FlipIt and the Cramér-Rao Bound

Peter Schmidt-Nielsen

MIT, Center for Bits and Atoms

August 16, 2011

FlipIt is a two player game (with players 0 and 1), with three basic rules:

FlipIt

FlipIt is a two player game (with players 0 and 1), with three basic rules:

- There is a common resource that player i may capture at any time by paying a cost of k_i.

FLIPIT

FLIPIT is a two player game (with players 0 and 1), with three basic rules:

- There is a common resource that player i may capture at any time by paying a cost of k_i.
- For each second that the resource is held by a player he or she gains one point.

FlipIt

FlipIt is a two player game (with players 0 and 1), with three basic rules:

- There is a common resource that player i may capture at any time by paying a cost of k_i.
- For each second that the resource is held by a player he or she gains one point.
- The players don't know who owns the resource at any given time, except that when a player captures the resource he or she learns something about the past history of who owned it.

FlipIt

FlipIt is a two player game (with players 0 and 1), with three basic rules:

- There is a common resource that player i may capture at any time by paying a cost of k_i.
- For each second that the resource is held by a player he or she gains one point.
- The players don't know who owns the resource at any given time, except that when a player captures the resource he or she learns something about the past history of who owned it.

Here we only consider the case where a player capturing the resource learns only the last move (and thus who made it).

Open Problems

What is the asymptotically optimal non-adaptive strategy against an adaptive opponent?

Open Problems

What is the asymptotically optimal non-adaptive strategy against an adaptive opponent?

What is the *non-asymptotic* optimal adaptive strategy against the currently best known non-adaptive strategy (playing as a Poisson process)?

The Setup

Let player 0 be a non-adaptive Poisson process playing at a rate of $\alpha = \lambda$, playing at least once in time period τ with probability $1 - e^{-\lambda\tau}$.

The Setup

Let player 0 be a non-adaptive Poisson process playing at a rate of $\alpha = \lambda$, playing at least once in time period τ with probability $1 - e^{-\lambda\tau}$.

Player 1 is adaptive, and would like to learn λ to figure out the optimal capture rate.

The Setup

Let player 0 be a non-adaptive Poisson process playing at a rate of $\alpha = \lambda$, playing at least once in time period τ with probability $1 - e^{-\lambda\tau}$.

Player 1 is adaptive, and would like to learn λ to figure out the optimal capture rate.

Player 1 can't learn λ without capturing, but each capture has a cost of k_1.

The Setup

Let player 0 be a non-adaptive Poisson process playing at a rate of $\alpha = \lambda$, playing at least once in time period τ with probability $1 - e^{-\lambda\tau}$.

Player 1 is adaptive, and would like to learn λ to figure out the optimal capture rate.

Player 1 can't learn λ without capturing, but each capture has a cost of k_1.

If player 1 plays too fast, his or her moves might be consecutive, which will waste points and give less information about λ.

The Setup

Let player 0 be a non-adaptive Poisson process playing at a rate of $\alpha = \lambda$, playing at least once in time period τ with probability $1 - e^{-\lambda\tau}$.

Player 1 is adaptive, and would like to learn λ to figure out the optimal capture rate.

Player 1 can't learn λ without capturing, but each capture has a cost of k_1.

If player 1 plays too fast, his or her moves might be consecutive, which will waste points and give less information about λ.

In a finite game player 1 has to trade off certainty of λ with spending points (needless captures) and wasting time (needless caution).

A Tentative Information Theoretic Bound

The Cramér-Rao bound limits the variance of an unbiased[1] estimator of an unknown parameter θ of a distribution X from observations of X.

[1] This is a big assumption!

A Tentative Information Theoretic Bound

The Cramér-Rao bound limits the variance of an unbiased[1] estimator of an unknown parameter θ of a distribution X from observations of X.
The variance of our estimate of θ is at least $1/\mathcal{I}$, where $\mathcal{I}$ is the Fisher information of our observations of X.

