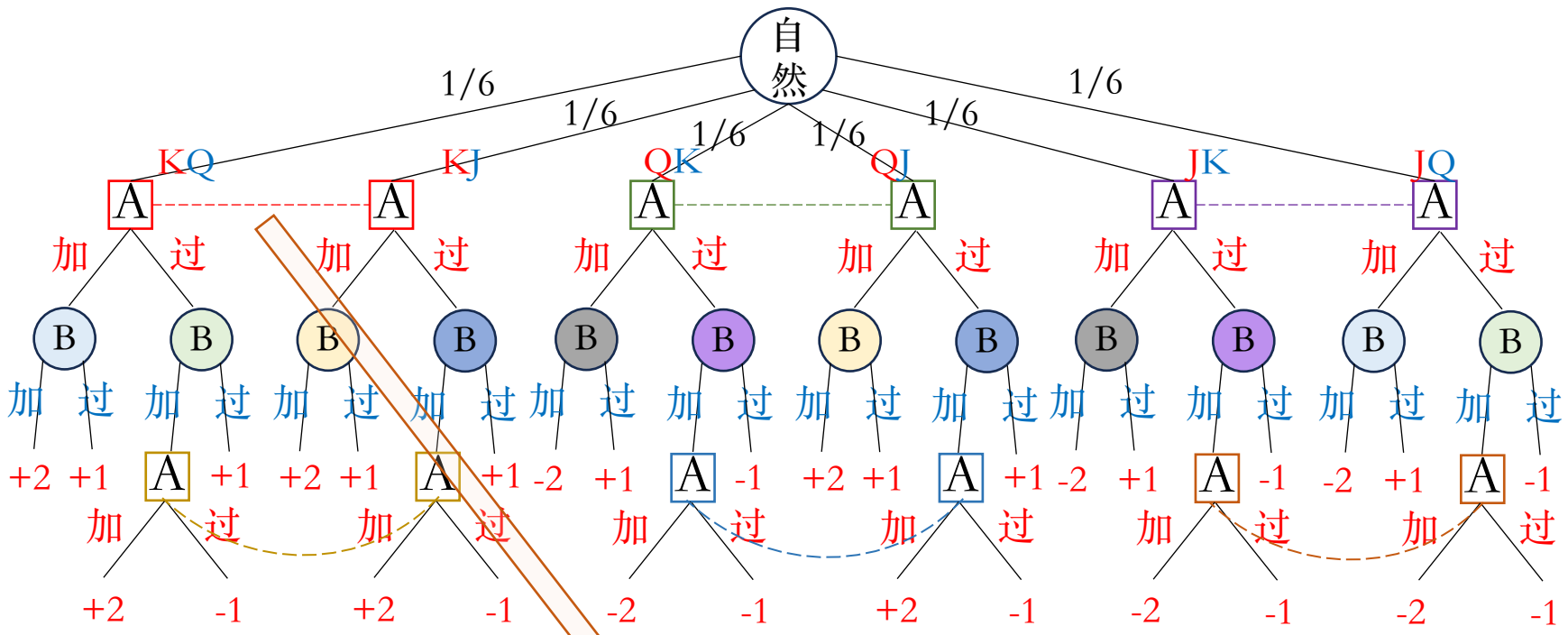
玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

计算过程

过牌的收益：

$$\frac{1}{2}*\left(\frac{2}{3}*(1*2+0*(-1))+\frac{1}{3}*1\right)+\frac{1}{2}*\left(\frac{1}{3}*(1*2+0*(-1))+\frac{2}{3}*1\right)=\frac{3}{2}$$



