Robust Strategies and Counter-Strategies

Building a Champion Level Computer Poker Player

Mike Johanson

September 27, 2007

University of Alberta
Computer Poker Research Group

One Sentence Summary

How can we create a poker program for competing against expert players?

One Sentence Summary

How can we create a poker program for competing against expert players?

- Three new techniques for finding game theoretic strategies

One Sentence Summary

How can we create a poker program for competing against expert players?

- Three new techniques for finding game theoretic strategies
- Useful for poker, applicable to other domains

One Sentence Summary

How can we create a poker program for competing against expert players?

- Three new techniques for finding game theoretic strategies
- Useful for poker, applicable to other domains
- Show the value of these approaches through competitions against expert humans and computers

The Computer Poker Research Group

- The CPRG's goal: Create poker programs to beat the world's best poker players
- Martin Zinkevich and I collaborated on this work

The Computer Poker Research Group

- The CPRG's goal: Create poker programs to beat the world's best poker players
- Martin Zinkevich and I collaborated on this work
    - This is a huge understatement

Texas Hold'em Poker

- Poker is a collection of wagering card games

Texas Hold'em Poker

- Poker is a collection of wagering card games
- Texas Hold'em is considered to be the most strategic variant

Texas Hold'em Poker

- Poker is a collection of wagering card games
- Texas Hold'em is considered to be the most strategic variant
- Players play a series of short games against each other

Texas Hold'em Poker

- Poker is a collection of wagering card games
- Texas Hold'em is considered to be the most strategic variant
- Players play a series of short games against each other
- Goal: Win as much money as possible from opponents over this series of games

- As the game progresses, more cards are revealed

Heads-Up Texas Hold'em Poker

- As the game progresses, more cards are revealed
  - Private cards that only one player can see and use
  - Public cards that all players can see and use

Heads-Up Texas Hold'em Poker

- As the game progresses, more cards are revealed
  - Private cards that only one player can see and use
  - Public cards that all players can see and use
- Players alternate taking actions:
  - Bet: Make a wager that their cards will be the best
  - Call: Match the opponent's wager
  - Fold: Surrender this game, and begin a new one.

Heads-Up Texas Hold'em Poker

Heads-Up Texas Hold'em Poker

- So, why do we care about poker?

Heads-Up Texas Hold'em Poker

- So, why do we care about poker?
- Poker is stochastic and has imperfect information

Heads-Up Texas Hold'em Poker

- So, why do we care about poker?
- Poker is stochastic and has imperfect information
- ...like the real world

Heads-Up Texas Hold'em Poker

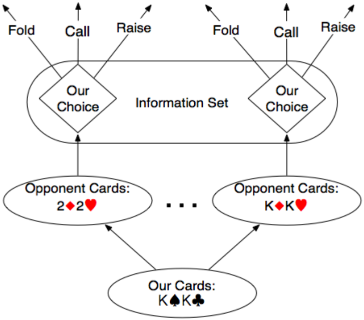
- So, why do we care about poker?
- Poker is stochastic and has imperfect information
- ...like the real world
- Exploitation is important

Heads-Up Texas Hold'em Poker

- So, why do we care about poker?
- Poker is stochastic and has imperfect information
- ...like the real world
- Exploitation is important
- Approaches for other games (such as alpha-beta) don't apply here — we need to find new techniques

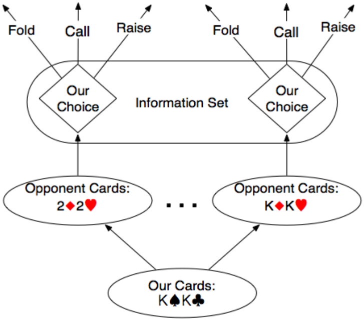
Heads-Up Texas Hold'em Poker

- So, why do we care about poker?
- Poker is stochastic and has imperfect information
- ...like the real world
- Exploitation is important
- Approaches for other games (such as alpha-beta) don't apply here — we need to find new techniques
- Our techniques are applicable beyond poker



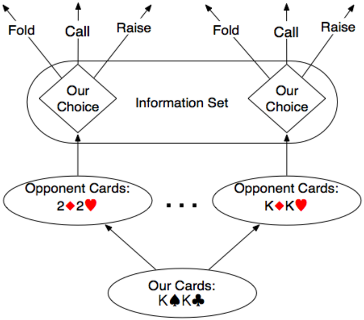
- Because of hidden information, some game states are indistinguishable

Strategies and Information Sets



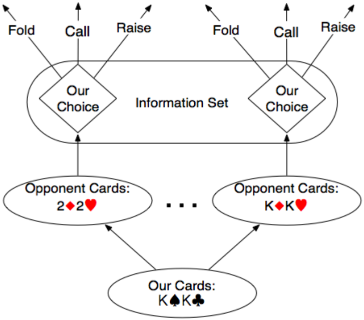
- Because of hidden information, some game states are indistinguishable
- An *information set* is a set of game states that we cannot tell apart

Strategies and Information Sets



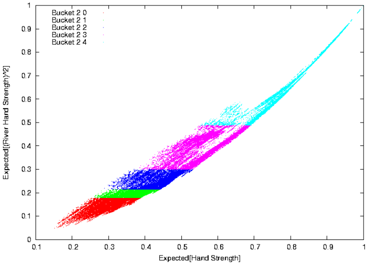
- Because of hidden information, some game states are indistinguishable
- An *information set* is a set of game states that we cannot tell apart
- We have to play the same way for every game state in an information set

Strategies and Information Sets



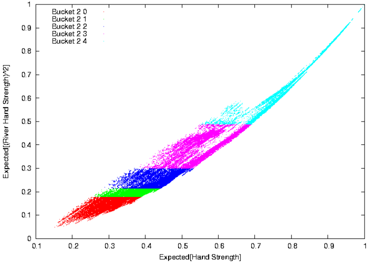
- Because of hidden information, some game states are indistinguishable
- An *information set* is a set of game states that we cannot tell apart
- We have to play the same way for every game state in an information set
- A *behavioral strategy* is a probability distribution over actions for each information set

Computer Poker



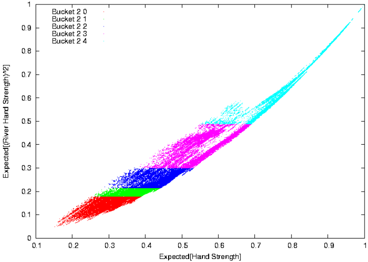
- Poker is big — 10^{18} game states

Computer Poker



- Poker is big — 10^{18} game states
- We abstract the cards into buckets to make the size more reasonable — 10^{12}

Computer Poker



- Poker is big — 10^{18} game states
- We abstract the cards into buckets to make the size more reasonable — 10^{12}
- Poker strategies for the abstract game are still powerful in the "real" game, but there is a loss

Counterfactual Regret Minimization

- First approach: a strategy that works against anyone

Counterfactual Regret Minimization

- First approach: a strategy that works against anyone
- Nash Equilibrium: strategy for each player, where no player can do better by unilaterally changing their strategy

Counterfactual Regret Minimization

- First approach: a strategy that works against anyone
- Nash Equilibrium: strategy for each player, where no player can do better by unilaterally changing their strategy
- Approximation to a Nash equilibrium: no player can do better than ϵ by switching

ϵ-Nash equilibria

- Unbeatable (within its abstraction)

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes
- ...thus "playing to not lose"

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes
- ...thus "playing to not lose"
- (We still use these strategies to win)

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes
- ...thus "playing to not lose"
- (We still use these strategies to win)
- Can be found through linear programming, requires memory proportional to number of game states

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes
- ...thus "playing to not lose"
- (We still use these strategies to win)
- Can be found through linear programming, requires memory proportional to number of game states
- Counterfactual Regret Minimization requires memory proportional to number of *information sets* — much smaller.

ϵ-Nash equilibria

- Unbeatable (within its abstraction)
- The strategy can win if the opponent makes mistakes
- ...thus "playing to not lose"
- (We still use these strategies to win)
- Can be found through linear programming, requires memory proportional to number of game states
- Counterfactual Regret Minimization requires memory proportional to number of *information sets* — much smaller.
- Poker has $3.16 * 10^{17}$ game states and $3.19 * 10^{14}$ information sets

Counterfactual Regret Minimization: Theory

- Play T games of poker, updating your strategy on each round

Counterfactual Regret Minimization: Theory

- Play T games of poker, updating your strategy on each round
- Find the best strategy you could have used for all of those games

Counterfactual Regret Minimization: Theory

- Play T games of poker, updating your strategy on each round
- Find the best strategy you could have used for all of those games
- Define Average Overall Regret as:
  $\frac{1}{T}\sum_{t=1}^{T}((\text{Value of best strategy}) - (\text{Value of your strategy}))$

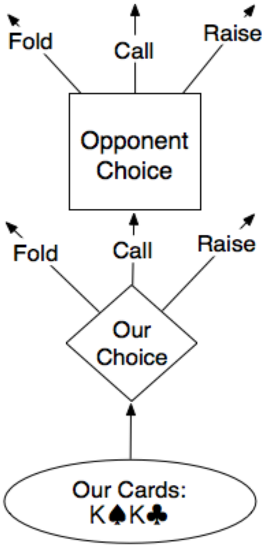
Counterfactual Regret Minimization: Theory

- Play T games of poker, updating your strategy on each round
- Find the best strategy you could have used for all of those games
- Define Average Overall Regret as:
  $\frac{1}{T}\sum_{t=1}^{T}((\text{Value of best strategy}) - (\text{Value of your strategy}))$
- If we minimize Average Overall Regret, the *average strategy* used over the T games approaches a Nash equilibrium

Counterfactual Regret Minimization: Theory

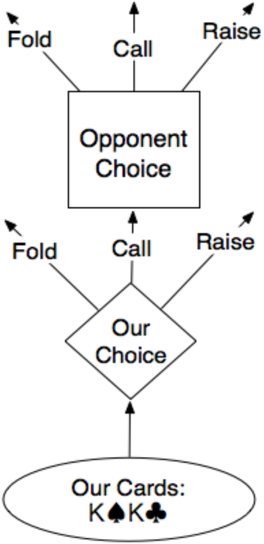
- Play T games of poker, updating your strategy on each round
- Find the best strategy you could have used for all of those games
- Define Average Overall Regret as:
  $\frac{1}{T}\sum_{t=1}^{T}((\text{Value of best strategy}) - (\text{Value of your strategy}))$
- If we minimize Average Overall Regret, the *average strategy* used over the T games approaches a Nash equilibrium
- How do we minimize Average Overall Regret?

