# Bayesian Opponent Exploitation in Imperfect-Information Games

**Sam Ganzfried**

**Assistant Professor**

**Florida International University, Miami, FL**

**School of Computing and Information Sciences**

http://www.ganzfriedresearch.com/

sam.ganzfried@gmail.com

# Constructing an opponent model

E.g., if opponent has played Rock 10 times Paper 7 times Scissors 3 times, can predict he will play R with prob 10/20, etc.

In imperfect-information games more challenging but doable to approximate

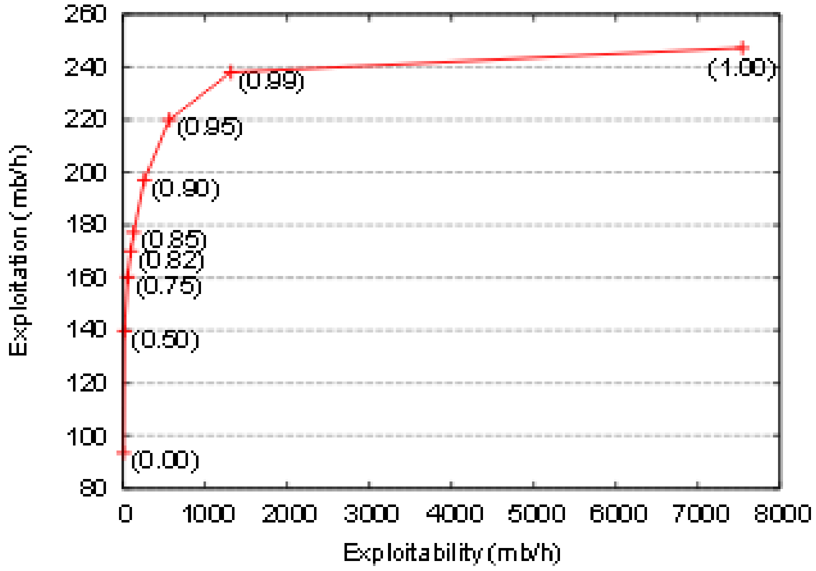
- e.g., Ganzfried/Sandholm AAMAS 2011

But is it really valid to assign a single “model”? What if he isn’t following that exact strategy?

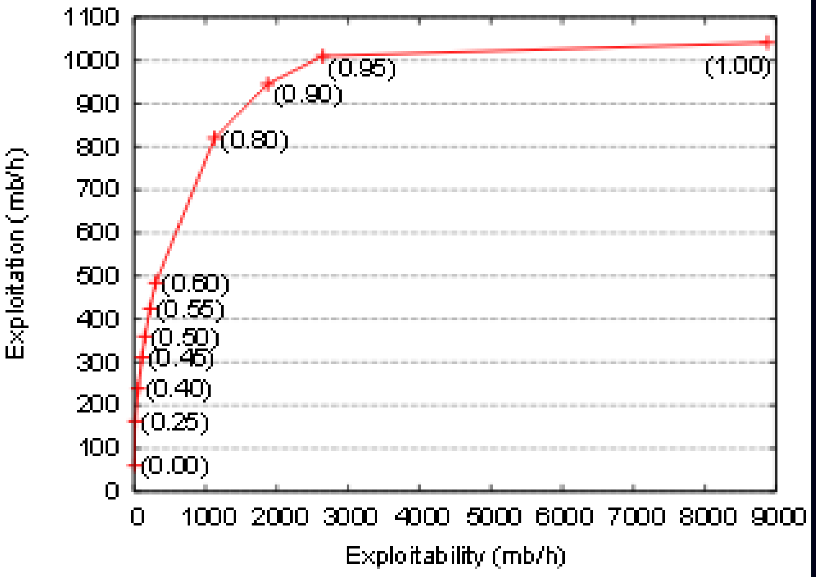
- Maybe he is playing R with prob 0.49!!