玩家A　随机初始化　　计算行为收益　　玩家B　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

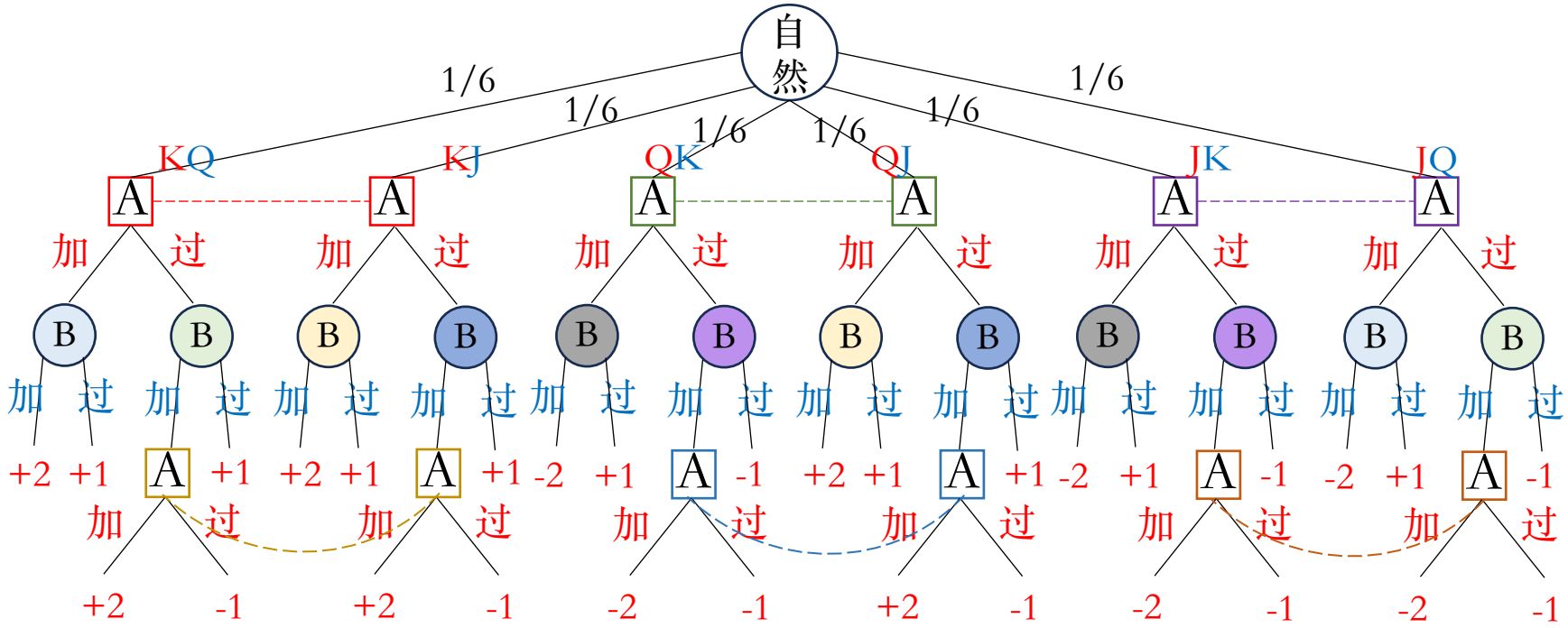
JK/过/加 3/5　JQ/过/加 2/5

加注	过牌
5/4	3/2
2	-1
-1	-1
-2	-1

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤三 计算行为对应收益



根据贝叶斯定理计算概率

玩家A　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

KQ 1/2　KJ 1/2

QK 1/2　QJ 1/2

JK 1/2　JQ 1/2

KQ/过/加 2/3　KJ/过/加 1/3