Immediate Counterfactual Regret



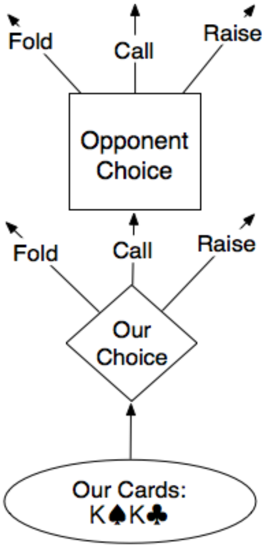
- Break down overall regret into the regret for each action at each information set

Immediate Counterfactual Regret



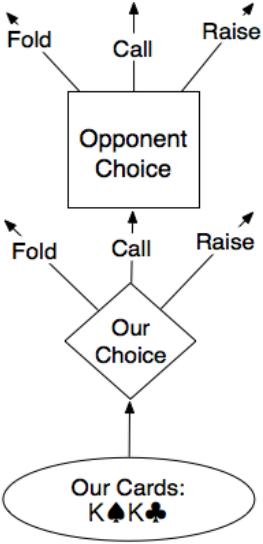
- Break down overall regret into the regret for each action at each information set
- Regret: How much *more* utility we could have had if we always took some action instead of using our strategy

Immediate Counterfactual Regret



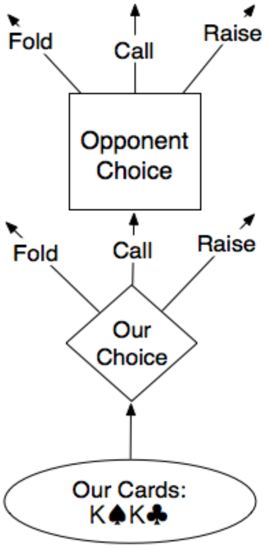
- Break down overall regret into the regret for each action at each information set
- Regret: How much *more* utility we could have had if we always took some action instead of using our strategy
- Immediate Counterfactual Regret: Weight this regret by the probability of the opponent reaching the information set

Immediate Counterfactual Regret



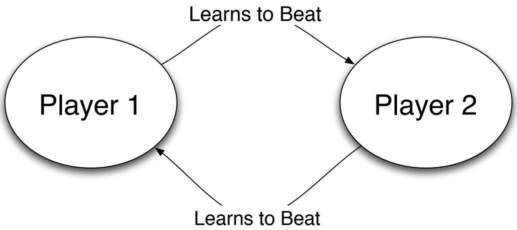
- Break down overall regret into the regret for each action at each information set
- Regret: How much *more* utility we could have had if we always took some action instead of using our strategy
- Immediate Counterfactual Regret: Weight this regret by the probability of the opponent reaching the information set
- Average Overall Regret is less than the sum of Immediate Counterfactual Regret

Immediate Counterfactual Regret



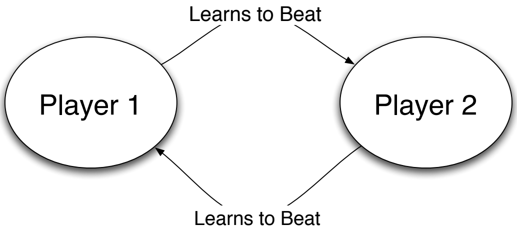
- Break down overall regret into the regret for each action at each information set
- Regret: How much *more* utility we could have had if we always took some action instead of using our strategy
- Immediate Counterfactual Regret: Weight this regret by the probability of the opponent reaching the information set
- Average Overall Regret is less than the sum of Immediate Counterfactual Regret
- So, if we can minimize our immediate counterfactual regret *at each information set* , then we approach a Nash equilibrium

Counterfactual Regret Minimization: Basic Idea



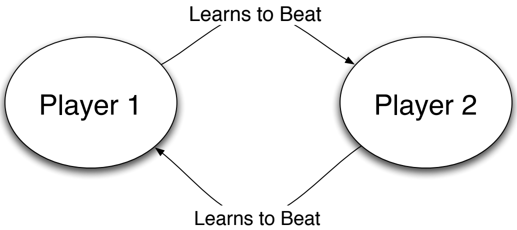
- Initialize the strategies' action probabilities to a uniform distribution

Counterfactual Regret Minimization: Basic Idea



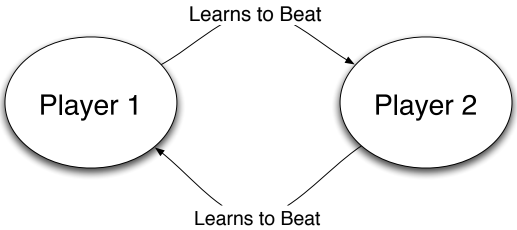
- Initialize the strategies' action probabilities to a uniform distribution
- Repeat:
    - (General) Iterate over all chance outcomes

Counterfactual Regret Minimization: Basic Idea



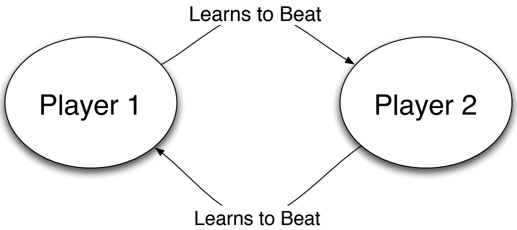
- Initialize the strategies' action probabilities to a uniform distribution
- Repeat:
  - (General) Iterate over all chance outcomes
  - (Poker-specific) Deal cards to each player, as if playing the game

Counterfactual Regret Minimization: Basic Idea



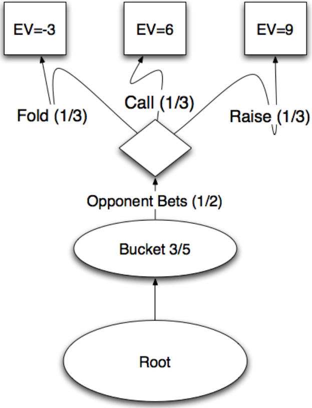
- Initialize the strategies' action probabilities to a uniform distribution
- Repeat:
  - (General) Iterate over all chance outcomes
  - (Poker-specific) Deal cards to each player, as if playing the game
  - Recurse over all choice nodes. Update the action probabilities at each choice node to minimize regret at that node.

Counterfactual Regret Minimization: Basic Idea

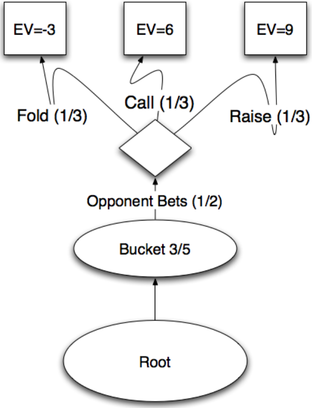


- Initialize the strategies' action probabilities to a uniform distribution
- Repeat:
  - (General) Iterate over all chance outcomes
  - (Poker-specific) Deal cards to each player, as if playing the game
  - Recurse over all choice nodes. Update the action probabilities at each choice node to minimize regret at that node.
- How do we update the action probabilities after each game?

Counterfactual Regret



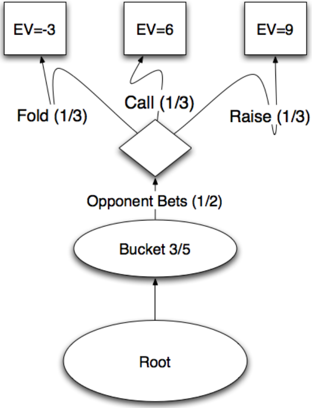
Counterfactual Regret



- Compute expected value of each action

- Strategy's EV: 4

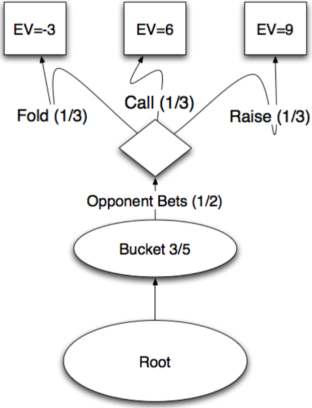
Counterfactual Regret



- Compute expected value of each action
- Calculate the *regret* for not taking each action
- (Regret: Difference between the EV for taking an action and the strategy's EV)

- Strategy's EV: 4
- Regret: (-7, 2, 5)

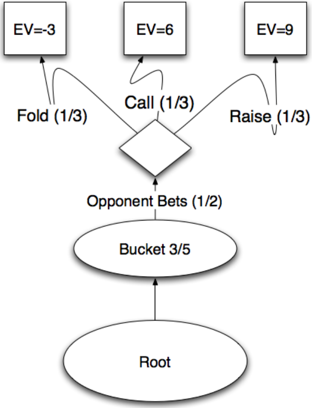
Counterfactual Regret



- Compute expected value of each action
- Calculate the *regret* for not taking each action
- (Regret: Difference between the EV for taking an action and the strategy's EV)
- Counterfactual Regret: Regret weighted by opponent's probability of reaching this state
- Add up Counterfactual Regret over all games

- Strategy's EV: 4
- Regret: (-7, 2, 5)
- Total CFR: (-3.5, 1, 2.5)