[1] This is a big assumption!

A Tentative Information Theoretic Bound

Generous assumption:

A Tentative Information Theoretic Bound

Generous assumption:

- Every time player 1 captures, he or she learns a valid sampling of the mean time to a player 0's plays – player 1 never plays "too fast".

A Tentative Information Theoretic Bound

Generous assumption:

- Every time player 1 captures, he or she learns a valid sampling of the mean time to a player 0's plays – player 1 never plays "too fast".

If we make n plays, we get n observations of player 0's playing rate, and the minimum variance in our estimate of λ possible is...

$$\mathcal{I}(X,\lambda) = \left\langle \left(\frac{\partial}{\partial\lambda} \ln \left(\prod_{i=1}^{n} \lambda e^{-\lambda x_i} \right) \right)^2 \right\rangle \tag{1}$$

$$= \left\langle \left(\frac{\partial}{\partial\lambda} \left(n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \right) \right)^2 \right\rangle \tag{2}$$

$$= \left\langle \left(\frac{n}{\lambda} - \sum_{i=1}^{n} x_i \right)^2 \right\rangle \tag{3}$$

$$= \frac{n^2}{\lambda^2} - \frac{2n}{\lambda} \sum_{i=1}^{n} \langle x_i \rangle + \sum_{i=1}^{n} \sum_{j=1}^{n} \langle x_i x_j \rangle \tag{4}$$

$$= \frac{n^2}{\lambda^2} - \frac{2n^2}{\lambda^2} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left\langle \lambda^2 e^{-\lambda(x_i + x_j)} \right\rangle + \sum_{i=1}^{n} \langle x_i^2 \rangle \tag{5}$$

$$\frac{1}{\mathcal{I}(X,\lambda)} = \frac{\lambda^2}{n} \tag{6}$$

Tactic One

As player 0's strategy is memoryless, and by Ron's first theorem, player 1 can't play better after n plies than a non-adaptive player who guesses λ from a distribution of variance λ^2/n. Thus, we can bound the quality of player 1's play at any point in time.

As player 0's strategy is memoryless, and by Ron's first theorem, player 1 can't play better after n plies than a non-adaptive player who guesses λ from a distribution of variance λ^2/n. Thus, we can bound the quality of player 1's play at any point in time.

This proof strategy works against any non-adaptive player 0 strategy parameterized by a set of unknowns to learn, assuming we can make a reasonable approximation of the Fisher information.

Tactic One

As player 0's strategy is memoryless, and by Ron's first theorem, player 1 can't play better after n plies than a non-adaptive player who guesses λ from a distribution of variance λ^2/n. Thus, we can bound the quality of player 1's play at any point in time.

This proof strategy works against any non-adaptive player 0 strategy parameterized by a set of unknowns to learn, assuming we can make a reasonable approximation of the Fisher information.

... *assuming* that the optimal adaptive player estimates λ with no bias, the Cramér-Rao bound holds, and we have bounded the quality of the best player 1 strategy.

Tactic One

As player 0's strategy is memoryless, and by Ron's first theorem, player 1 can't play better after n plies than a non-adaptive player who guesses λ from a distribution of variance λ^2/n. Thus, we can bound the quality of player 1's play at any point in time.

This proof strategy works against any non-adaptive player 0 strategy parameterized by a set of unknowns to learn, assuming we can make a reasonable approximation of the Fisher information.

... *assuming* that the optimal adaptive player estimates λ with no bias, the Cramér-Rao bound holds, and we have bounded the quality of the best player 1 strategy.

But we don't want to assume the best strategy consists of an unbiased estimator, so we try a different tactic...

Tactic Two

Define a given bound we would like to prove on player 1's quality of play, and let any player 1 strategy that is greater than this bound a "FlipIt shark".