QK/过/加 3/4　QJ/过/加 1/4

JK/过/加 3/5　JQ/过/加 2/5

计算行为收益

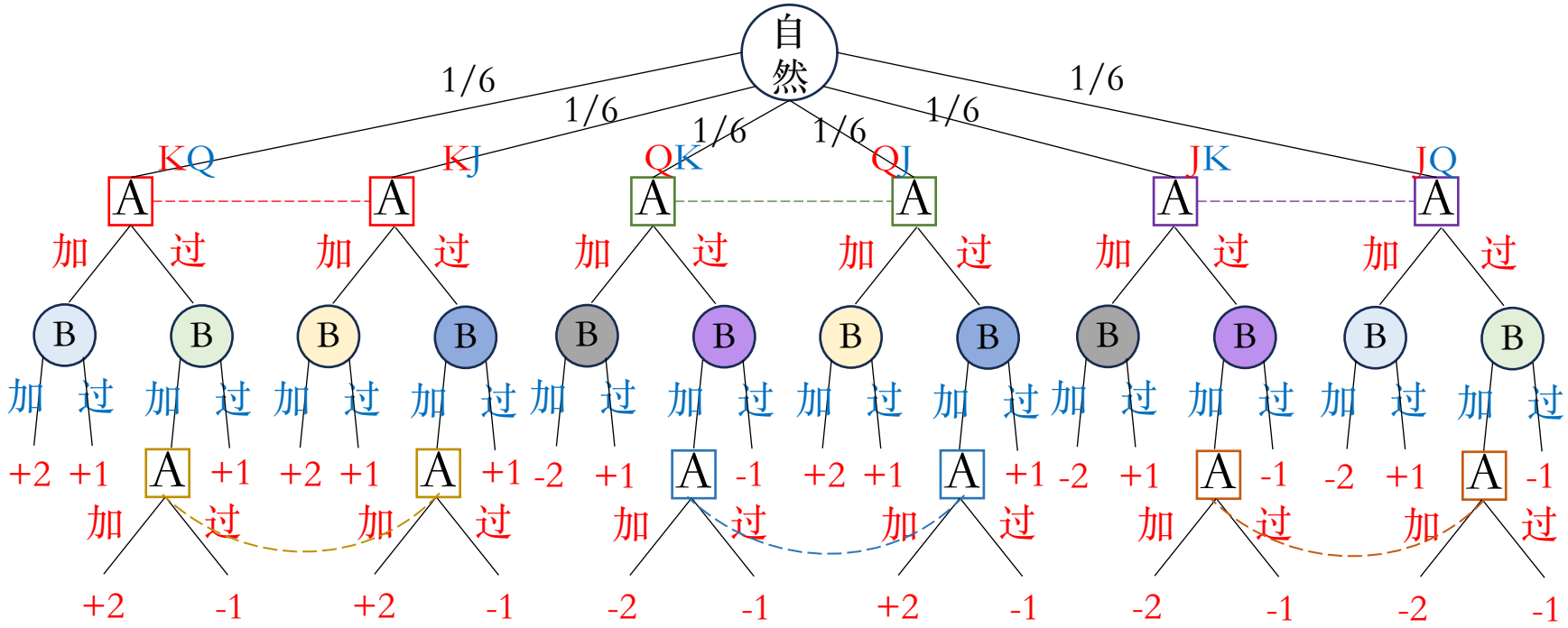
加注	过牌
5/4	3/2
-1/2	-1/3
-5/4	-1
2	-1
-1	-1
-2	-1

玩家B　随机初始化

	加注	过牌
K/加	1	0
K/过	1	0
Q/加	1/2	1/2
Q/过	2/3	1/3
J/加	0	1
J/过	1/3	2/3

双人库恩扑克

步骤四 计算混合策略收益



根据贝叶斯定理计算概率

玩家A

随机初始化

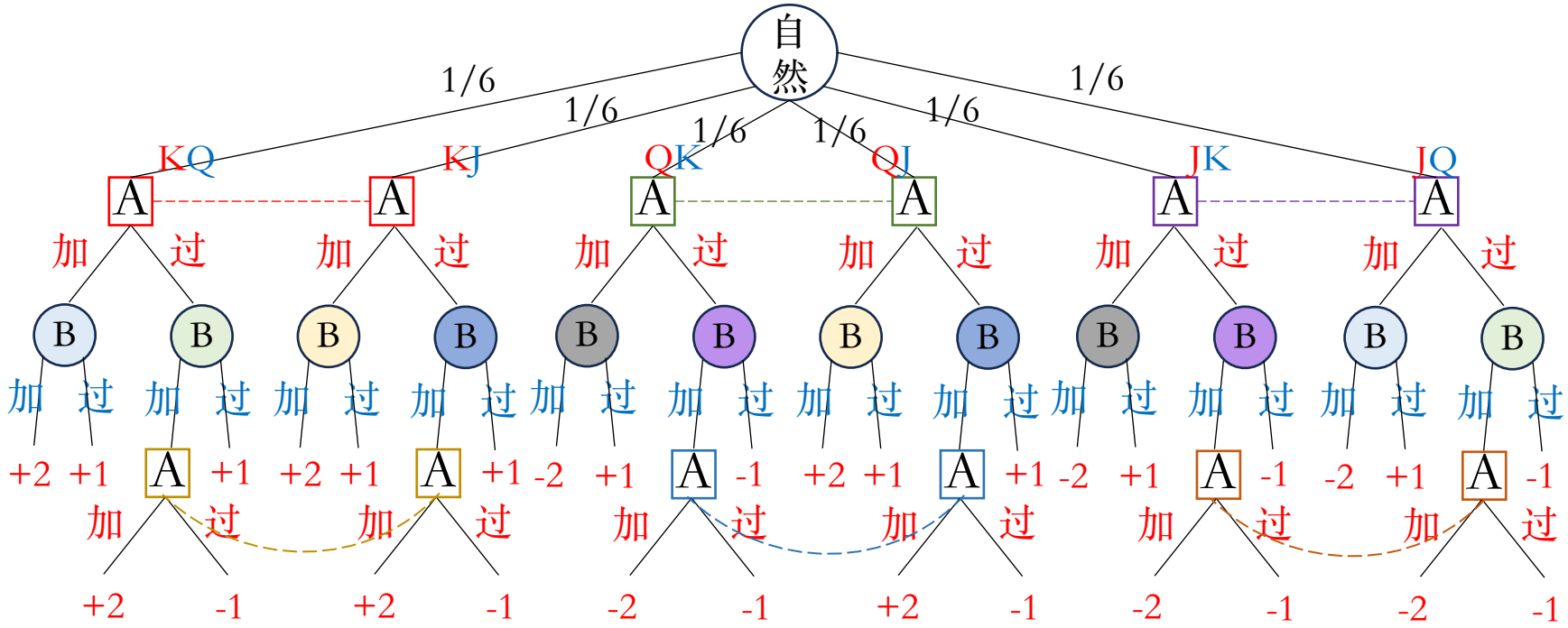
	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为收益

