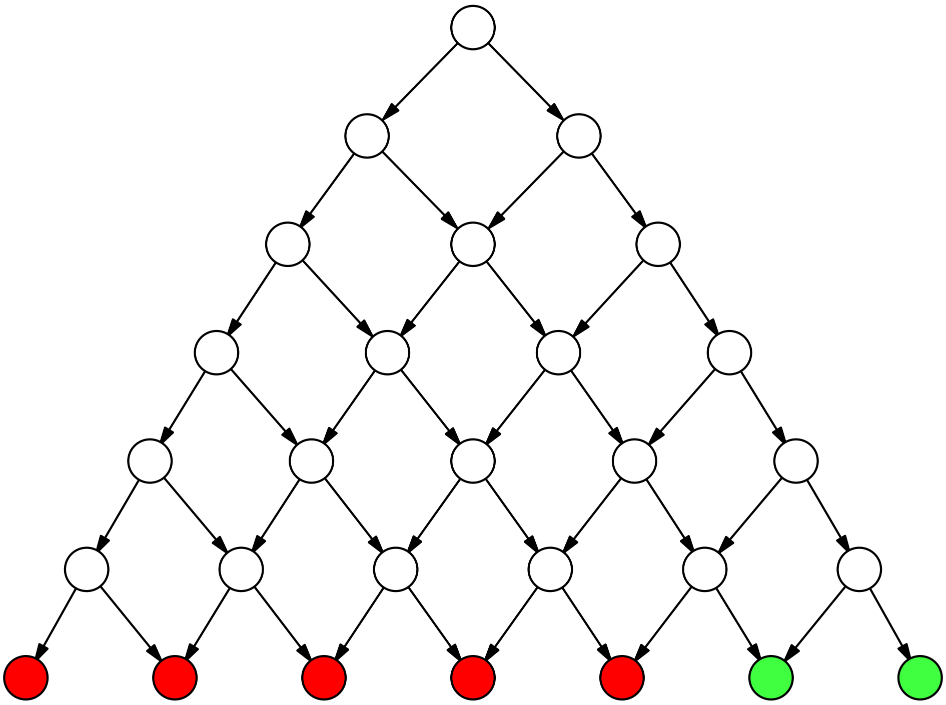
CHALMERS
UNIVERSITY OF TECHNOLOGY



Optimal Restart Games

General Theory and Near-Optimal Strategies for Rivest's Coin Game

Master's thesis in Complex Adaptive Systems
Anton Älgmyr

Master's thesis in Engineering Mathematics and Computational Science
Björn Martinsson

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis 2019

Optimal Restart Games

General Theory and Near-Optimal Strategies for Rivest's Coin Game

ANTON ÄLGMYR   BJÖRN MARTINSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Chalmers University of Technology
Gothenburg, Sweden 2019

Optimal Restart Games
General Theory and Near-Optimal Strategies for Rivest's Coin Game
ANTON ÄLGMYR
BJÖRN MARTINSSON

Supervisor: Jeffrey Steif, Department of Mathematical Sciences
Examinator: Jeffrey Steif, Department of Mathematical Sciences

Master's Thesis 2019
Department of Mathematical Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Shows the state space of the Rivest coin game with $N = 6$ and $K = 5$. The green states are the winning states and the red states are the losing states.

Typeset in LaTeX
Gothenburg, Sweden 2019

Optimal Restart Games
General Theory and Near-Optimal Strategies for Rivest's Coin Game
ANTON ÄLGMYR, BJÖRN MARTINSSON
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

The purpose of this thesis was to analyze the Rivest coin game. To do this we formally introduced a type of game which we call restart games and built a theoretical framework for analysis of such games. Most of this theory was derived by connecting restart games to quitting games (optimal stopping problems) with ideas based on [1].

Through this framework we have developed two strategies for playing the Rivest coin game which we have shown to be constant-factor from optimal. These strategies seem to generalize well to other similar games, in particular we have shown them to be constant-factor from optimal in another related coin game.

The theoretical framework also provided means to analyze the Rivest coin game numerically and even construct an optimal strategy for any particular instance of the game. The constructed optimal strategy, along with some other simple strategies, was analyzed and reaffirms the optimality analysis while also highlighting some important differences between our strategies and an optimal strategy.

Keywords: Rivest coin game, restart game, quitting game, optimal stopping, Markov decision process, dynamic programming, complexity analysis, speedrun

Acknowledgements

We would like to thank Anders Martinsson for proposing the problem that turned into this master thesis, as well as taking time for helpful discussions and guidance when we needed it the most. We would like to thank Jeffrey Steif for taking on the role of supervisor and examiner for the project as well as giving valuable and thorough feedback on our work. We would also like to thank our opponent Henrik Gingsjö for a well executed opposition along with detailed and valuable feedback on our work. In conclusion we would like to thank friends and family for enduring our incessant blabbering about our work.

Anton Älgmyr and Björn Martinsson, Gothenburg, September 2019

Contents

List of Figures

1
Introduction

Suppose you have a computer program whose runtime is stochastic. Sometimes the program might have a bad run that ends up taking a lot of time. In that case it might make sense to abort the program and start anew, sacrificing your invested time with the hope of getting a better start. In such a situation, when should you restart the program to minimize the expected total runtime? This scenario would be an instance of what we will call an optimal restart problem.

The idea of augmenting a random process with restarts to improve running times is not a new concept, but has been studied in different contexts.

In the case where the internal state of the underlying random process is hidden, there is a well-known result that says that there exist a universal strategy that comes within a logarithmic factor of the expected optimal run time [2].

In the case where the internal state is observable, you can do better. For example, Janson and Peres investigated adding restarts to random walks on lattices to minimize hitting time, where they found the asymptotic behavior of the expected hitting time for an optimal strategy [3].