# Restricted Nash Response

## Johanson, Zinkevich, Bowling NIPS 2007



(a) Versus PsOpti4



(b) Versus A80

- Suppose opponent is playing $\sigma_{-i}$, where $\sigma_{-i}(s_{-j})$ is probability that he plays pure strategy $s_{-j}$ in $S_{-j}$

  $u_i(\sigma_i, \sigma_{-i}) = \sum_{s-j} [\sigma_{-i}(s_{-j}) * u_i(\sigma_i, s_{-j})]$

- Now suppose opponent is playing a probability distribution $f_{-i}$ over *mixed strategies*

  $u_i(\sigma_i, f_{-i}) = \int_{\sigma-i} [f_{-i}(\sigma_{-i}) * u_i(\sigma_i, \sigma_{-i})]$

- Let $f^*_{-i}$ denote the mean of $f_{-i}$. Selects $s_{-j}$ with prob

  $\int_{\sigma-i} [\sigma_{-i}(s_{-j}) * f_{-i}(\sigma_{-i})]$

Theorem: $u_i(\sigma_i, f^*_{-i}) = u_i(\sigma_i, f_{-i})$

Proof:

$$
\begin{aligned}
u_i(\sigma_i, f^*_{-i}) &= \textstyle\sum_{s\text{-}j} [u_i(\sigma_i, s_{-j}) \int_{\sigma\text{-}i} [\sigma_{-i}(s_{-j}) * f_{-i}(\sigma_{-i})]] \\
&= \textstyle\sum_{s\text{-}j} [\int_{\sigma\text{-}i} [u_i(\sigma_i, s_{-j}) * \sigma_{-i}(s_{-j}) * f_{-i}(\sigma_{-i})]] \\
&= \int_{\sigma\text{-}i} [\textstyle\sum_{s\text{-}j} [u_i(\sigma_i, s_{-j}) * \sigma_{-i}(s_{-j}) * f_{-i}(\sigma_{-i})]] \\
&= \int_{\sigma\text{-}i} [u_i(\sigma_i, \sigma_{-i}) * f_{-i}(\sigma_{-i})] \\
&= u_i(\sigma_i, f_{-i})
\end{aligned}
$$

Corollary: $u_i(\sigma_i, p^*(\sigma_{-i}|x)) = u_i(\sigma_i, p(\sigma_{-i}|x))$

- $p(\sigma_{-i})$ denotes prior (probability distribution over mixed strategies) and $p(\sigma_{-i}|x)$ denote posterior given some observations x
- $p^*(\sigma_{-i}|x)$ is mean of $p(\sigma_{-i}|x)$

- Theorem and corollary apply to normal-form and extensive-form (both perfect and imperfect information) for any number of players (can let $\sigma_{-i}$ be joint strategy profile for all other agents)

# Meta-algorithm for Bayesian opponent exploitation

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
  $x_t \leftarrow$ observations of opponent's play at time step $t$
  $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$
  $M_t \leftarrow$ expectation of $p_t$
  $R_t \leftarrow r_t(M_t)$
  Play according to $R_t$

# Challenges

- #1: Assumes we can compactly represent prior and posterior distributions $p_t$, which have infinite domain

---

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

---

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$

$R_0 \leftarrow r_0(M_0)$

Play according to $R_0$

**for** $t = 1$ to $T$ **do**

 $x_t \leftarrow$ observations of opponent's play at time step $t$

 $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$

 $M_t \leftarrow$ expectation of $p_t$

 $R_t \leftarrow r_t(M_t)$

 Play according to $R_t$

---

# Challenge #2

- Requires procedure to efficiently compute posterior distributions given prior and observations, which will involve having to update potentially infinitely-many strategies

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \le t \le T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
  $x_t \leftarrow$ observations of opponent's play at time step $t$
  $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$
  $M_t \leftarrow$ expectation of $p_t$
  $R_t \leftarrow r_t(M_t)$
  Play according to $R_t$

# #3

Requires efficient procedure to compute mean of $p_t$

---

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

---

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
  $x_t \leftarrow$ observations of opponent's play at time step $t$
  $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$
  $M_t \leftarrow$ expectation of $p_t$
  $R_t \leftarrow r_t(M_t)$
  Play according to $R_t$