加注	过牌
5/4	3/2
-1/2	-1/3
-5/4	-1
2	-1
-1	-1
-2	-1

双人库恩扑克

步骤四 计算混合策略收益



根据贝叶斯定理计算概率

玩家A　　随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

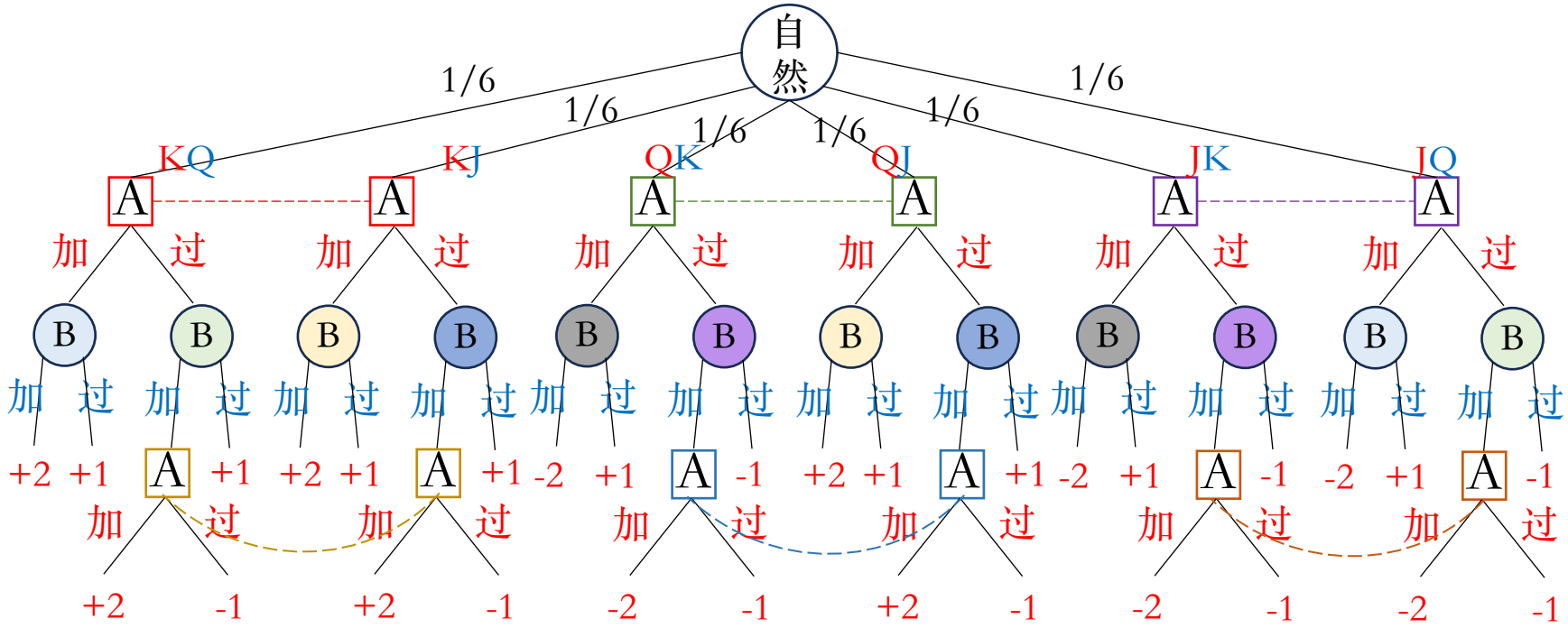
计算行为和混合策略收益

加注	过牌	混合策略
5/4	3/2	4/3
-1/2	-1/3	
-5/4	-1	