Dixon's factorization algorithm is another example where this kind of random process occurs. To analyze this process in isolation Rivest (of RSA fame) introduced a restarting coin game as a simple model [4][5]. In the paper of Hu and Venkatesh [5] they include a succinct summary of the premise of the game:

> An individual has 20 fair coins in his pocket. He takes coins out of his pocket one at a time and tosses them, his objective being to obtain 15 heads. If fewer than 15 heads transpire in any round of 20 tosses, he must return all 20 coins to his pocket and restart the game. He also has the option of restarting the game at any point by ending a round of tosses and returning all 20 coins to his pocket before starting anew. The problem facing our protagonist is to choose an optimal restart strategy which would minimize the expected number of tosses he has to make before achieving his goal of 15 heads.

The Rivest coin game is good toy problem to analyze to gain insight for a category of restart problems.

Our motivation for investigating the Rivest coin game comes from the idea of modeling speedruns, the task of completing some game as quickly as possible. In particular we want to model the act of speedrunning to beat some record time. In the progression of the run you might gain or lose time relative to your target. At which point is it better to abort the current run and restart and reset the timer? This is an optimal restart problem.

The connection with the Rivest coin game is modeling fixed time gains/losses as coin tosses (heads/tails). The analogue of beating some record time is to end up with heads exceeding some goal amount.

1.1 Aim of the project

The aim of this thesis is to analyze the Rivest coin game. What is required to be optimal or close to optimal? Does insights for this game generalize to other similar games?

1.2 Methodology

The theoretical analysis will define what we call restart games and develop a framework for analyzing such games. For this formulation and framework, inspiration has been taken from [1], which investigates Markov decision processes where the player is given some number of Markov chains and is allowed to pick which Markov chain to advance, with the goal of minimizing the expected time until a terminating state is reached in any of the Markov chains. The restart game concept that will be presented here have a lot in common with this, and could conceptually be thought of as the limit where there are infinitely many identical Markov chains to pick from.

The previous investigation of the Rivest coin game done by Hu and Venkatesh [5] investigates the performance of a number of simple strategies in depth (down to explicit formulas for the simplest ones). However, in their paper (and other papers on the subject) there has been no attempt at proving any sense of optimality for a strategy. In contrast, this thesis we will, from the ground up, prove the constant-factor from optimal behavior for two simple strategies, the *promising strategy* and the *probabilistic strategy*, for the Rivest coin game (along with a related coin game).

Like in the aforementioned article, we will also analyze the Rivest coin game numerically using some insights from the restart game framework to prove correctness of the methods used. Dynamic programming and a variant of binary search will be employed to allow for performant calculations. The calculations involve very wide dynamic range of numbers and therefore we will make use of arbitrary precision arithmetic to eliminate any arithmetical errors. The numerical results will help visualize the behavior of various algorithms and reaffirm the optimality of the promising strategy and the probabilistic strategy, while also highlighting the differences between the strategies that are constant-factor from optimal, as well as shedding light on the lack of sharpness of our theoretical bounds.

1.3 Outline

In Chapter 2 we introduce restart games and a framework for analyzing them. In Chapter 3 two specific games are introduced, the Rivest coin game and another related game (Good Coin Bad Coin). In Chapter 4 all strategies that will be considered in the thesis will be presented, most importantly the *promising strategy* and the *probabilistic strategy* whose properties will be analyzed. In Chapter 5 the promising strategy is shown to be constant-factor from optimal in the two coin games, which concludes the theoretical portion of the thesis. In Chapter 6 numerical methods for analysing the Rivest coin game are presented, along with numerical results which allows for comparison of key properties for different strategies. Finally, discussion and conclusions will be presented in Chapters 7 and 8 respectively.

2

The theory of restart games

To begin we will introduce a base game which will later be augmented with restarts to create restart games. The base games will also be augmented to a quitting game as a tool for proving theorems about the restart games.

2.1 Base game

Definition 2.1 (Base game)**.** A base game is a Markov chain $G = (X \cup L \cup W, P, s_0)$, where $X \cup L \cup W$ are the states of the game (X, L, W are disjoint), P is the probability transition matrix for transitions from X to $X \cup L \cup W$, and $s_0 \in X$ is the initial state. Since the base game is a Markov chain it can not be controlled in any way, only observed.

You start out in the initial state $s_0 \in X$, then using the probabilities given by the probability transition matrix P you transition until you reach a state in L or W. If you reach W then the game is won, otherwise the game is lost.

There are two additional constraints we will require a base game to have. The state space $\text{State}(G) := X$ is required to be finite, and for any state in X there must be a non-zero probability of being able to reach a state in $L \cup W$ with one or more transitions.

Definition 2.2 (Base game conditioned on win/loss)**.** Given a base game G, define the base game G_W as the game G conditioned on winning, meaning that had you played the entire game through you are conditioned to end up in W. Likewise define G_L as conditioned on losing, meaning that you will end up in L.

2.2 Restart game

The base game can be augmented into a restart game as follows.

Definition 2.3 (Restart game)**.** Given a base game G we define the corresponding restart game $\text{Restart}(G)$.

$\text{Restart}(G)$ is played in rounds. Each round starts out with a new instance of the base game. Each round will consist of a sequence of choices. You have the choice to either play, meaning you do a transition in the base game, or restart, meaning you move on to the next round.

If you ever reach a losing state in the base game then you are forced to restart. If you reach a winning state in the base game then you are said to have finished the current round, which in turn also ends the entire restart game.

For a round in the restart game, let t denote the number of times you played. Let T be the sum of t over all rounds in the restart game. The goal is to minimize $\mathrm{E}(T)$, the expected duration of the game.

2.2.1 Restart strategies

The way of making the choices in Definition 2.3 is what makes a *restart strategy*. A restart at the beginning of a round does nothing and just adds the complication of infinitely many zero cost rounds, hence w.l.o.g. we can disallow such restarts.