---

# #4

Requires that the full posterior distribution from one round be compactly represented to be used as the prior distribution in the next round

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
  $x_t \leftarrow$ observations of opponent's play at time step $t$
  $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$
  $M_t \leftarrow$ expectation of $p_t$
  $R_t \leftarrow r_t(M_t)$
  Play according to $R_t$

Can solve #4 by using the following modification:

$p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_0$ and observations $x_1, \ldots, x_t$

**Algorithm 1** Meta-algorithm for Bayesian opponent exploitation

**Inputs:** Prior distribution $p_0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow \overline{p_0(\sigma_{-i})}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
  $x_t \leftarrow$ observations of opponent's play at time step $t$
  $p_t \leftarrow$ posterior distribution of opponent's strategy given prior $p_{t-1}$ and observations $x_t$
  $M_t \leftarrow$ expectation of $p_t$
  $R_t \leftarrow r_t(M_t)$
  Play according to $R_t$

# Dirichlet distribution

- pdf of the Dirichlet distribution returns the belief that the probabilities of K rival events are $x_i$ given that each event has been observed $\alpha_i$ - 1 times:
  - $f(x, \alpha) = [\prod xi^{\alpha i-1}] / B(\alpha)$
- Normalization B(α) is beta function
  - $B(\alpha) = \prod_i \Gamma(\alpha_i)/\Gamma(\sum_i \alpha_i)$, where $\Gamma(n) = (n-1)!$ is Gamma function
- $E[x_i] = \alpha_i / \sum_k \alpha_k$
- Assuming multinomial sampling, the posterior distribution after including new observations is also a Dirichlet distribution with parameters updated based on the new observations.