2	-1	
-1	-1	
-2	-1	

$$\frac{2}{3}*\frac{5}{4}+\frac{1}{3}*\frac{3}{2}=\frac{4}{3}$$

双人库恩扑克

步骤四 计算混合策略收益



根据贝叶斯定理计算概率

玩家A　　随机初始化

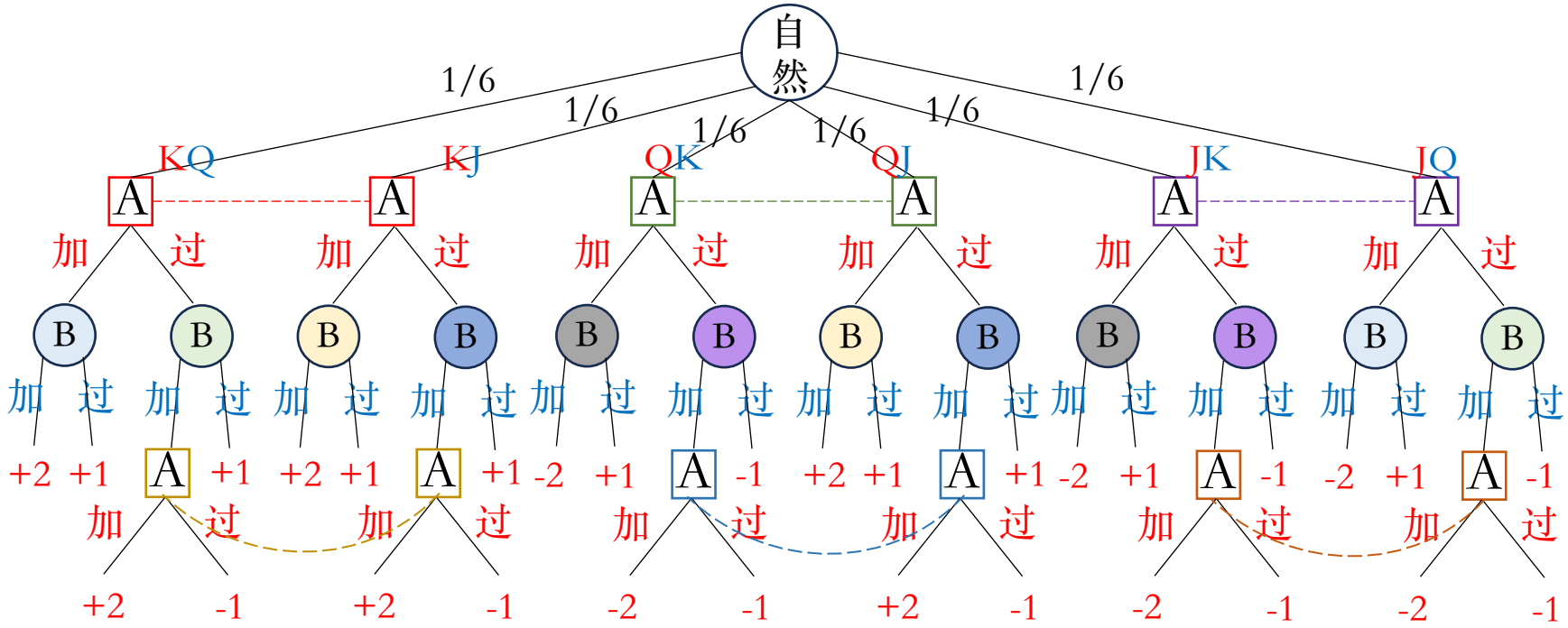
	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为和混合策略收益