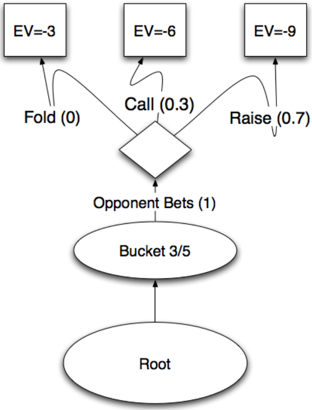
Counterfactual Regret



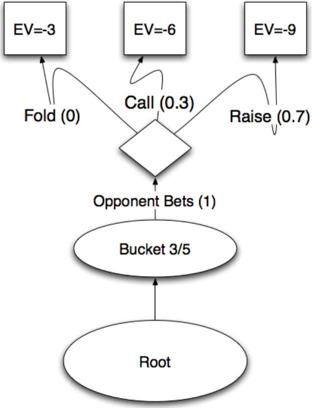
- Compute expected value of each action
- Calculate the *regret* for not taking each action
- (Regret: Difference between the EV for taking an action and the strategy's EV)
- Counterfactual Regret: Regret weighted by opponent's probability of reaching this state
- Add up Counterfactual Regret over all games
- Assign new probabilities proportional to accumulated positive CFR

- Strategy's EV: 4
- Regret: (-7, 2, 5)
- Total CFR: (-3.5, 1, 2.5)
- New Probabilities: (0, 0.3, 0.7)

Counterfactual Regret Example 2

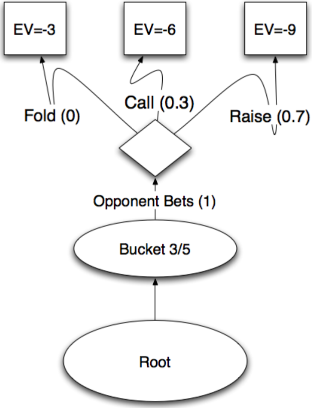


Counterfactual Regret Example 2



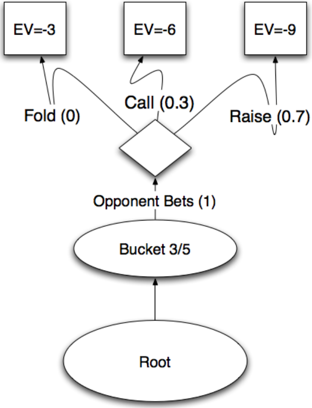
- Strategy's EV: -8.1

Counterfactual Regret Example 2



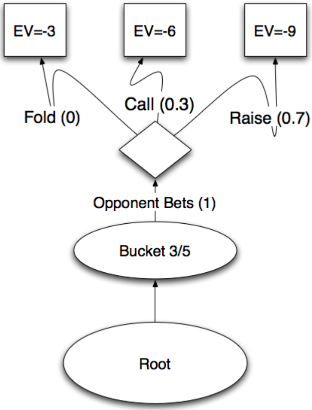
- Strategy's EV: -8.1
- Regret: (5.1, 2.1, -0.9)

Counterfactual Regret Example 2



- Strategy's EV: -8.1
- Regret: (5.1, 2.1, -0.9)
- Total CFR: (1.6, 3.1, 1.6)

Counterfactual Regret Example 2



- Strategy's EV: -8.1
- Regret: (5.1, 2.1, -0.9)
- Total CFR: (1.6, 3.1, 1.6)
- New Probabilities: (0.25, 0.5, 0.25)

Performance Bounds

- Counterfactual Regret Minimization approaches a Nash equilibrium - how fast does it get there?

Performance Bounds

- Counterfactual Regret Minimization approaches a Nash equilibrium - how fast does it get there?
  - General: # iterations grows quadratically with # information sets

Performance Bounds

- Counterfactual Regret Minimization approaches a Nash equilibrium - how fast does it get there?
  - General: # iterations grows quadratically with # information sets
  - Poker: # iterations grows *linearly* with # information sets

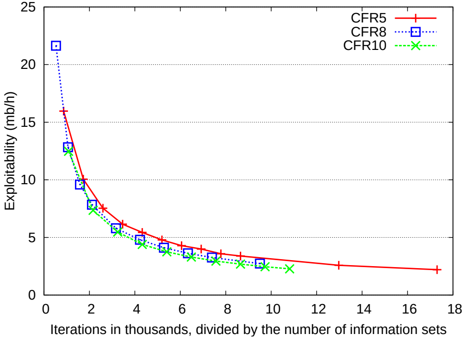
Performance Bounds

- Counterfactual Regret Minimization approaches a Nash equilibrium - how fast does it get there?
  - General: # iterations grows quadratically with # information sets
  - Poker: # iterations grows *linearly* with # information sets
  - (Because seeing a few samples of the states in an information set is enough to choose a good strategy for that information set)

Performance Bounds

- Counterfactual Regret Minimization approaches a Nash equilibrium - how fast does it get there?
  - General: # iterations grows quadratically with # information sets
  - Poker: # iterations grows *linearly* with # information sets
  - (Because seeing a few samples of the states in an information set is enough to choose a good strategy for that information set)
- In practical terms: we can solve very large games (10^{12} *states*) in under two weeks
- That's two orders of magnitude larger than was previously possible

Convergence to a Nash Equilibrium



Abstraction	Size (game states) ($\times 10^9$)	Iterations ($\times 10^6$)	Time (h)	Exp (mb/h)
5	6.45	100	33	3.4
6	27.7	200	75	3.1
8	276	750	261	2.7
10	1646	2000	326	2.2

Comparison to the 2006 AAAI Competition

	Hyperborean	Bluffbot	Monash	Teddy	Average
Smallbot2298	61	113	695	474	336
CFR8	106	170	746	517	385

Counterfactual Regret Minimization: Conclusions

- Approaches Nash Equilibria faster and with less memory than older techniques

Counterfactual Regret Minimization: Conclusions

- Approaches Nash Equilibria faster and with less memory than older techniques
- The resulting strategies are *robust* — they work well against any opponent

Counterfactual Regret Minimization: Conclusions

- Approaches Nash Equilibria faster and with less memory than older techniques
- The resulting strategies are *robust* — they work well against any opponent
- But... How exploitable are the opponents?

Counterfactual Regret Minimization: Conclusions

- Approaches Nash Equilibria faster and with less memory than older techniques
- The resulting strategies are *robust* — they work well against any opponent
- But... How exploitable are the opponents?
- How much better could an exploitive strategy do?

Counterfactual Regret Minimization: Conclusions

- Approaches Nash Equilibria faster and with less memory than older techniques
- The resulting strategies are *robust* — they work well against any opponent
- But... How exploitable are the opponents?
- How much better could an exploitive strategy do?
- "Playing to Not Lose"

Frequentist Best Response

- Best Response: best possible counter-strategy to some strategy

Frequentist Best Response

- Best Response: best possible counter-strategy to some strategy
- Useful for a few reasons:
  - Tells you how exploitable that strategy is
  - Could use it during a match to win

Best Response Challenges

- “real” best response is intractable

Best Response Challenges

- "real" best response is intractable
- abstract game best response is easy, but has some challenges:
  - Need to actually have the opponent's strategy
  - Resulting counter-strategy plays in the same abstraction as the strategy

Best Response Challenges

- "real" best response is intractable
- abstract game best response is easy, but has some challenges:
  - Need to actually have the opponent's strategy
  - Resulting counter-strategy plays in the same abstraction as the strategy
    - (Bigger abstraction == better counter-strategy)

Motivating Frequentist Best Response

- We'd like to make best response counter-strategies with fewer restrictions:
  - What if we don't have the actual strategy, only observations?
  - What if we want to choose the abstraction that the counter-strategy uses?

Frequentist Best Response: Basic Idea

- Observe lots of real-game data — say, 1 million hands

Frequentist Best Response: Basic Idea

- Observe lots of real-game data — say, 1 million hands
- Abstract the data, and do frequency counts on how often actions are taken in each choice node

Frequentist Best Response: Basic Idea

- Observe lots of real-game data — say, 1 million hands
- Abstract the data, and do frequency counts on how often actions are taken in each choice node
- Construct an opponent model, where action probabilities are just the action frequencies

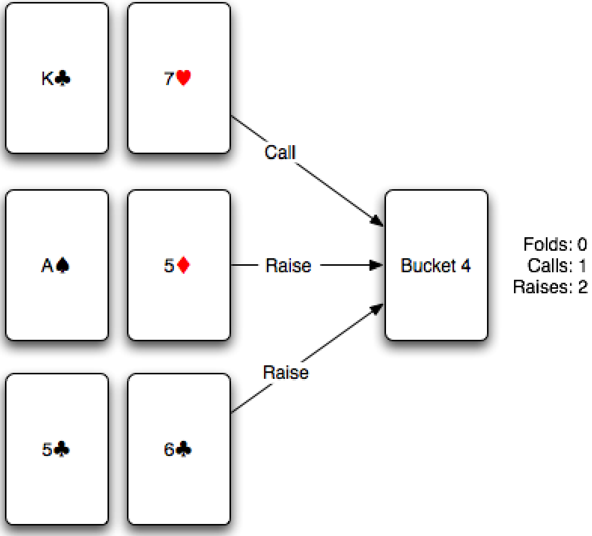
Frequentist Best Response: Basic Idea

- Observe lots of real-game data — say, 1 million hands
- Abstract the data, and do frequency counts on how often actions are taken in each choice node
- Construct an opponent model, where action probabilities are just the action frequencies
- Find the abstract game best response to the opponent model

Frequentist Best Response: Basic Idea

- Observe lots of real-game data — say, 1 million hands
- Abstract the data, and do frequency counts on how often actions are taken in each choice node
- Construct an opponent model, where action probabilities are just the action frequencies
- Find the abstract game best response to the opponent model
- Use the counter-strategy to play against the strategy in the real game