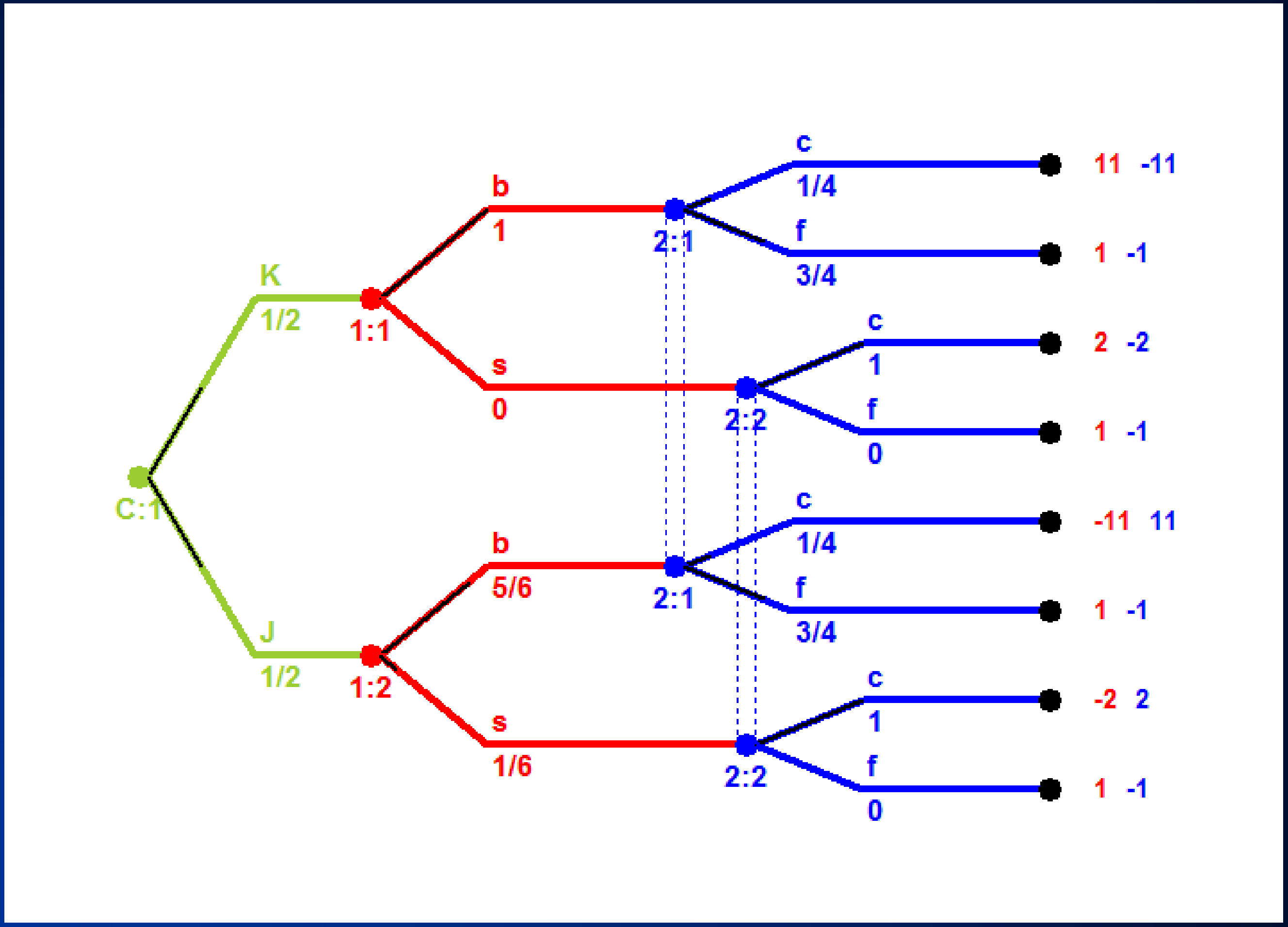
# Dirichlet distribution

- Very natural distribution, has been previously used for modeling in large imperfect-information games
- Dirichlet is conjugate prior for multinomial distribution, and therefore posterior is also Dirichlet
  - Opponent plays in proportion to updated weights
- So simple closed form for mean of posterior
  - Alg 1 gives exact efficient algorithm for computing Bayesian Best Response [Fudenberg/Levine ‘98]
  - “Fictitious play” [Brown ‘51]
- This applies to normal-form games and extensive-form games with perfect information
  - Zero-sum, general-sum, and any number of players

# Imperfect information

- It would also apply to imperfect-information games if the opponent's private information was observed after each round (so we knew exactly what information set he took observed action from)
- But not to imperfect-information games where opponent's private information is not (or is only sometimes) observed.
- Algorithm exists using importance sampling to approximate value of infinite integral [Southey et.al UAI '05]
  - Has been applied to limit Texas hold 'em successfully
  - But has no guarantees, and does not provide much intuition

- P1 given private information state $x_i$ according to distribution.
- P1 takes publicly observable action $a_i$.
- P2 observes $a_i$ but not $x_i$. Then P2 acts and players get payoff.



- If we observe opponent’s hand after each play, we could just maintain counter for each action/info set and update appropriate one
- But if we don’t observe his card, we wouldn’t know which counter to increment

- To simplify analysis assume we never see opponent’s card after a hand (and also assume we don’t observe our payoff until the end so that we could not draw inferences about his card).
- This is not realistic, but no known exact algorithms even for this simplified setting
  - Suspect approach can extend straightforwardly to case of partial observability

- Let $\alpha_{Kb}$ -1denote number of “fictitious” times we have observed opponent play b with K according to our prior
- Now assume we observe him take action b, but don’t observe his card

- Mean of posterior for probability he bets big with J:
- $[B(\alpha_{Kb}+1, \alpha_{Ks})B(\alpha_{Jb}+1, \alpha_{Js}) + B(\alpha_{Kb}, \alpha_{Ks})B(\alpha_{Jb}+2, \alpha_{Js})]/Z$

- $Z = B(\alpha_{Kb}+1, \alpha_{Ks})B(\alpha_{Jb}+1, \alpha_{Js}) + B(\alpha_{Kb}, \alpha_{Ks})B(\alpha_{Jb}+2, \alpha_{Js})$
- $\quad + B(\alpha_{Kb}+1, \alpha_{Ks})B(\alpha_{Jb}, \alpha_{Js}+1) + B(\alpha_{Kb}, \alpha_{Ks})B(\alpha_{Jb}+1, \alpha_{Js}+1)$