加注	过牌	混合策略
5/4	3/2	4/3
-1/2	-1/3	-5/12
-5/4	-1	-13/12
2	-1	2
-1	-1	-1
-2	-1	-1

双人库恩扑克

步骤五 计算行为遗憾值



根据贝叶斯定理计算概率

玩家A

随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为和混合策略收益

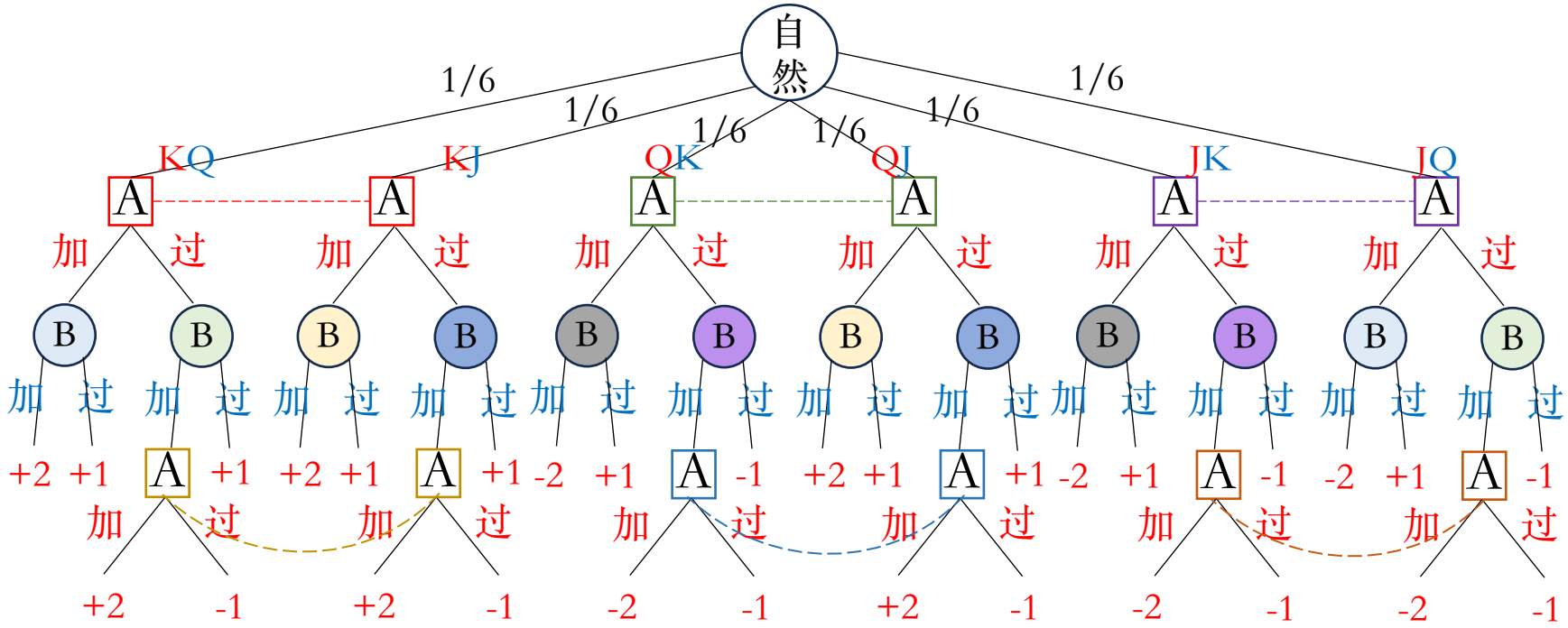
加注	过牌	混合策略
5/4	3/2	4/3
-1/2	-1/3	-5/12
-5/4	-1	-13/12
2	-1	2
-1	-1	-1
-2	-1	-1