It is also reasonable to restrict ourselves to restart strategies that will always make the same choice in a given state, which is formalized in the following definition.

Definition 2.4 (Stochastic Markovian restart strategies)**.** A stochastic Markovian restart strategy in $\mathrm{Restart}(G)$ is a restart strategy that for each state $s \in \mathrm{State}(G)$ plays with some fixed probability (only dependent on s).

Definition 2.5 (Determinsitic Markovian restart strategies)**.** A deterministic Markovian (DM) restart strategy in $\mathrm{Restart}(G)$ is a restart strategy that plays in some $A \subseteq \mathrm{State}(G)$ and restarts in A^{c}.

The goal is to find an *optimal restart strategy*, i.e. a strategy that attains the infimum of $\mathrm{E}(T)$. The restriction to DM restart strategies turns out to not be limiting in this regard.

Theorem 2.1 (Existence of an optimal DM restart strategy)**.** *For every restart game there exists an optimal restart strategy that is a DM restart strategy.*

Lemma 2.2 (Deterministic strategies are at least as good as stochastic)**.** *For every stochastic Markovian strategy in the restart game there exists a deterministic Markovian strategy for which* $\mathrm{E}(T)$ *is either equal or smaller.*

Proof of Lemma 2.2. Start out with any stochastic Markovian strategy $\mathcal{S}$. Construct the DM restart strategy $\mathcal{S}'$ by for every state, make $\mathcal{S}'$ play the action that minimizes the duration of the game had you played $\mathcal{S}$ in the future (if there are many such actions then pick any).

Claim. *Let the restart strategy $\mathcal{Z}_m$ be the strategy that plays $\mathcal{S}'$ for the first m actions and plays $\mathcal{S}$ after that point. Then for $m \geq 0$,* $\mathrm{E}(T)$ *is equal or smaller when playing $\mathcal{Z}_m$ compared to $\mathcal{S}$.*

The proof of the claim is done by induction. The statement is trivially true for $m = 0$. Assume it holds for $m - 1$, then note that the mth action done by $\mathcal{S}'$ is the optimal action assuming you will play S from action $m + 1$ and onwards. So if the claim is true for $m - 1$ it is true for m.

The claim shows that it is possible to postpone the transition to the stochastic strategy arbitrarily far while not increasing the expected duration. Assuming the limit behavior of the expected value is not pathological this means that the DM strategy $\mathcal{Z}_\infty = \mathcal{S}'$ is at least as good as $\mathcal{S}$. $\square$

Proof of Theorem 2.1. In this proof only the subset of Markovian strategies need to be analyzed, because the only information that matters during the game is the current state of the base game. So the only thing that needs to be shown is the existence of an optimal DM strategy over the set of all Markovian strategies.

From Lemma 2.2 it follows that only DM strategies need to be considered, and as the set of DM strategies is finite there must exist an optimal one. $\square$

2.2.2 Master theorem for restart games

There is a fundamental formulation for the expected duration of a restart game. This formulation comes from separating realizations of the base game into win-conditioned games G_W and loss-conditioned games G_L, and rounds into restarting rounds R and non-restarting/finishing rounds $\neg R$.

To state the theorem some additional notation will be introduced, let $\mathrm{E}_\mathcal{S}(\cdot)$ and $\mathrm{P}_\mathcal{S}(\cdot)$ be the expected value and probability when using strategy $\mathcal{S}$. For this to be well-defined we will restrict ourselves to DM strategies. Some illustrative examples of the notation are

$\mathrm{E}_\mathcal{S}(t \mid R)$
: Playing with strategy $\mathcal{S}$, the expected round length given that a restart occurs.

$\mathrm{P}_\mathcal{S}(\neg R \mid G_W)$
: Playing with strategy $\mathcal{S}$, the probability to finish a round (winning the game) given that the base game is win-conditioned.

$\mathrm{P}(G_L)$
: The probability of the base game being loss-conditioned, i.e. had you played without restarts it would have ended in a loss.

Lemma 2.3 (Duration of a restart game)**.** *Given a DM restart strategy $\mathcal{S}$ with $\mathrm{P}_\mathcal{S}(\neg R) > 0$, i.e. the strategy $\mathcal{S}$ has a non-zero probability of finishing, then*

$$\mathrm{E}_\mathcal{S}(T) = \mathrm{E}_\mathcal{S}(t \mid R)\frac{\mathrm{P}_\mathcal{S}(R)}{\mathrm{P}_\mathcal{S}(\neg R)} + \mathrm{E}_\mathcal{S}(t \mid \neg R).$$

Theorem 2.4 (Master theorem for restart games)**.** *Lemma 2.3 can be further split conditioned on winning/losing the base game as*

$$\mathrm{E}_\mathcal{S}(T) = \mathrm{E}_\mathcal{S}(t \mid G_L)\frac{1}{\mathrm{P}_\mathcal{S}(\neg R \mid G_W)}\frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} + \mathrm{E}_\mathcal{S}(t \mid G_W, R)\frac{\mathrm{P}_\mathcal{S}(R \mid G_W)}{\mathrm{P}_\mathcal{S}(\neg R \mid G_W)} + \mathrm{E}_\mathcal{S}(t \mid \neg R).$$