- Recall $B(\alpha) = \prod_i \Gamma(\alpha_i)/\Gamma(\sum_i \alpha_i)$, where $\Gamma(n) = (n-1)!$ is Gamma function

# General solution

- Assume we observe him play b $\theta_b$ times and s $\theta_s$ times
- Mean of posterior of probability of betting big with Jack:
- $\sum_i \sum_j B(\alpha_{Kb}+i, \alpha_{Ks}+j)\ B(\alpha_{Jb}+\theta_b - i+1, \alpha_{Js}+\theta_s - j) / Z$
- $Z = \sum_i \sum_j [B(\alpha_{Kb}+i, \alpha_{Ks}+j)\ B(\alpha_{Jb}+\theta_b - i+1, \alpha_{Js}+\theta_s - j) + B(\alpha_{Kb}+i, \alpha_{Ks}+j)\ B(\alpha_{Jb}+\theta_b - i, \alpha_{Js} + \theta_s - j+1)]$

# Example

- Suppose prior is that opponent played b with K 10 times, played s with K 3 times, played b with J 4 times, played s with J 9 times.
- Now suppose we see him play b at next iteration
- Previously we thought probability of betting big with a jack was 4/13 = 0.308
- Now: $p(b|O,J) = B(11,3)B(5,9) + B(10,3)(6,9)/Z$
- $p(s|O,J) = B(11,3)B(4,10) + B(10,3)(5,10)/Z$
- -> $p(b|O,J) = p(b|O,J)/[p(b|O,J)+p(s|O,J)] = \ldots$

- $p(b|O,J) = 0.322$
- Previously we thought probability of betting with a jack was $4/13 = 0.308$

- What if we observed his card after game play and observed he had a jack?

- $p(b|O,J) = 0.322$
- Previously we thought probability of betting with a jack was $4/13 = 0.308$

- What if we always observed his card after game play and observed he had a jack?
  - $5/14 = 0.357$

- What about “naïve” approach where we increment counter for $\alpha_{Jb}$ by $\alpha_{Jb}/(\alpha_{Jb} + \alpha_{Kb})$?

- $p(b|O,J) = 0.322$
- Previously we thought probability of betting with a jack was $4/13 = 0.308$
- What if we always observed his card after game play and observed he had a jack?
  - $5/14 = 0.357$
- “Naïve” approach: $(4 + 4/13)/14 = 0.308$

# “Naïve” approach

- “Naïve” approach: (4 + 4/13)/14 = 0.308
- It turns out that this is equivalent to just using prior

$$\frac{x + \frac{x}{x+y}}{x+y+1} \cdot \frac{x+y}{x+y} = \frac{x(x+y)+x}{(x+y+1)(x+y)}$$

$$= \frac{x(x+y+1)}{(x+y+1)(x+y)} = \frac{x}{x+y}$$

# Uniform prior over polyhedron

- Opponent playing uniformly at random within region of fixed strategy, e.g., specific NE or "population mean" strategy
- E.g., "sophisticated" Rock-Paper-Scissors opponents who play uniformly at random out of strategies with probability within [0.31,0.35], instead of completely random over [0,1].
  - Ganzfried/Sandholm used similar opponents for poker, EC12/TEAC15

**Algorithm 2** Algorithm for opponent exploitation with uniform prior distribution over polyhedron

**Inputs**: Prior distribution over vertices $p^0$, response functions $r_t$ for $0 \leq t \leq T$

$M_0 \leftarrow$ strategy profile assuming opponent $i$ plays each vertex $v_{i,j}$ with probability $p^0_{i,j} = \frac{1}{V_i}$
$R_0 \leftarrow r_0(M_0)$
Play according to $R_0$
**for** $t = 1$ to $T$ **do**
    **for** $i = 1$ to $N$ **do**
        $a_i \leftarrow$ action taken by player $i$ at time step $t$
        **for** $j = 1$ to $V_i$ **do**
            $p^t_{i,j} \leftarrow p^{t-1}_{i,j} \cdot v_{i,j}(a_i)$
        Normalize the $p^t_{i,j}$'s so they sum to 1
    $M_t \leftarrow$ strategy profile assuming opponent $i$ plays each vertex $v_{i,j}$ with probability $p^t_{i,j}$
    $R_t \leftarrow r_t(M_t)$
    Play according to $R_t$