计算行为遗憾值

加注	过牌
0	1/6
0	1/12
0	1/12
0	0
0	0
0	0

双人库恩扑克

步骤五 计算行为遗憾值



根据贝叶斯定理计算概率

玩家A

随机初始化

	加注	过牌
K	2/3	1/3
Q	1/2	1/2
J	1/3	2/3
K/过/加	1	0
Q/过/加	1/2	1/2
J/过/加	0	1

计算行为和混合策略收益

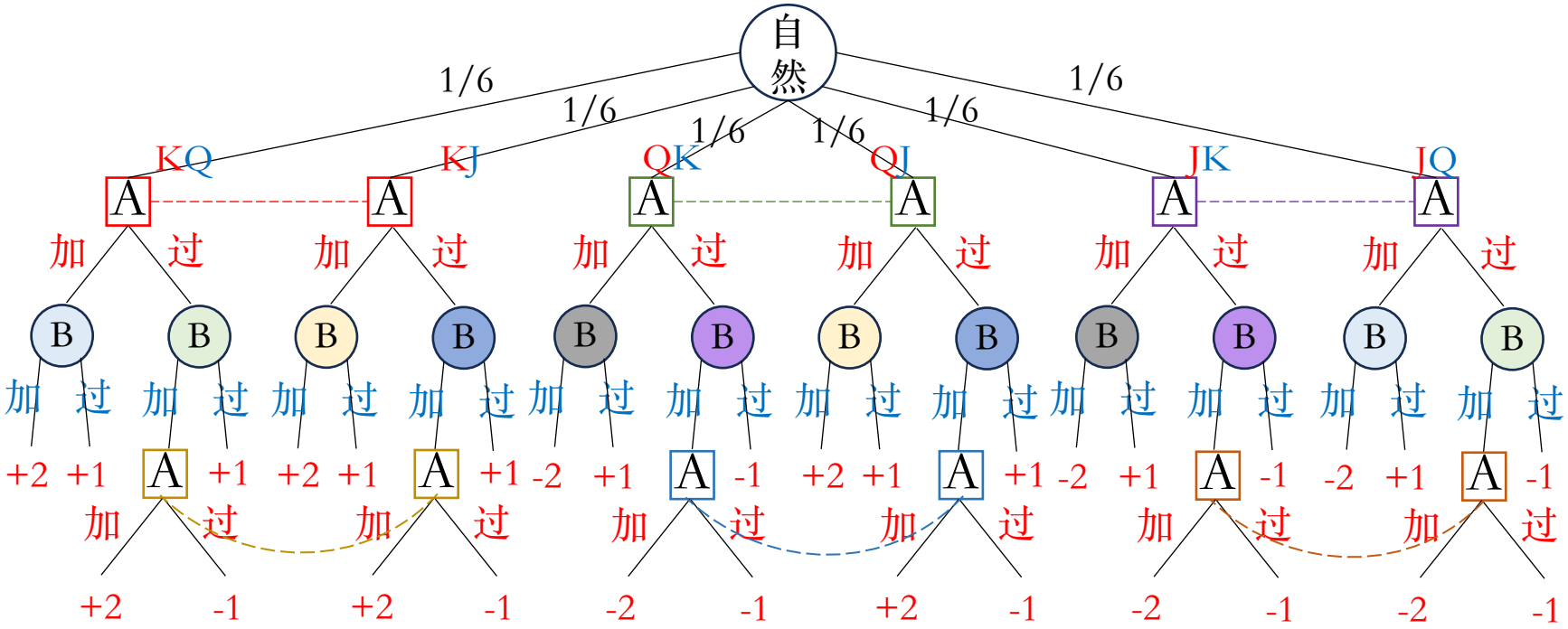
加注	过牌	混合策略
5/4	3/2	4/3
-1/2	-1/3	-5/12
-5/4	-1	-13/12
2	-1	2
-1	-1	-1
-2	-1	-1