Proof of Lemma 2.3. Let N denote the number of rounds played by the strategy and let t_i denote the length of round i. The duration of the game T can be written as

$$T = \left(\sum_{i=1}^{N-1} t_i\right) + t_N.$$

The first $N-1$ rounds are restart rounds and the last round is a winning round. Note that the t_i for $i < N$ are independent and identically distributed and also independent of $N-1$. Therefore we can calculate the expected value of T using Wald's equation. The expected value is

$$\begin{aligned}
\mathrm{E}_{\mathcal{S}}(T) &= \mathrm{E}_{\mathcal{S}}\left(\sum_{i=1}^{N-1} t_i\right) + \mathrm{E}_{\mathcal{S}}(t_N) \\
\{\text{Wald's equation}\} &= \mathrm{E}_{\mathcal{S}}(t_1)\mathrm{E}_{\mathcal{S}}(N-1) + \mathrm{E}_{\mathcal{S}}(t_N) \\
&= \mathrm{E}_{\mathcal{S}}(t \mid R)\left(\frac{1}{\mathrm{P}_{\mathcal{S}}(\neg R)} - 1\right) + \mathrm{E}_{\mathcal{S}}(t \mid \neg R) \\
&= \mathrm{E}_{\mathcal{S}}(t \mid R)\frac{\mathrm{P}_{\mathcal{S}}(R)}{\mathrm{P}_{\mathcal{S}}(\neg R)} + \mathrm{E}_{\mathcal{S}}(t \mid \neg R).
\end{aligned}$$

□

Proof of Theorem 2.4. This can be shown by either rewriting using Bayes' formula or using a derivation similar to the one in Lemma 2.3. The proof itself is left out for the sake of brevity. □

The point of the rewriting in Theorem 2.4 is to separate the strategy-independent part of the expression, $\mathrm{P}(G_L)/\mathrm{P}(G_W)$, from the rest. There is also another more fundamental reason why the rewriting is natural.

An instance of a base game G can thought of as having a predetermined outcome unknown to the player. This means that it is possible to split the analysis of G into two different games G_W and G_L, and then combine the two to recreate G. Informally this can be described as

$$G = \begin{cases} G_W & \text{with probability } \mathrm{P}(G_W) \\ G_L & \text{with probability } \mathrm{P}(G_L). \end{cases}$$

It turns out that the probabilities can be varied independently of G_W and G_L.

Theorem 2.5. *Given a base game G, the game*

$$G' := \begin{cases} G_W & \textit{with probability } \mathrm{P}(G'_W) \\ G_L & \textit{with probability } \mathrm{P}(G'_L) \end{cases} \qquad \mathrm{P}(G'_W) = 1 - \mathrm{P}(G'_L) \in (0,1)$$

is also a base game.

Remark. G' is different from G unless $\mathrm{P}(G'_W) = \mathrm{P}(G_W)$.

Remark. G and G' share the same state space, with the intepretation that you are in a superposition of playing either G_W or G_L. To show that G' is a base game one just needs to show that G' is a Markov chain, i.e. the transition probability matrix exists, i.e. transition probabilities are only dependent on the current state of the game.

Proof of Theorem 2.5. Firstly note that G_W and G_L are Markov chains since it is well known that a Markov chain conditioned on reaching a certain state is in itself a Markov chain.

Secondly, suppose that when playing G' the first $n+1$ states reached are $s_0, s_1, \ldots, s_n$. Note that the conditional probability of G'_W given this history only depends on the present state s_n and not the full history. One way to show this is to apply Bayes' theorem to the ratio

$$\begin{aligned}\frac{\mathrm{P}(G'_W \mid \text{history})}{\mathrm{P}(G'_L \mid \text{history})} &= \frac{\mathrm{P}(G'_W)}{\mathrm{P}(G'_L)} \frac{\mathrm{P}(\text{history} \mid G'_W)}{\mathrm{P}(\text{history} \mid G'_L)} \\ &= \frac{\mathrm{P}(G'_W)}{\mathrm{P}(G'_L)} \frac{\mathrm{P}(\text{history} \mid G_W)}{\mathrm{P}(\text{history} \mid G_L)} \\ &= \frac{\mathrm{P}(G'_W)}{\mathrm{P}(G'_L)} \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \frac{\mathrm{P}(G_W \mid \text{history})}{\mathrm{P}(G_L \mid \text{history})}.\end{aligned}$$

Since G is a Markov chain, the RHS only depends on the present state s_n. So $\mathrm{P}(G'_W \mid \text{history})$ and $\mathrm{P}(G'_L \mid \text{history})$ on the LHS must also only depend on s_n and not the full history.

Lastly, note that playing G' given history can be seen as playing G_W with $\mathrm{P}(G'_W \mid \text{history})$ and playing G_L with $\mathrm{P}(G'_L \mid \text{history})$. As the conditioned games are Markov chains and the conditional probabilities given history only depend on the present state, G' is a Markov chain. $\square$

So changing the win/loss probabilities in the base game only affects the fraction $\mathrm{P}(G_L)/\mathrm{P}(G_W)$ in Theorem 2.4. Everything else remains the same.

In the case of hard games, $\mathrm{P}(G_W) \ll 1$, the first term will dominate and the only thing the strategy can affect is $\frac{\mathrm{E}_S(t \mid G_L)}{\mathrm{P}_S(\neg R \mid G_W)}$.

2.2.3 Grade/optimality

This section only aims to state a number of theorems about how to optimally play restart games by introducing something we call the grade of a state in a restart game. The proofs of these theorems will be done in Section 2.5 using the connections between restart games and their corresponding quitting games established in Section 2.4.

Definition 2.6 (Minimum duration)**.** For a restart game $\mathrm{Restart}(G)$, define the minimum duration $\mathrm{E}^*(T) := \min_S \mathrm{E}_S(T)$ over all possible DM restart strategies in $\mathrm{Restart}(G)$.

Note that this definition of minimum duration coincides with the infimum of $\mathrm{E}(T)$ over all restart strategies according to Theorem 2.1.

Definition 2.7 (Grade)**.** Given a base game G and $s \in \mathrm{State}(G)$ define the grade $\gamma(s)$ as the minimum duration of $\mathrm{Restart}(G')$ where G' is a modified version of G with the initial state set to s.