Abstracting the data



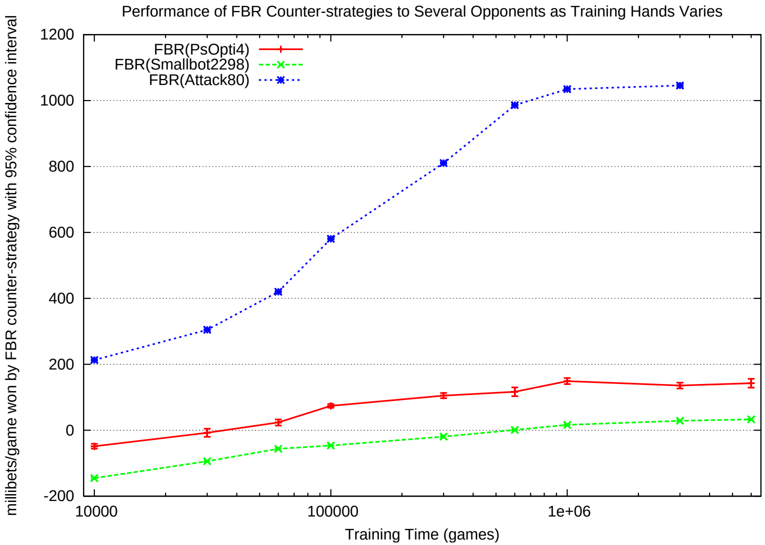
Frequentist Best Response

- There's a few variables you need to get right:

Frequentist Best Response

- There's a few variables you need to get right:
  - Who is the strategy playing against for the million hands? (Self play is bad, because it doesn't explore the whole strategy space)
  - What do you do in states you never observe? (We assume they call)

Frequentist Best Response



Frequentist Best Response

	PsOpti4	PsOpti6	Attack60	Attack80	Smallbot1239	Smallbot1399	Smallbot2298	CFR5	Average
FBR-PsOpti4	137	-163	-227	-231	-106	-85	-144	-210	-129
FBR-PsOpti6	-79	330	-68	-89	-36	-23	-48	-97	-14
FBR-Attack60	-442	-499	2170	-701	-359	-305	-377	-620	-142
FBR-Attack80	-312	-281	-557	1048	-251	-231	-266	-331	-148
FBR-Smallbot1239	-20	105	-89	-42	106	91	-32	-87	3
FBR-Smallbot1399	-43	38	-48	-77	75	118	-46	-109	-11
FBR-Smallbot2298	-39	51	-50	-26	42	50	33	-41	2
CFR5	36	123	93	41	70	68	17	0	56
Max	137	330	2170	1048	106	118	33	0	

Table: Each entry is the result of a match between the row and the column player. Score is the average amount won by the row player, in millibets / hand. One millibet is 0.001 small bets.

- Columns are poker strategies we've produced in the past
- Rows are counter-strategies to each strategy
- CFR5 is a Counterfactual Regret Minimization strategy

Frequentist Best Response

	PsOpti4	PsOpti6	Attack60	Attack80	Smallbot1239	Smallbot1399	Smallbot2298	CFR5	Average
FBR-PsOpti4	137	-163	-227	-231	-106	-85	-144	-210	-129
FBR-PsOpti6	-79	330	-68	-89	-36	-23	-48	-97	-14
FBR-Attack60	-442	-499	2170	-701	-359	-305	-377	-620	-142
FBR-Attack80	-312	-281	-557	1048	-251	-231	-266	-331	-148
FBR-Smallbot1239	-20	105	-89	-42	106	91	-32	-87	3
FBR-Smallbot1399	-43	38	-48	-77	75	118	-46	-109	-11
FBR-Smallbot2298	-39	51	-50	-26	42	50	33	-41	2
CFR5	36	123	93	41	70	68	17	0	56
Max	137	330	2170	1048	106	118	33	0	

Table: Each entry is the result of a match between the row and the column player. Score is the average amount won by the row player, in millibets / hand. One millibet is 0.001 small bets.

- Columns are poker strategies we've produced in the past
- Rows are counter-strategies to each strategy
- CFR5 is a Counterfactual Regret Minimization strategy
- Two observations:
  - The diagonal has the matches where the counter-strategy plays against its intended opponent. These scores are all good - significantly higher than the CFR strategy does

Frequentist Best Response: Conclusions

- "Playing to Win"

Frequentist Best Response: Conclusions

- "Playing to Win"
- Frequentist Best Response counter-strategies are useful for defeating specific opponents

Frequentist Best Response: Conclusions

- "Playing to Win"
- Frequentist Best Response counter-strategies are useful for defeating specific opponents
- We also use them to evaluate our strategies, to see how weak they are

Frequentist Best Response: Conclusions

- "Playing to Win"
- Frequentist Best Response counter-strategies are useful for defeating specific opponents
- We also use them to evaluate our strategies, to see how weak they are
- However, they are *brittle* — when used against other opponens, even weak ones, they can lose badly.

Frequentist Best Response: Conclusions

- "Playing to Win"
- Frequentist Best Response counter-strategies are useful for defeating specific opponents
- We also use them to evaluate our strategies, to see how weak they are
- However, they are *brittle* — when used against other opponens, even weak ones, they can lose badly.
- Is there a way to keep the exploitiveness of FBR counter-strategies, while also gaining the robustness of CFR strategies?

Restricted Nash Response

- Exploiting opponents is important — we'd like to win more money than the Counterfactual Regret Minimization strategies do

Restricted Nash Response

- Exploiting opponents is important — we'd like to win more money than the Counterfactual Regret Minimization strategies do
- Frequentist Best Response strategies win lots of money, but are terrible against the wrong opponent

Restricted Nash Response

- Exploiting opponents is important — we'd like to win more money than the Counterfactual Regret Minimization strategies do
- Frequentist Best Response strategies win lots of money, but are terrible against the wrong opponent
- We'd like a compromise: a strategy that exploits an opponent (or class of opponents), but is also *robust* against arbitrary opponents

Restricted Nash Response: Motivation

- We suspect our opponent will use some strategy

Restricted Nash Response: Motivation

- We suspect our opponent will use some strategy
- What if they only used it, say, 75% of the time?

Restricted Nash Response: Motivation

- We suspect our opponent will use some strategy
- What if they only used it, say, 75% of the time?
- The other 25% of the time, they can do anything...

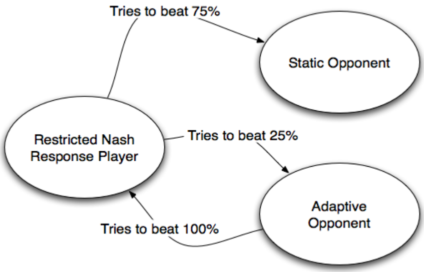
Restricted Nash Response: Motivation

- We suspect our opponent will use some strategy
- What if they only used it, say, 75% of the time?
- The other 25% of the time, they can do anything...
- ...but lets assume they play a best response to whatever we do

Restricted Nash Response: Motivation

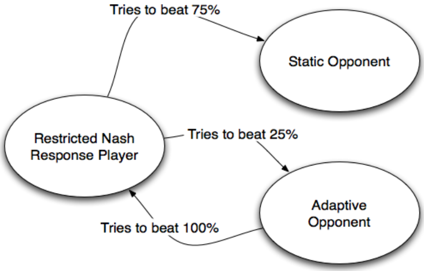
- We suspect our opponent will use some strategy
- What if they only used it, say, 75% of the time?
- The other 25% of the time, they can do anything...
- ...but lets assume they play a best response to whatever we do
- We now have two goals: attack the 75% “weak” strategy, and defend against the 25% “adaptive” strategy

Restricted Nash Response: Basic Idea



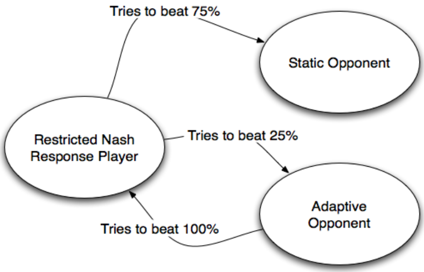
- In CFR, we had two strategies that adapt to beat each other

Restricted Nash Response: Basic Idea



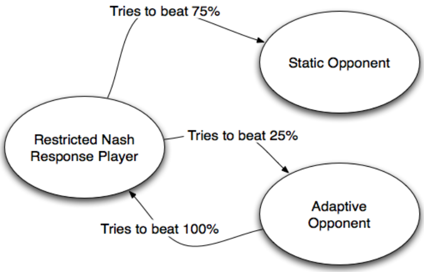
- In CFR, we had two strategies that adapt to beat each other
- In RNR, we have one strategy for our player, and two for our opponent

Restricted Nash Response: Basic Idea



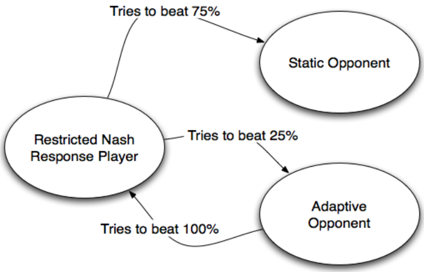
- In CFR, we had two strategies that adapt to beat each other
- In RNR, we have one strategy for our player, and two for our opponent
- The opponent's static strategy is the model we get from Frequentist Best Response

Restricted Nash Response: Basic Idea



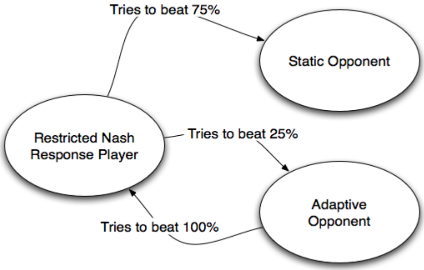
- In CFR, we had two strategies that adapt to beat each other
- In RNR, we have one strategy for our player, and two for our opponent
- The opponent's static strategy is the model we get from Frequentist Best Response
- We play millions of games, where our player minimizes regret when playing against both the static and adaptive opponent