# Run time of basic algorithm

- Colt Java math library for Beta computation
- Dirichlet parameters uniformly random in {1,n}
  - n = 100 corresponds to 400 prior observations
  - Previous work (Southey et al) used 200 hands per match
- Computation very fast but numerical instability for large n

| $n$ | 10 | 20 | 50 | 100 | 200 | 500 |
|---|---|---|---|---|---|---|
| Time | 0.0005 | 0.0008 | 0.0018 | 0.0025 | 0.0034 | 0.0076 |
| NaN | 0 | 0 | 0 | 0.0883 | 0.8694 | 0.9966 |

Table 1: Results of modifying Dirichlet parameters to be U{1,n} over one million samples. First row is average runtime in milliseconds. Second row is percentage of the trials that output "NaN."

# Run time of generalized algorithm

- Tested generalized algorithm for different numbers of observations keeping prior fixed
- Used Dirichlet prior with all parameters equal to 2 (as done in prior work Southey et al)
- For $\theta_b = 101$, $\theta_s = 100$, ran in 19 milliseconds.

| $n$ | 10 | 20 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|---|
| Time | 0.015 | 0.03 | 0.36 | 2.101 | 10.306 | 128.165 | 728.383 |
| NaN | 0 | 0 | 0 | 0 | 0.290 | 0.880 | 0.971 |

Table 2: Results using Dirichlet prior with all parameters equal to 2 and $\theta_b$, $\theta_s$ in U{1,n} averaged over one thousand samples.

# Comparison to other approaches

- EBBR: our Exact Bayesian Best Response
- BBR: Bayesian Best Response
  - samples strategies from prior, best responds to posterior mean
- MAP: Max A Posteriori Response
  - samples from prior, computes posteriors, best response to max
- Thompson's Response
  - Sample from prior, compute posteriors, best response to sample

| Algorithm | Initial | 10 | 25 |
|---|---|---|---|
| **EBBR** | **$0.0003 \pm 0.0009$** | **-0.0024** | **0.0012** |
| BBR | $0.0002 \pm 0.0009$ | -0.0522 | -0.138 |
| MAP | $-0.2701 \pm 0.0008$ | -0.2848 | -0.2984 |
| Thompson | $-0.2593 \pm 0.0007$ | -0.2760 | -0.3020 |
| FullBR | $0.4976 \pm 0.0006$ | 0.4956 | 0.4963 |
| Nash | $-0.3750 \pm 0.0001$ | -0.3751 | -0.3745 |

Table 3: Comparison of our algorithm with algorithms from prior work (BBR, MAP, Thompson), full best response, and Nash equilibrium. Prior is Dirichlet with parameters equal to 2. For the initial column we sampled ten million opponents from the prior, for 10 rounds we sampled one million opponents, and for 25 rounds 100,000. Results are average winrate per hand over all opponents. For initial column 95% confidence intervals are reported.

# Generalizations

- Generalized model to n different states according to arbitrary distribution $\pi$ and can take m actions
- Have closed-form solution, but contains number of terms exponential in n and m (though polynomial in T).
- Can approach or analysis be improved?

# Conclusions and directions

- First exact algorithm for Bayesian opponent exploitation in class of imperfect-information games
- Runs quickly experimentally and outperforms prior approaches, but frequent numerical instability for large n
- General meta-algorithm and new theoretical framework
- Studied Dirichlet prior and uniform over polyhedron
- Future research and extensions:
  - Partial observability (likely straightforward)
  - General game trees with sequential actions (likely hard)
  - Any number of agents (alg not specialized for 2 pl zero-sum)
  - Other important and tractable prior distributions