Theorem 2.6 (Optimal restart criterion). *A DM restart strategy $\mathcal{S}$ plays Restart(G) optimally if and only if*

$$\begin{cases} \textit{Plays in } \{s \in \textit{State}(G) : \gamma(s) < \mathrm{E}^*(T)\} \\ \textit{Restarts in } \{s \in \textit{State}(G) : \gamma(s) > \mathrm{E}^*(T)\} \end{cases}$$

In the case of $\gamma(s) = \mathrm{E}^*(T)$ the action taken is arbitrary.

The optimal restart criterion says what a DM strategy needs to fulfill in order to be optimal. It is also possible to state a similar theorem pertaining to forcing the player to not restart in a set of states.

Theorem 2.7 (Forced to play). *Given a restart game Restart(G) with minimum duration $\mathrm{E}^*(T)$, then*

$$\min_{\mathcal{S}\in F} \mathrm{E}_{\mathcal{S}}(T) \leq \max(\mathrm{E}^*(T), \gamma(s_1), \ldots, \gamma(s_n)),$$

where F is the set of all DM restart strategies $\mathcal{S}$ that does not restart in $\{s_1, \ldots, s_n\} \subseteq$ State(G).

This theorem is the key insight used to do constant-factor from optimal analysis. The idea is that from knowing that the grade of some set of states is a constant factor from the minimum duration, then it follows that it is possible to find a constant-factor from optimal strategy that plays in those states.

2.3 Quitting game

Restart games are just one of the possible games you can create using a base game. The problem with restart games is that it is not possible to view playing from any state as a restart game in itself because you need to keep track of which state to restart to (s_0). Due to this we will introduce a new category of games similar to restart games, which we will call quitting games.

Definition 2.8 (Quitting game). Given a base game G and *quitting cost* $q \geq 0$, define the corresponding quitting game Quitting(G, q) as playing a single round of the base game. The round will consist of a sequence of choices. You have the choice to either quit, meaning you pay the quitting cost q and terminate, or play, meaning you pay 1 and do a transition in the base game.

If you ever reach a losing state in the base game then you are forced to quit. If you reach a winning state in the base game then you terminate without having to pay the quitting cost.

The goal of the game is to minimize the expected *total cost* C_q. This is an optimal stopping problem.

For a quitting game, playing from any state is in itself a quitting game with the initial state set to this state. This symmetry means that as long as we know the optimal action for an arbitrary initial state, we know how to play the quitting game optimally. There is also a strong connection between restart games and quitting games which will allow us to prove the optimality theorems about restart games.

2.3.1 Quitting strategies

The way of making the choices in Definition 2.8 is what makes a *quitting strategy*. It is possible to directly translate DM restart strategies into quitting strategies by switching restarting and quitting. Note however that there is a single quitting strategy that cannot be translated into a restart strategy, the strategy that quits at the initial state.

The parallel to DM restart strategies will be called DM quitting strategies and the existence of an optimal DM quitting strategy follows from the same argument as in Theorem 2.1.

2.3.2 Master theorem for quitting games

To formulate the master theorem for quitting games we will use similar notation as the one for restart games. Just as with restart games, the number of times the play action is taken before termination is denoted by t. The analogue of the duration of the game T of a restart game is the total cost C_q. Instead of denoting the outcome of a round as R (restarted) or $\neg R$ (finished), we will for quitting games use Q (quitted) and $\neg Q$ (finished).

Theorem 2.8 (Duration of a quitting game). *Given a quitting game $Quitting(G, q)$ and a DM quitting strategy $\mathcal{S}$*

$$\mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{P}_{\mathcal{S}}(Q)(\mathrm{E}_{\mathcal{S}}(t \mid Q) + q) + \mathrm{P}_{\mathcal{S}}(\neg Q)\mathrm{E}_{\mathcal{S}}(t \mid \neg Q).$$

Proof. This is simply a decomposition of $\mathrm{E}_{\mathcal{S}}(C_q)$ based on the outcome (quitted/finished). □

2.3.3 Grade/optimality

This is the optimality criterion for quitting games which depends on the grade defined for restart games. This will be proved in Section 2.5.

Theorem 2.9 (Optimal quitting criterion). *A DM quitting strategy $\mathcal{S}$ plays $Quitting(G, q)$ optimally if and only if*

$$\begin{cases} \textit{Plays in } \{s \in \textit{State}(G) : \gamma(s) < q\} \\ \textit{Quits in } \{s \in \textit{State}(G) : \gamma(s) > q\}, \end{cases}$$

where γ is the grade defined for $Restart(G)$. In the case of $\gamma(s) = q$ the action taken is arbitrary.

Corollary 2.10 (Alternative definition of grade). *Despite the grade being defined for restart games, from Theorem 2.9 it is possible to give many equivalent definitions of γs purely in terms of quitting games.*

1. *$\gamma(s)$ is the smallest quitting cost q such that it is optimal to play in s.*

2. $\gamma(s)$ *is the largest quitting cost* q *such that it is optimal to quit in* s*.*

3. $\gamma(s)$ *is the quitting cost* q *such that the optimal action at* s *is arbitrary.*

Remark. Any of these three alternative definitions could be used as the basis for analysis of restart and quitting games. We originally got the idea of using grade from [1], which defines the grade only in terms of quitting games. Since this thesis aims to analyze restart games, we have arbitrarily chosen to define γ in terms of restart games.

2.4 Connection between restart games and quitting games

This section aims to prove all of the optimality theorems for restart games and quitting games by using the connection between them. A big part of the reasoning used is based on using DM restart strategies to play quitting games.