Restricted Nash Response: Basic Idea



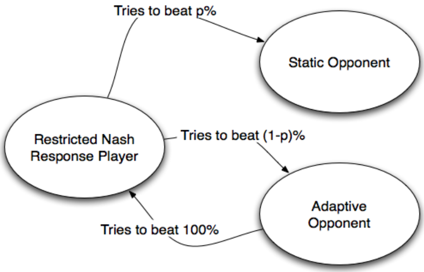
- In CFR, we had two strategies that adapt to beat each other
- In RNR, we have one strategy for our player, and two for our opponent
- The opponent's static strategy is the model we get from Frequentist Best Response
- We play millions of games, where our player minimizes regret when playing against both the static and adaptive opponent
- The adaptive opponent minimizes regret when playing against us

Restricted Nash Response: Basic Idea



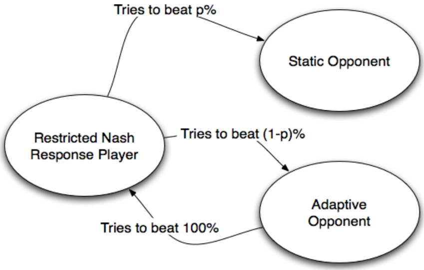
- "Restricted Nash Response": our opponent is *restricted* to playing the static strategy some of the time.
- We approach a Nash equilibrium in this restricted game.

Restricted Nash Response: Picking the Percentage



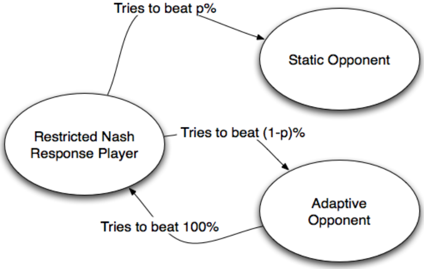
- In the last example, we said the opponent uses the static strategy 75% of the time

Restricted Nash Response: Picking the Percentage



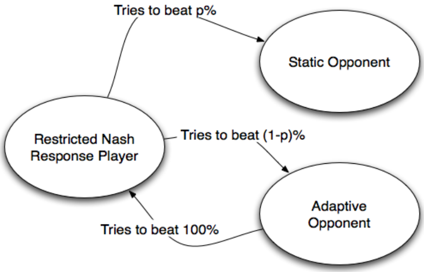
- In the last example, we said the opponent uses the static strategy 75% of the time
- This is actually just a variable, p.

Restricted Nash Response: Picking the Percentage



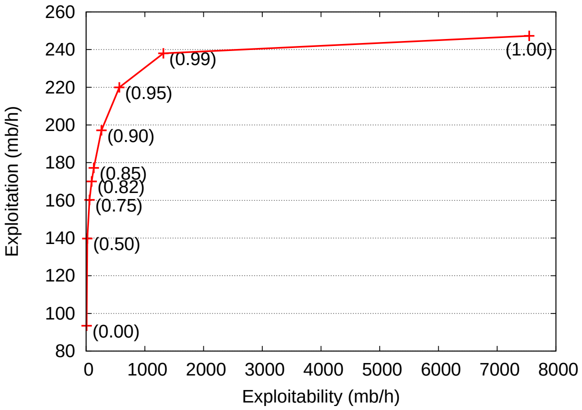
- In the last example, we said the opponent uses the static strategy 75% of the time
- This is actually just a variable, p.
- Interpretations of p:
  - How much you care about exploiting the static strategy
  - How confident you are that the opponent will actually use the static strategy

Restricted Nash Response: Picking the Percentage



- If p is low, then the resulting counter-strategy is more like a Nash equilibrium
- If p is high, then the resulting counter-strategy is more like a best response

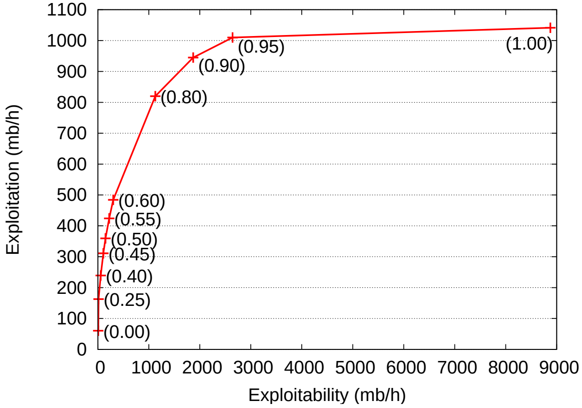
Restricted Nash Response: Picking the Percentage



PsOpti4

- X-Axis: How exploitable the counter-strategy is
- Y-Axis: How much we beat the opponent
- Labels: The value of p used to generate the strategy

Restricted Nash Response: Picking the Percentage



Attack80

- Don't use a Nash equilibrium - you can win a lot by giving up a tiny amount!
- Don't use a Best Response - you can save a lot by giving up a tiny amount!

Restricted Nash Response: Results

Frequentist Best Response:

	PsOpti4	PsOpti6	Attack60	Attack80	Smallbot1239	Smallbot1399	Smallbot2298	CFR5	Average
FBR-PsOpti4	137	-163	-227	-231	-106	-85	-144	-210	-129
FBR-PsOpti6	-79	330	-68	-89	-36	-23	-48	-97	-14
FBR-Attack60	-442	-499	2170	-701	-359	-305	-377	-620	-142
FBR-Attack80	-312	-281	-557	1048	-251	-231	-266	-331	-148
FBR-Smallbot1239	-20	105	-89	-42	106	91	-32	-87	3
FBR-Smallbot1399	-43	38	-48	-77	75	118	-46	-109	-11
FBR-Smallbot2298	-39	51	-50	-26	42	50	33	-41	2
CFR5	36	123	93	41	70	68	17	0	56
Max	137	330	2170	1048	106	118	33	0	

Table: Each entry is the result of a match between the row and the column player. Score is the average amount won by the row player, in millibets / hand. One millibet is 0.001 small bets.

Restricted Nash Response:

	Opponents								
	PsOpti4	PsOpti6	Attack60	Attack80	Smallbot1239	Smallbot1399	Smallbot2298	CFR5	Average
RNR-PsOpti4	85	112	39	9	63	61	-1	-23	43
RNR-PsOpti6	26	234	72	34	59	59	1	-28	57
RNR-Attack60	-17	63	582	-22	37	39	-9	-45	78

Restricted Nash Response: Conclusions

- “Playing to Win, Carefully”

Restricted Nash Response: Conclusions

- "Playing to Win, Carefully"
- Restricted Nash Response makes *robust* counter-strategies

Restricted Nash Response: Conclusions

- "Playing to Win, Carefully"
- Restricted Nash Response makes *robust* counter-strategies
- Exploits one opponent, minimizes weakness against all others

Restricted Nash Response: Conclusions

- "Playing to Win, Carefully"
- Restricted Nash Response makes *robust* counter-strategies
- Exploits one opponent, minimizes weakness against all others
- If you ever have to compute a best response offline, you can do this instead. It's not so bad if you're right, and a life saver if you're wrong.

Competition Results

- We competed in two competitions at AAAI this year:

Competition Results

- We competed in two competitions at AAAI this year:
- Second AAAI Computer Poker Competition
  - 3 events, 15 competitors, 43 bots
  - Used CFR strategies to get a 1st, a 2nd, and a 3rd

Competition Results

- We competed in two competitions at AAAI this year:
- Second AAAI Computer Poker Competition
  - 3 events, 15 competitors, 43 bots
  - Used CFR strategies to get a 1st, a 2nd, and a 3rd
- First Man-Machine Poker Championship
  - Played against two poker pros, Phil Laak and Ali Eslami
  - Used CFR and RNR strategies to win one, tie one, and lose two
  - Post-game analysis suggests a different result

3 new techniques for stochastic, imperfect information games:

- Counterfactual Regret Minimization
  - "Playing to Not Lose"
- Frequentist Best Response
  - "Playing to Win"
- Restricted Nash Response
  - "Playing to Win, Carefully"

3 new techniques for stochastic, imperfect information games:

- Counterfactual Regret Minimization
  - "Playing to Not Lose"
  - Approximate Nash Equilibrium strategies
  - Runs faster and with lower memory requirements than past techniques
- Frequentist Best Response
  - "Playing to Win"
- Restricted Nash Response
  - "Playing to Win, Carefully"

3 new techniques for stochastic, imperfect information games:

- Counterfactual Regret Minimization
  - "Playing to Not Lose"
  - Approximate Nash Equilibrium strategies
  - Runs faster and with lower memory requirements than past techniques
- Frequentist Best Response
  - "Playing to Win"
  - Finds exploitive counter-strategies for specific opponents
  - Useful for finding maximum exploitability of an opponent
  - Brittle when used against other opponents
- Restricted Nash Response
  - "Playing to Win, Carefully"

3 new techniques for stochastic, imperfect information games:

- Counterfactual Regret Minimization
  - "Playing to Not Lose"
  - Approximate Nash Equilibrium strategies
  - Runs faster and with lower memory requirements than past techniques
- Frequentist Best Response
  - "Playing to Win"
  - Finds exploitive counter-strategies for specific opponents
  - Useful for finding maximum exploitability of an opponent
  - Brittle when used against other opponents
- Restricted Nash Response
  - "Playing to Win, Carefully"
  - Finds *robust* counter-strategies for specific opponents
  - Useful for exploiting a suspected tendency
  - Robust when used against other opponents

3 new techniques for stochastic, imperfect information games:

- Counterfactual Regret Minimization
  - "Playing to Not Lose"
  - Approximate Nash Equilibrium strategies
  - Runs faster and with lower memory requirements than past techniques
- Frequentist Best Response
  - "Playing to Win"
  - Finds exploitive counter-strategies for specific opponents
  - Useful for finding maximum exploitability of an opponent
  - Brittle when used against other opponents
- Restricted Nash Response
  - "Playing to Win, Carefully"
  - Finds *robust* counter-strategies for specific opponents
  - Useful for exploiting a suspected tendency
  - Robust when used against other opponents
- We proved the value of these techniques through competitive play

Concluding Thoughts

- There's another Computer Poker Competition next year, and we're hoping for another Man-Machine match

Concluding Thoughts

- There's another Computer Poker Competition next year, and we're hoping for another Man-Machine match
- We have many directions to take this work
  - Better ways to manage a team of strategies
  - Counter-strategies that exploit a wide variety of opponents
  - ...and many other parts of the problem

Concluding Thoughts

- There's another Computer Poker Competition next year, and we're hoping for another Man-Machine match
- We have many directions to take this work
    - Better ways to manage a team of strategies
    - Counter-strategies that exploit a wide variety of opponents
    - ...and many other parts of the problem
- The CFR and RNR techniques described in this thesis are iterative
    - The longer you run the program, the better they get
    - Over the next year, we can produce much stronger poker programs
    - The quality of human play will not improve much this year

Concluding Thoughts

- There's another Computer Poker Competition next year, and we're hoping for another Man-Machine match
- We have many directions to take this work
  - Better ways to manage a team of strategies
  - Counter-strategies that exploit a wide variety of opponents
  - ...and many other parts of the problem
- The CFR and RNR techniques described in this thesis are iterative
  - The longer you run the program, the better they get
  - Over the next year, we can produce much stronger poker programs
  - The quality of human play will not improve much this year
- The next Man-Machine match might have a different outcome!

Questions?






AAAI Computer Poker Competition

- Second year it's been run

AAAI Computer Poker Competition

- Second year it's been run
- Last year: 2 events, 5 competitors, 5 bots
- This year: 3 events, 15 competitors, 43 bots

AAAI Computer Poker Competition

- Second year it's been run
- Last year: 2 events, 5 competitors, 5 bots
- This year: 3 events, 15 competitors, 43 bots
- 3 Events:
  - Heads-Up Limit Equilibrium
  - Heads-Up Limit Online Learning
  - Heads-Up No-Limit

AAAI: Heads-Up Limit Equilibrium

- Winner determined by total matches (not dollars!) won

	Hyperborean07EQ	IanBot	GS3	PokeMinn	Quick	Gomel-2	DumboEQ	DumboEQ-2	Sequel	Sequel-2	PokeMinn-2	UNCC	Gomel	LeRenard	MonashBPP	MilanoEQ	Average
Hyperborean07EQ		21	32	136	115	110	193	182	165	166	131	454	115	138	465	428	194
IanBot	-21		4	130	99	85	142	119	131	140	142	472	88	130	408	398	164
GS3	-32	-4		150	73	112	160	149	140	148	154	467	107	142	412	445	175
PokeMinn	-136	-130	-150		40	144	80	76	-33	-22	-24	373	265	127	627	421	111
Quick	-115	-99	-73	-40		19	235	135	125	121	134	298	149	15	564	489	131
Gomel-2	-110	-85	-112	-144	-19		206	200	135	150	16	275	232	136	802	859	169
DumboEQ	-193	-142	-160	-80	-235	-206		133	67	64	55	23	300	13	774	672	72
DumboEQ-2	-182	-119	-149	-76	-135	-200	-133		87	82	83	-52	271	54	808	762	74
Sequel	-165	-131	-140	33	-125	-135	-67	-87		19	130	167	-17	92	556	556	46
Sequel-2	-166	-140	-148	22	-121	-150	-64	-82	-19		125	174	-4	74	583	526	41
PokeMinn-2	-131	-142	-154	24	-134	-16	-55	-83	-130	-125		96	123	60	770	748	57
UNCC	-454	-472	-467	-373	-298	-275	-23	52	-167	-174	-96		95	-281	553	503	-125
Gomel	-115	-88	-107	-265	-149	-232	-300	-271	17	4	-123	-95		96	779	993	10
LeRenard	-138	-130	-142	-127	-15	-136	-13	-54	-92	-74	-60	281	-96		478	354	2
MonashBPP	-465	-408	-412	-627	-564	-802	-774	-808	-556	-583	-770	-553	-779	-478		489	-539
MilanoEQ	-428	-398	-445	-421	-489	-859	-672	-762	-556	-526	-748	-503	-993	-354	-489		-576

AAAI: Heads-Up Limit Equilibrium

- Winner determined by total matches (not dollars!) won
- Emphasizes winning, not exploiting

	Hyperborean07EQ	IanBot	GS3	PokeMinn	Quick	Gomel-2	DumboEQ	DumboEQ-2	Sequel	Sequel-2	PokeMinn-2	UNCC	Gomel	LeRenard	MonashBPP	MilanoEQ	Average
Hyperborean07EQ		21	32	136	115	110	193	182	165	166	131	454	115	138	465	428	194
IanBot	-21		4	130	99	85	142	119	131	140	142	472	88	130	408	398	164
GS3	-32	-4		150	73	112	160	149	140	148	154	467	107	142	412	445	175
PokeMinn	-136	-130	-150		40	144	80	76	-33	-22	-24	373	265	127	627	421	111
Quick	-115	-99	-73	-40		19	235	135	125	121	134	298	149	15	564	489	131
Gomel-2	-110	-85	-112	-144	-19		206	200	135	150	16	275	232	136	802	859	169
DumboEQ	-193	-142	-160	-80	-235	-206		133	67	64	55	23	300	13	774	672	72
DumboEQ-2	-182	-119	-149	-76	-135	-200	-133		87	82	83	-52	271	54	808	762	74
Sequel	-165	-131	-140	33	-125	-135	-67	-87		19	130	167	-17	92	556	556	46
Sequel-2	-166	-140	-148	22	-121	-150	-64	-82	-19		125	174	-4	74	583	526	41
PokeMinn-2	-131	-142	-154	24	-134	-16	-55	-83	-130	-125		96	123	60	770	748	57
UNCC	-454	-472	-467	-373	-298	-275	-23	52	-167	-174	-96		95	-281	553	503	-125
Gomel	-115	-88	-107	-265	-149	-232	-300	-271	17	4	-123	-95		96	779	993	10
LeRenard	-138	-130	-142	-127	-15	-136	-13	-54	-92	-74	-60	281	-96		478	354	2
MonashBPP	-465	-408	-412	-627	-564	-802	-774	-808	-556	-583	-770	-553	-779	-478		489	-539
MilanoEQ	-428	-398	-445	-421	-489	-859	-672	-762	-556	-526	-748	-503	-993	-354	-489		-576

AAAI: Heads-Up Limit Equilibrium

- Winner determined by total matches (not dollars!) won
- Emphasizes winning, not exploiting
- Took first place, using a CFR bot

	Hyperborean07EQ	IanBot	GS3	PokeMinn	Quick	Gomel-2	DumboEQ	DumboEQ-2	Sequel	Sequel-2	PokeMinn-2	UNCC	Gomel	LeRenard	MonashBPP	MilanoEQ	Average
Hyperborean07EQ		21	32	136	115	110	193	182	165	166	131	454	115	138	465	428	194
IanBot	-21		4	130	99	85	142	119	131	140	142	472	88	130	408	398	164
GS3	-32	-4		150	73	112	160	149	140	148	154	467	107	142	412	445	175
PokeMinn	-136	-130	-150		40	144	80	76	-33	-22	-24	373	265	127	627	421	111
Quick	-115	-99	-73	-40		19	235	135	125	121	134	298	149	15	564	489	131
Gomel-2	-110	-85	-112	-144	-19		206	200	135	150	16	275	232	136	802	859	169
DumboEQ	-193	-142	-160	-80	-235	-206		133	67	64	55	23	300	13	774	672	72
DumboEQ-2	-182	-119	-149	-76	-135	-200	-133		87	82	83	-52	271	54	808	762	74
Sequel	-165	-131	-140	33	-125	-135	-67	-87		19	130	167	-17	92	556	556	46
Sequel-2	-166	-140	-148	22	-121	-150	-64	-82	-19		125	174	-4	74	583	526	41
PokeMinn-2	-131	-142	-154	24	-134	-16	-55	-83	-130	-125		96	123	60	770	748	57
UNCC	-454	-472	-467	-373	-298	-275	-23	52	-167	-174	-96		95	-281	553	503	-125
Gomel	-115	-88	-107	-265	-149	-232	-300	-271	17	4	-123	-95		96	779	993	10
LeRenard	-138	-130	-142	-127	-15	-136	-13	-54	-92	-74	-60	281	-96		478	354	2
MonashBPP	-465	-408	-412	-627	-564	-802	-774	-808	-556	-583	-770	-553	-779	-478		489	-539
MilanoEQ	-428	-398	-445	-421	-489	-859	-672	-762	-556	-526	-748	-503	-993	-354	-489		-576

AAAI: Heads-Up Limit Online Learning

- Winner determined by total winnings (in dollars)

	Hyperborean07OL-2	Hyperborean07OL	GS3	IanBot	Quick	Gomel-2	PokeMinn	Sequel	Sequel-2	LeRenard	DumboOL-2	Average
Hyperborean07OL-2		-37	-27	-37	138	155	172	166	178	170	259	114
Hyperborean07OL	37		21	27	116	108	141	153	175	132	207	112
GS3	27	-21		6	73	112	150	140	148	142	199	98
IanBot	37	-27	-6		99	85	130	131	140	130	157	87
Quick	-138	-116	-73	-99		19	-40	125	121	15	129	-6
Gomel-2	-155	-108	-112	-85	-19		-144	135	150	136	123	-8
PokeMinn	-172	-141	-150	-130	40	144		-33	-22	127	-15	-35
Sequel	-166	-153	-140	-131	-125	-135	33		19	92	-1	-71
Sequel-2	-178	-175	-148	-140	-121	-150	22	-19		74	17	-82
LeRenard	-170	-132	-142	-130	-15	-136	-127	-92	-74		21	-100
DumboOL-2	-259	-207	-199	-157	-129	-123	15	1	-17	-21		-110

AAAI: Heads-Up Limit Online Learning

- Winner determined by total winnings (in dollars)
- Took second place with a CFR bot. We just barely lost to...