计算行为遗憾值

加注	过牌
0	1/6
0	1/12
0	1/12
0	0
0	0
0	0

根据类似步骤，可以为玩家B计算每个信息集下各行为的遗憾值

双人库恩扑克



反事实遗憾最小化

0 为每个玩家、每个信息集、每个行为初始化累加遗憾值

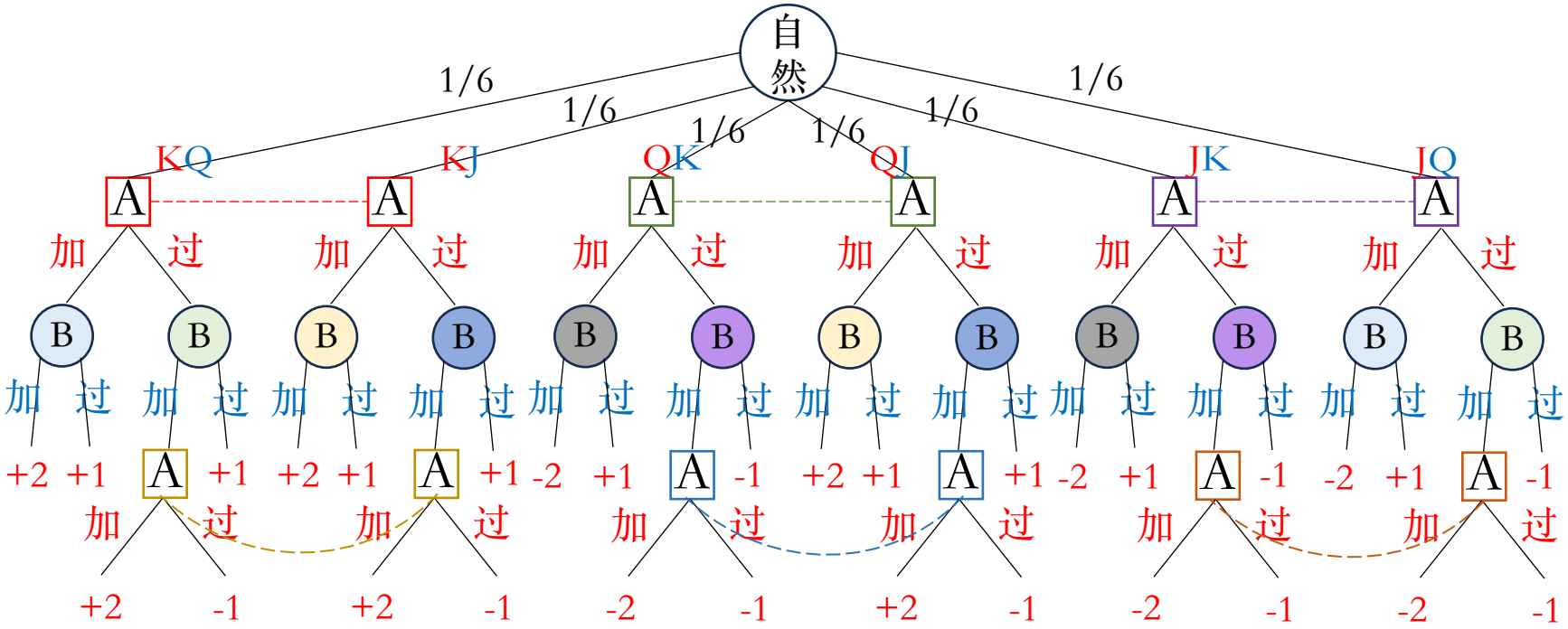
1 为每个玩家、每个信息集，对累加遗憾值向量归一化。由此得到当前的混合策略(σ_1, σ_2)

2 为每个玩家、每个信息集，根据(σ_1, σ_2)计算玩家在当前混合策略下的收益、各行为的遗憾值

3 为每个玩家、每个信息集，利用各行为的遗憾值，更新累加遗憾值向量

4 返回1，直到混合策略(σ_1, σ_2)收敛

双人库恩扑克



反事实遗憾最小化

0 为每个玩家、每个信息集、每个行为初始化累加遗憾值

1 为每个玩家、每个信息集，对累加遗憾值向量归一化。由此得到当前的混合策略(σ_1, σ_2)

2 为每个玩家、每个信息集，根据(σ_1, σ_2)计算玩家在当前混合策略下的收益、各行为的遗憾值

3 为每个玩家、每个信息集，利用各行为的遗憾值，更新累加遗憾值向量

4 返回1，直到混合策略(σ_1, σ_2)收敛

$$R_i^T(I,a) = \frac{1}{T}\sum_{t=1}^{T}\pi_{-i}^{\sigma^t}(I)\left(u_i(\sigma^t|_{I\to a}, I) - u_i(\sigma^t, I)\right)$$

在NIPS07论文中，更新累加遗憾值时，额外乘以一个系数

本讲小结

- 棋牌AI设计
- 矩阵博弈的混合策略均衡、遗憾值匹配算法
- 反事实遗憾最小化算法、库恩扑克的混合策略均衡

主要参考资料

Bryce Wiedenbeck, <Algorithmic Game Theory> Fictitious Play and Regret Matching (online course)

Bryce Wiedenbeck, <Algorithmic Game Theory> Counterfactual Regret Minimization (online course)

Martin Zinkevich et al. <Regret Minimization in Games with Incomplete Information> NIPS 2007 Paper

计算机学院
余皓然
2024/5/14

谢谢！

计算机学院
佘皓然
2024/5/14