Theorem 2.11 (Relationship between restart and quitting games)**.** *Let* G *be a base game and let* $\mathcal{S}$ *be a DM restart strategy for Restart*(G)*. Then for* $q \geq 0$

$$\mathrm{E}_{\mathcal{S}}(T) < q \iff \mathrm{E}_{\mathcal{S}}(C_q) < q \tag{2.1}$$
$$\mathrm{E}_{\mathcal{S}}(T) > q \iff \mathrm{E}_{\mathcal{S}}(C_q) > q \tag{2.2}$$
$$\mathrm{E}_{\mathcal{S}}(T) = q \iff \mathrm{E}_{\mathcal{S}}(C_q) = q \tag{2.3}$$

where T *is the duration of Restart*(G) *and* C_q *is the total cost of Quitting*(G, q)*.*

Corollary 2.12 (Corollary of Theorem 2.11)**.** *For a base game* G *and a DM restart strategy* $\mathcal{S}$ *on Restart*(G)*,* $\mathcal{S}$ *is optimal at Restart*(G) *if and only if it is optimal at Quitting*$(G, q = \mathrm{E}^*(T))$*.*

Lemma 2.13. *Let* G *be a base game and let* $\mathcal{S}$ *be a DM restart strategy for Restart*(G) *such that* $\mathrm{P}_{\mathcal{S}}(\neg R) > 0$*. Then for* $q \geq 0$

$$\mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{P}_{\mathcal{S}}(R)q + \mathrm{P}_{\mathcal{S}}(\neg R)\mathrm{E}_{\mathcal{S}}(T).$$

Proof of Lemma 2.13. The proof is just algebraic manipulation of Lemma 2.3 and Theorem 2.8. The two theorems state that

$$\mathrm{E}_{\mathcal{S}}(T) = \mathrm{E}_{\mathcal{S}}(t \mid R)\frac{\mathrm{P}_{\mathcal{S}}(R)}{\mathrm{P}_{\mathcal{S}}(\neg R)} + \mathrm{E}_{\mathcal{S}}(t \mid \neg R)$$
$$\mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{P}_{\mathcal{S}}(Q)(\mathrm{E}_{\mathcal{S}}(t \mid Q) + q) + \mathrm{P}_{\mathcal{S}}(\neg Q)\mathrm{E}_{\mathcal{S}}(t \mid \neg Q).$$

Note that all the Qs can be replaced with Rs as the quantities only involve one round in the quitting/restart game and as such are equivalent. Combining the two expressions gives

$$\begin{aligned}
\mathrm{E}_{\mathcal{S}}(C_q) &= \mathrm{P}_{\mathcal{S}}(Q)(\mathrm{E}_{\mathcal{S}}(t \mid Q) + q) + \mathrm{P}_{\mathcal{S}}(\neg Q)\mathrm{E}_{\mathcal{S}}(t \mid \neg Q) \\
&= \mathrm{P}_{\mathcal{S}}(R)(\mathrm{E}_{\mathcal{S}}(t \mid R) + q) + \mathrm{P}_{\mathcal{S}}(\neg R)\mathrm{E}_{\mathcal{S}}(t \mid \neg R) \\
&= \mathrm{P}_{\mathcal{S}}(R)q + \mathrm{P}_{\mathcal{S}}(R)\mathrm{E}_{\mathcal{S}}(t \mid R) + \mathrm{P}_{\mathcal{S}}(\neg R)\mathrm{E}_{\mathcal{S}}(t \mid \neg R) \\
&= \mathrm{P}_{\mathcal{S}}(R)q + \mathrm{P}_{\mathcal{S}}(\neg R)\left(\mathrm{E}_{\mathcal{S}}(t \mid R)\frac{\mathrm{P}_{\mathcal{S}}(R)}{\mathrm{P}_{\mathcal{S}}(\neg R)} + \mathrm{E}_{\mathcal{S}}(t \mid \neg R)\right) \\
&= \mathrm{P}_{\mathcal{S}}(R)q + \mathrm{P}_{\mathcal{S}}(\neg R)\mathrm{E}_{\mathcal{S}}(T).
\end{aligned}$$

□

Proof of Theorem 2.11. In the case $\mathrm{P}_{\mathcal{S}}(\neg R) = 0$ then $\mathrm{E}_{\mathcal{S}}(T)$ is infinite which is greater than any q. Also $\mathrm{E}_{\mathcal{S}}(C_q) \geq q + 1$ as $\mathcal{S}$ is a restart strategy which will always quit in the quitting game, so the total cost will be at least $q + 1$.

In the case of $\mathrm{P}_{\mathcal{S}}(\neg R) > 0$ we can apply Lemma 2.13. So we know that

$$\mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{P}_{\mathcal{S}}(R)q + \mathrm{P}_{\mathcal{S}}(\neg R)\mathrm{E}_{\mathcal{S}}(T).$$

Note that $\mathrm{E}_{\mathcal{S}}(C_q)$ is a weighted mean of q and $\mathrm{E}_{\mathcal{S}}(T)$. Thus the statement in the theorem follows trivially as $\mathrm{P}_{\mathcal{S}}(\neg R) > 0$. □

Proof of Corollary 2.12. For any DM restart strategy $\mathcal{S}$ playing Restart(G) and Quitting($G, q = \mathrm{E}^*(T)$), by Theorem 2.11,

$$\mathrm{E}_{\mathcal{S}}(T) = \mathrm{E}^*(T) \iff \mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{E}^*(T).$$

Note that by definition $\mathrm{E}_{\mathcal{S}}(T) = \mathrm{E}^*(T)$ implies optimality at Restart(G).