Tactic Two

Define a given bound we would like to prove on player 1's quality of play, and let any player 1 strategy that is greater than this bound a "FlipIt shark".

Construct a unbiased estimator that can estimate λ from the history of any game between a FlipIt shark and player 0.

Tactic Two

Define a given bound we would like to prove on player 1's quality of play, and let any player 1 strategy that is greater than this bound a "FlipIt shark".
Construct a unbiased estimator that can estimate λ from the history of any game between a FlipIt shark and player 0.
If this unbiased estimator beats the Cramér-Rao bound, by contradiction, FlipIt sharks don't exist, and the bound is proven – *regardless* of how the given shark actually works.

Our new Cramér-Rao bound, expanded for two player 1 plies:

$$\frac{1}{\mathcal{I}(X_0, X_1, \lambda)} = 1/\left\langle \left(\frac{\partial}{\partial \lambda} \ln(P[X_0 = x_0, X_1 = x_1]) \right)^2 \right\rangle \quad (7)$$

$$\vdots \quad (8)$$

$$= 1/\left\langle \left(\frac{2}{\lambda} - x_0 - x_1 + w \right)^2 \right\rangle \quad (9)$$

Where x_i are the times of the first two observations, and where w is the period of time waited between the first two plies.

Our new Cramér-Rao bound, expanded for two player 1 plies:

$$\frac{1}{\mathcal{I}(X_0, X_1, \lambda)} = 1/\left\langle \left(\frac{\partial}{\partial \lambda} \ln(P[X_0 = x_0, X_1 = x_1]) \right)^2 \right\rangle \quad (7)$$

$$\vdots \quad (8)$$

$$= 1/\left\langle \left(\frac{2}{\lambda} - x_0 - x_1 + w \right)^2 \right\rangle \quad (9)$$

Where x_i are the times of the first two observations, and where w is the period of time waited between the first two plies.

A FLIPIT shark that manages to estimate λ well enough from the first ply to gain the maximal amount of information from the second ply without wasting too much time waiting *might* end up violating this bound.

The End of the Proof – (sketch)

Let our FlipIt shark be any strategy that holds the resource more than . . . of the time on average.

The End of the Proof – (sketch)

Let our FLIPIT shark be any strategy that holds the resource more than . . . of the time on average.
This implies that it only attempts to capture the resource while it already controls the resource . . . proportion of the time, and w is no larger than . . . .

The End of the Proof – (sketch)

Let our FlipIt shark be any strategy that holds the resource more than . . . of the time on average.
This implies that it only attempts to capture the resource while it already controls the resource . . . proportion of the time, and w is no larger than . . . .
Thus, we construct an estimator of λ that gets . . . samples of player 0's Poisson process after . . . plies, and thus violates the Cramér-Rao bound.

The End of the Proof – (sketch)

Let our FLIPIT shark be any strategy that holds the resource more than . . . of the time on average.
This implies that it only attempts to capture the resource while it already controls the resource . . . proportion of the time, and w is no larger than . . . .
Thus, we construct an estimator of λ that gets . . . samples of player 0's Poisson process after . . . plies, and thus violates the Cramér-Rao bound.
Sharks don't exist, QED.

Relevant Conclusions

On Ron's open problem: What's the optimal non-adaptive strategy against an adaptive opponent?

Relevant Conclusions

On Ron's open problem: What's the optimal non-adaptive strategy against an adaptive opponent?

A Poisson process minimizes the Fisher information of a non-adaptively chosen observations, and I think it minimizes the Fisher information given by a series of adaptively timed observations.

Relevant Conclusions

On Ron's open problem: What's the optimal non-adaptive strategy against an adaptive opponent?

A Poisson process minimizes the Fisher information of a non-adaptively chosen observations, and I think it minimizes the Fisher information given by a series of adaptively timed observations.

If true, this implies that a Poisson process kills as many sharks as possible with the Cramér-Rao bound in the non-asymptotic version of FlipIt.