	Hyperborean07OL-2	Hyperborean07OL	GS3	IanBot	Quick	Gomel-2	PokeMinn	Sequel	Sequel-2	LeRenard	DumboOL-2	Average
Hyperborean07OL-2		-37	-27	-37	138	155	172	166	178	170	259	114
Hyperborean07OL	37		21	27	116	108	141	153	175	132	207	112
GS3	27	-21		6	73	112	150	140	148	142	199	98
IanBot	37	-27	-6		99	85	130	131	140	130	157	87
Quick	-138	-116	-73	-99		19	-40	125	121	15	129	-6
Gomel-2	-155	-108	-112	-85	-19		-144	135	150	136	123	-8
PokeMinn	-172	-141	-150	-130	40	144		-33	-22	127	-15	-35
Sequel	-166	-153	-140	-131	-125	-135	33		19	92	-1	-71
Sequel-2	-178	-175	-148	-140	-121	-150	22	-19		74	17	-82
LeRenard	-170	-132	-142	-130	-15	-136	-127	-92	-74		21	-100
DumboOL-2	-259	-207	-199	-157	-129	-123	15	1	-17	-21		-110

AAAI: Heads-Up Limit Online Learning

- Winner determined by total winnings (in dollars)
- Took second place with a CFR bot. We just barely lost to...
- ...the other U of A bot

	Hyperborean07OL-2	Hyperborean07OL	GS3	IanBot	Quick	Gomel-2	PokeMinn	Sequel	Sequel-2	LeRenard	DumboOL-2	Average
Hyperborean07OL-2		-37	-27	-37	138	155	172	166	178	170	259	114
Hyperborean07OL	37		21	27	116	108	141	153	175	132	207	112
GS3	27	-21		6	73	112	150	140	148	142	199	98
IanBot	37	-27	-6		99	85	130	131	140	130	157	87
Quick	-138	-116	-73	-99		19	-40	125	121	15	129	-6
Gomel-2	-155	-108	-112	-85	-19		-144	135	150	136	123	-8
PokeMinn	-172	-141	-150	-130	40	144		-33	-22	127	-15	-35
Sequel	-166	-153	-140	-131	-125	-135	33		19	92	-1	-71
Sequel-2	-178	-175	-148	-140	-121	-150	22	-19		74	17	-82
LeRenard	-170	-132	-142	-130	-15	-136	-127	-92	-74		21	-100
DumboOL-2	-259	-207	-199	-157	-129	-123	15	1	-17	-21		-110

AAAI: Heads-Up Limit Online Learning

- Winner determined by total winnings (in dollars)
- Took second place with a CFR bot. We just barely lost to...
- ...the other U of A bot
- (Darse Billings and Morgan Kan)

	Hyperborean07OL-2	Hyperborean07OL	GS3	IanBot	Quick	Gomel-2	PokeMinn	Sequel	Sequel-2	LeRenard	DumboOL-2	Average
Hyperborean07OL-2		-37	-27	-37	138	155	172	166	178	170	259	114
Hyperborean07OL	37		21	27	116	108	141	153	175	132	207	112
GS3	27	-21		6	73	112	150	140	148	142	199	98
IanBot	37	-27	-6		99	85	130	131	140	130	157	87
Quick	-138	-116	-73	-99		19	-40	125	121	15	129	-6
Gomel-2	-155	-108	-112	-85	-19		-144	135	150	136	123	-8
PokeMinn	-172	-141	-150	-130	40	144		-33	-22	127	-15	-35
Sequel	-166	-153	-140	-131	-125	-135	33		19	92	-1	-71
Sequel-2	-178	-175	-148	-140	-121	-150	22	-19		74	17	-82
LeRenard	-170	-132	-142	-130	-15	-136	-127	-92	-74		21	-100
DumboOL-2	-259	-207	-199	-157	-129	-123	15	1	-17	-21		-110

AAAI: Heads-Up No-Limit

- No-Limit is what you see on TV - bets can be any size

	BluffBot20	GS3	Hyperborean07	SlideRule	Gomel	Gomel-2	Milano	Manitoba	PokeMinn	Manitoba-2	Average
BluffBot20		267	380	576	2093	2885	3437	475	1848	2471	1603
GS3	-267		113	503	3161	124	1875	4204	-42055	5016	-3036
Hyperborean07	-380	-113		-48	6657	5455	6795	8697	12051	22116	6803
SlideRule	-576	-503	48		11596	9730	10337	10387	15637	10791	7494
Gomel	-2093	-3161	-6657	-11596		3184	8372	11450	62389	52325	12690
Gomel-2	-2885	-124	-5455	-9730	-3184		15078	11907	58985	40256	11650
Milano	-3437	-1875	-6795	-10337	-8372	-15078		5741	12719	27040	-44
Manitoba	-475	-4204	-8697	-10387	-11450	-11907	-5741		18817	50677	1848
PokeMinn	-1848	42055	-14051	-15637	-62389	-58985	-12719	-18817		34299	-12010
Manitoba-2	-2471	-5016	-22116	-10791	-52325	-40256	-27040	-50677	-34299		-27221

AAAI: Heads-Up No-Limit

- No-Limit is what you see on TV - bets can be any size
- This was our first time making a No-Limit bot

	BluffBot20	GS3	Hyperborean07	SlideRule	Gomel	Gomel-2	Milano	Manitoba	PokeMinn	Manitoba-2	Average
BluffBot20		267	380	576	2093	2885	3437	475	1848	2471	1603
GS3	-267		113	503	3161	124	1875	4204	-42055	5016	-3036
Hyperborean07	-380	-113		-48	6657	5455	6795	8697	12051	22116	6803
SlideRule	-576	-503	48		11596	9730	10337	10387	15637	10791	7494
Gomel	-2093	-3161	-6657	-11596		3184	8372	11450	62389	52325	12690
Gomel-2	-2885	-124	-5455	-9730	-3184		15078	11907	58985	40256	11650
Milano	-3437	-1875	-6795	-10337	-8372	-15078		5741	12719	27040	-44
Manitoba	-475	-4204	-8697	-10387	-11450	-11907	-5741		18817	50677	1848
PokeMinn	-1848	42055	-14051	-15637	-62389	-58985	-12719	-18817		34299	-12010
Manitoba-2	-2471	-5016	-22116	-10791	-52325	-40256	-27040	-50677	-34299		-27221

AAAI: Heads-Up No-Limit

- No-Limit is what you see on TV - bets can be any size
- This was our first time making a No-Limit bot
- Took third place, using a CFR bot with abstracted betting

	BluffBot20	GS3	Hyperborean07	SlideRule	Gomel	Gomel-2	Milano	Manitoba	PokeMinn	Manitoba-2	Average
BluffBot20		267	380	576	2093	2885	3437	475	1848	2471	1603
GS3	-267		113	503	3161	124	1875	4204	-42055	5016	-3036
Hyperborean07	-380	-113		-48	6657	5455	6795	8697	12051	22116	6803
SlideRule	-576	-503	48		11596	9730	10337	10387	15637	10791	7494
Gomel	-2093	-3161	-6657	-11596		3184	8372	11450	62389	52325	12690
Gomel-2	-2885	-124	-5455	-9730	-3184		15078	11907	58985	40256	11650
Milano	-3437	-1875	-6795	-10337	-8372	-15078		5741	12719	27040	-44
Manitoba	-475	-4204	-8697	-10387	-11450	-11907	-5741		18817	50677	1848
PokeMinn	-1848	42055	-14051	-15637	-62389	-58985	-12719	-18817		34299	-12010
Manitoba-2	-2471	-5016	-22116	-10791	-52325	-40256	-27040	-50677	-34299		-27221

AAAI: Heads-Up No-Limit

- No-Limit is what you see on TV - bets can be any size
- This was our first time making a No-Limit bot
- Took third place, using a CFR bot with abstracted betting
- We hope to do better next year! Lots of exciting work to be done here.

	BluffBot20	GS3	Hyperborean07	SlideRule	Gomel	Gomel-2	Milano	Manitoba	PokeMinn	Manitoba-2	Average
BluffBot20		267	380	576	2093	2885	3437	475	1848	2471	1603
GS3	-267		113	503	3161	124	1875	4204	-42055	5016	-3036
Hyperborean07	-380	-113		-48	6657	5455	6795	8697	12051	22116	6803
SlideRule	-576	-503	48		11596	9730	10337	10387	15637	10791	7494
Gomel	-2093	-3161	-6657	-11596		3184	8372	11450	62389	52325	12690
Gomel-2	-2885	-124	-5455	-9730	-3184		15078	11907	58985	40256	11650
Milano	-3437	-1875	-6795	-10337	-8372	-15078		5741	12719	27040	-44
Manitoba	-475	-4204	-8697	-10387	-11450	-11907	-5741		18817	50677	1848
PokeMinn	-1848	42055	-14051	-15637	-62389	-58985	-12719	-18817		34299	-12010
Manitoba-2	-2471	-5016	-22116	-10791	-52325	-40256	-27040	-50677	-34299		-27221

First Man-Machine Poker Championship

- Beating human experts is a big milestone

First Man-Machine Poker Championship

- Beating human experts is a big milestone
- Tough to get statistical significance against humans

First Man-Machine Poker Championship

- Beating human experts is a big milestone
- Tough to get statistical significance against humans
- So we played two at once with the same cards

First Man-Machine Poker Championship

- Beating human experts is a big milestone
- Tough to get statistical significance against humans
- So we played two at once with the same cards
- Four matches of 500 hands each

First Man-Machine Poker Championship

- Beating human experts is a big milestone
- Tough to get statistical significance against humans
- So we played two at once with the same cards
- Four matches of 500 hands each
- Have to be ahead by 25 small bets to win a match

Phil Laak

- Background: Mechanical Engineer
- Started gambling in competitive backgammon
- Competes in the world's biggest poker tournaments

Ali Eslami

- Background: Computer consultant
- Started out by playing...
- Plays in $1000-$2000 Limit games

Ali Eslami

- Background: Computer consultant
- Started out by playing... Magic: The Gathering
- Plays in $1000-$2000 Limit games

Ali Eslami

- Background: Computer consultant
- Started out by playing... Magic: The Gathering
- Plays in $1000-$2000 Limit games
- (This is a lot of money!)