If $\mathrm{E}_{\mathcal{S}}(C_q) = \mathrm{E}^*(T)$ then $\mathcal{S}$ must also be optimal at Quitting($G, q = \mathrm{E}^*(T)$). This is because any better DM quitting strategy cannot afford to quit immediately, which would be a DM restart strategy and by Theorem 2.11 would violate the optimality of $\mathcal{S}$ in the restart game. □

2.5 Proofs of optimality theorems

We are finally at the point where we can prove all the optimality theorems. First we prove the optimality criterion for quitting games Theorem 2.9, which is then used to prove the optimality criterion for restart games Theorem 2.6 and the impact of forcing play in restart games Theorem 2.7.

Proof of quitting criterion Theorem 2.9. Any state in the quitting game is in itself a quitting game with the same quitting cost q. This means that it is enough to prove that the quitting criterion is correct for the first move at the initial state s_0, since then it will be correct for every move.

Criterion implies optimality There are two distinct cases to handle. Namely that $\gamma(s_0) \leq q$ implies that there exists an optimal quitting strategy that plays as its first action, and that $\gamma(s_0) \geq q$ implies that there exists an optimal quitting strategy that quits as its first action.

Case $\gamma(s_0) \leq q$ The only strategy that does not play as its initial move quits at cost q. So the only thing to prove is the existence of a strategy that plays as its first action and that has cost $\leq q$.

One such strategy is $\mathcal{S}^*$, the restart strategy optimal in Restart(G). By definition $\gamma(s_0) \equiv \mathrm{E}_{\mathcal{S}^*}(T)$ which was assumed to be $\leq q$. So Theorem 2.11 implies that $\mathrm{E}_{\mathcal{S}^*}(C_q) \leq q$.

Case $\gamma(s_0) \geq q$ In this case any restart strategy $\mathcal{S}$ fulfills $\mathrm{E}_{\mathcal{S}}(T) \geq \gamma(s_0) \geq q$. Again Theorem 2.11 together with $\mathrm{E}_{\mathcal{S}}(T) \geq q$ implies that $\mathrm{E}_{\mathcal{S}}(C_q) \geq q$. So all quitting strategies that play in s_0 have $\mathrm{E}_{\mathcal{S}}(C_q) \geq q$. The strategy that quits immediately has cost q and is therefore optimal.

Optimality implies criterion Again there are two distinct cases to handle. Namely that optimality implies playing if $\gamma(s_0) < q$, and that optimality implies quitting if $\gamma(s_0) > q$.

Case $\gamma(s_0) < q$ Assume for the sake of contradiction that an optimal strategy quits in s_0 at cost q. From the definition of grade there exists a restart strategy with $\mathrm{E}_{\mathcal{S}}(T) = \gamma(s_0) < q$, and by Theorem 2.11 $\mathrm{E}_{\mathcal{S}}(C_q) < q$. This contradicts the assumption of optimality, hence any optimal strategy must play in s_0.

Case $\gamma(s_0) > q$ Assume for the sake of contradiction that an optimal strategy $\mathcal{S}$ plays in s_0. From the definition of grade, using $\mathcal{S}$ in the restart game results in $\mathrm{E}_{\mathcal{S}}(T) \geq \gamma(s_0) > q$, and by Theorem 2.11 $\mathrm{E}_{\mathcal{S}}(C_q) > q$. This contradicts the assumption of optimality as quitting at s_0 would be a better strategy, hence any optimal strategy must quit in s_0. $\square$

Proof of restart criterion Theorem 2.6. According to Corollary 2.12, any DM restart strategy $\mathcal{S}$ that is optimal at either Restart(G) or Quitting$(G, \mathrm{E}^*(T))$ must be optimal at both. As such the quitting criterion Theorem 2.9 with $q = \mathrm{E}^*(T)$ will also work as a restart criterion. $\square$

Proof of forced to play Theorem 2.7. Note that in the quitting game Quitting(G, q) the DM quitting strategy $\mathcal{S}$ which

$$\begin{cases} \text{Plays in } \{s \in \mathrm{State}(G) : \gamma(s) \leq q\} \\ \text{Quits in } \{s \in \mathrm{State}(G) : \gamma(s) > q\} \end{cases}$$

is optimal according to the quitting criterion Theorem 2.9. As $\mathcal{S}$ is optimal at Quitting(G, q)

$$\mathrm{E}_{\mathcal{S}}(C_q) \leq q. \tag{2.4}$$

Now consider $q = \max(\mathrm{E}^*(T), \gamma(s_1), \ldots, \gamma(s_n))$. As $\mathrm{E}^*(T)$ is the grade of the initial state and $\mathcal{S}$ only quits if $\mathrm{E}^*(T) > q$, the strategy will not quit immediately and as such is a restart strategy. For restart strategies we can apply Theorem 2.11 which together with (2.4) gives

$$\mathrm{E}_{\mathcal{S}}(T) \leq q = \max(\mathrm{E}^*(T), \gamma(s_1), \ldots, \gamma(s_n)).$$

$\square$

3

Coin games

In this section two coin games will be introduced. The primary game is very similar to the coin game proposed by R. L. Rivest as a model for one part of the well-known Dixon's factorization algorithm [4][5], for use in analyzing possible runtime improvements of the algorithm. The secondary game is a related game introduced by us that is easier to analyze, in the hope that the proof techniques will carry over to the primary game (which turned out to be correct).

For both games the probability $p \in (0, 1)$ of a coin flip being heads will be considered a hyperparameter, meaning it will be thought of as a constant when analyzing the games.

3.1 The Rivest coin game

The original Rivest Coin game is getting $\geq K$ heads in a sequence of N coin flips. If this is not achieved during N tosses the sequence is discarded and you get to try again. As soon as K heads is achieved the game stops. Note that this could happen in the middle of the sequence. On top of this you are allowed to voluntarily discard an unfavorable sequence at any time and try again. The goal is to minimize the expected total number of coin tosses until the game terminates.