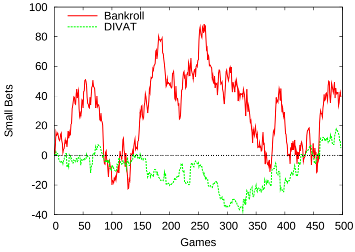
Day 1, Session 1

- We had 10 different bots to use:
  - Several Counterfactual Regret Minimization approximate Nash equilibria
  - Flavours of Restricted Nash Response counter-strategies

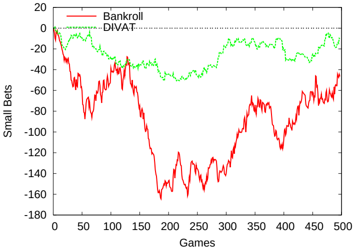
Day 1, Session 1

- We had 10 different bots to use:
  - Several Counterfactual Regret Minimization approximate Nash equilibria
  - Flavours of Restricted Nash Response counter-strategies
- We wanted a baseline to compare future bots against
- Bot used: Mr. Pink, our finest abstraction CFR approximate Nash equilibrium

Day 1, Session 1

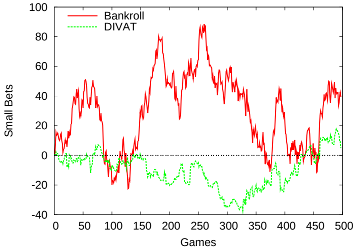


On Stage: Ali Eslami

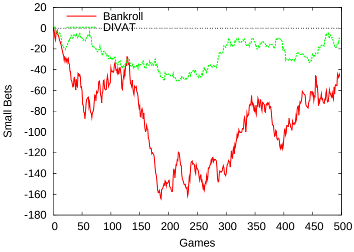


Hotel: Phil Laak

Day 1, Session 1



On Stage: Ali Eslami



Hotel: Phil Laak

- Ali: $395
- Phil: -$465
- Polaris ends ahead by $70
- Result: Tie

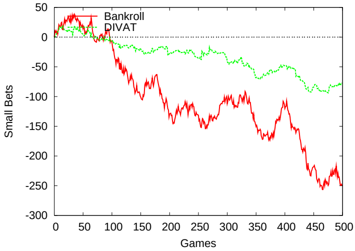
Day 1, Session 2

- Score so far: 1 Tie
- The careful choice (Mr. Pink) did OK, so lets try something crazy!

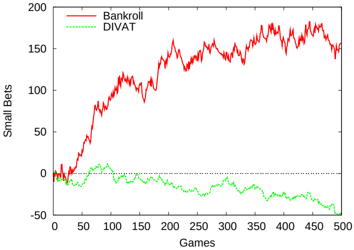
Day 1, Session 2

- Score so far: 1 Tie
- The careful choice (Mr. Pink) did OK, so lets try something crazy!
- Bot used: Mr. Orange / Crazy 8s
- It's a CFR approximate Nash equilibrium in a broken game that encourages aggression

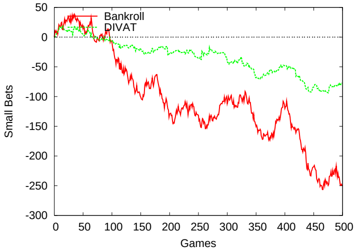
Day 1, Session 2



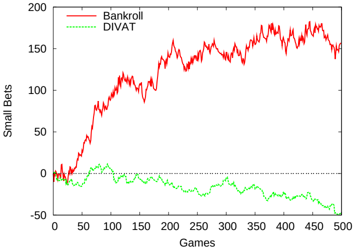
Hotel: Ali Eslami



On Stage: Phil Laak



Hotel: Ali Eslami



On Stage: Phil Laak

- Ali:-$2495
- Phil: $1570
- Polaris ends ahead by $925
- Result: Win

Day 2, Session 1

- Score so far: 1 Win, 1 Tie
- Which of our 10 bots to use this time?

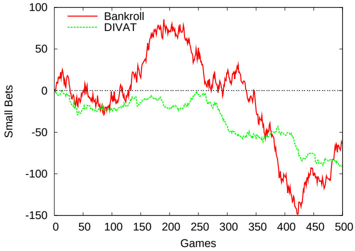
Day 2, Session 1

- Score so far: 1 Win, 1 Tie
- Which of our 10 bots to use this time?
- We pulled an all nighter and ran importance sampling on the last 1000 hands
- Predicted the best 3 bots to use against each player

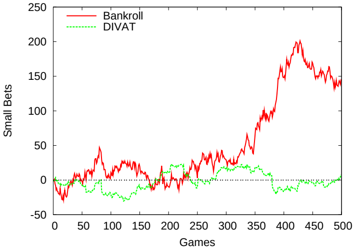
Day 2, Session 1

- Score so far: 1 Win, 1 Tie
- Which of our 10 bots to use this time?
- We pulled an all nighter and ran importance sampling on the last 1000 hands
- Predicted the best 3 bots to use against each player
- Used a coach that chose between these 3 during the match

Day 2, Session 1

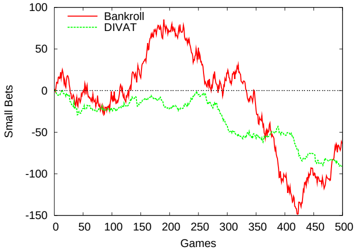


Hotel: Ali Eslami

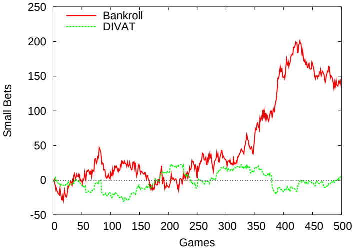


On Stage: Phil Laak

Day 2, Session 1



Hotel: Ali Eslami



On Stage: Phil Laak

- Ali:-$635
- Phil: $1455
- Polaris ends behind by $820
- Result: Loss

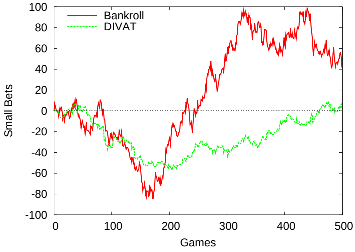
Day 2, Session 2

- Score so far: 1 Win, 1 Tie, 1 Loss
- Decided to play it safe and go for a tie

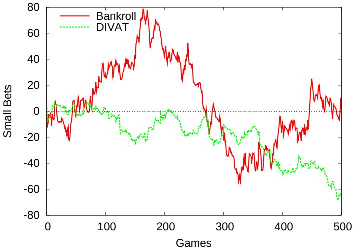
Day 2, Session 2

- Score so far: 1 Win, 1 Tie, 1 Loss
- Decided to play it safe and go for a tie
- Bot used: Mr. Pink, the approximate Nash equilibrium from the first match

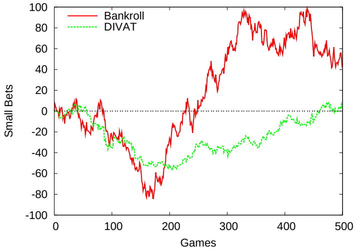
Day 2, Session 2



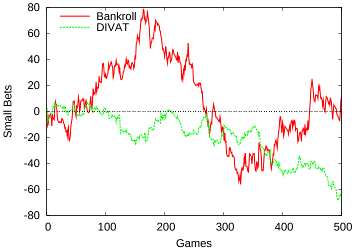
Onstage: Ali Eslami



Hotel: Phil Laak



Onstage: Ali Eslami



Hotel: Phil Laak

- Ali: $460
- Phil: $110
- Polaris ends behind by $570
- Result: Loss

Man-Machine Match Conclusions

- Very close game — we lost by 0.01 small bets/game, less than the tie margin

Man-Machine Match Conclusions

- Very close game — we lost by 0.01 small bets/game, less than the tie margin
- Ali: "This was not a win for us...I played the best heads-up poker I've ever played...we just barely won"

Man-Machine Match Conclusions

- Very close game — we lost by 0.01 small bets/game, less than the tie margin
- Ali: "This was not a win for us...I played the best heads-up poker I've ever played...we just barely won"
- Post-game analysis (DIVAT) suggests that we outplayed them

Man-Machine Match Conclusions

- Very close game — we lost by 0.01 small bets/game, less than the tie margin
- Ali: "This was not a win for us...I played the best heads-up poker I've ever played...we just barely won"
- Post-game analysis (DIVAT) suggests that we outplayed them
- We'd like to do another match next year

Man-Machine Match Conclusions

- Very close game — we lost by 0.01 small bets/game, less than the tie margin
- Ali: "This was not a win for us...I played the best heads-up poker I've ever played...we just barely won"
- Post-game analysis (DIVAT) suggests that we outplayed them
- We'd like to do another match next year
- There's lots of exciting work to do here, too!