For the sake of simplicity we make a minor change to this game, namely that the terminating condition is changed so that game ends if the sequence contains $\geq K$ heads after making all N coin tosses. The extra cost from this is that the winning round has to be played to completion. This is a cost of size $\leq N$ and will be quickly dominated as the game gets harder (when ϵ gets larger). This variation will be referred to as Rivest$(N, K; p)$.

Definition 3.1 (The base game of the Rivest coin game, Rivest$(N, K; p)$)**.** The game takes two parameters N and K where $N > 0$ and $0 \leq K \leq N$, and one hyperparameter $p \in (0, 1)$. A state in the game is represented by (n, H_n), the number of tossed coins and the number of heads among those tossed coins. The starting state of the game is $(0, 0)$. The transition from state (n, H_n) is decided by a coin flip with probability p. Getting a head means transition to $(n + 1, H_n + 1)$ and getting a tail means transition to $(n + 1, H_n)$. Winning states are $\{(N, H_N) : H_N \geq K\}$ and losing states are $\{(N, H_N) : H_N < K\}$. An illustration of a game instance can be seen in Figure 3.1.

Remark. An alternative parameterization of the game that is useful in analysis is Rivest$(N, \epsilon; p)$, where $\epsilon := K/N - p$. The formulation in ϵ captures how much

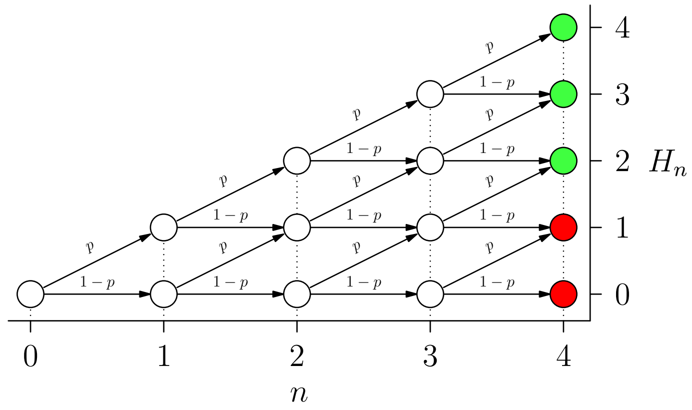


Figure 3.1: The (base game of the) Rivest coin game for $N = 4$, $K = 2$. Circles represent states, and the arrows between the states represent transition probabilities. At $n = 4$ the top three green states are the winning states and the bottom two red states are losing states.

better/worse than average you need to be, and is useful in the analysis because asymptotic behavior of optimal strategies turns out to be readily expressed in ϵ.

Dealing with $\epsilon \leq 0$ is something that is a bit bothersome in the analysis, in that the game behaves radically different than when $\epsilon > 0$. However, this case is trivial and can easily be dealt with. Importantly, if $\epsilon \leq 0$ then

$$P(G_W) \geq \frac{1}{2} \tag{3.1}$$

which follows from

$$\epsilon \leq 0 \implies P(G_W) \geq P(H_N \geq \lfloor Np \rfloor \mid H_N \sim \text{Binomial}(N, p))$$

and that the median of Binomial(N, p) is either $\lfloor np \rfloor$ or $\lceil np \rceil$. The means that for $\epsilon \leq 0$ when playing the game without voluntary restarts, the expected number of rounds played is at most 2.

3.2 Good coin bad coin (GCBC)

The formulation in ϵ invites one variation of the game that lends itself well to analysis. Rounds that win the Rivest base game could (in a heuristic sense) be seen as having been played with a lucky coin with a higher probability of heads, likewise playing a round of such a coin would most likely lead to winning the game. This can be formalised as follows.

Definition 3.2 (The base game of good coin bad coin, GCBC$(N, \epsilon, r;\, p)$)**.** The game has one hyperparameter $p \in (0, 1)$ and three parameters $N > 0$, $\epsilon \in (0, 1 - p]$ and $r \in (0, 1)$. At the beginning of the game you are given an unknown coin to play with. With probability r the coin is a *bad coin* with P(head) $= p$ and with probability $1 - r$ the coin is a *good coin* with P(head) $= p + \epsilon$. After a sequence of N tosses you win if you have a good coin, otherwise you lose.

The observable state is only (n, H_n), the number of tossed coins and the number of heads among those tossed coins, which will be the state in the base game. But the "true state" is $(n, H_n, \text{good/bad})$ where the states and state transitions for (n, H_n, good) and (n, H_n, bad) are defined in a similar manner as for the Rivest coin game, but with differently biased coins.

Remark. Note that

$$\text{State}(\text{GCBC}(N, \epsilon, r;\, p)) = \{(n, H_n) : n < N\}$$

where the observable states can be seen as being superposition of the "true states". In this interpretation, when you have made the Nth toss and arrive at state (N, H_N) you win or lose with some probability dependent on H_N. From the observer's point of view this probability is just the probability of having a good coin given the observed state. So while H_N by itself does not determine whether you win or lose the game, the higher H_N is, the higher the probability of having a good coin is.

Remark. One seemingly big difference between Rivest and GCBC is that only GCBC has the parameter r. However according to Theorem 2.5 it is possible to generalize Rivest in order to add an r parameter to it as well. So the parameters of two games are not that different.

3.3 Ease of analysis or ease of numerical simulation

One interesting thing to note is that the two games are very much opposites when it comes to ease of analysis. Finding an optimal strategy to the Rivest coin game is straightforward to do numerically, but it is more difficult to analyze in terms of win-conditioned and loss-conditioned rounds. On the other hand GCBC is straightforward to analyze as win-conditioned rounds have good coins and loss-conditioned rounds have bad coins, but finding an optimal strategy for it numerically is a bit more involved.

4

Strategies

This section contains all the strategies considered in this thesis. The strategies are intended for use in both the good coin bad coin game and the Rivest coin game introduced in Chapter 3. All the strategies introduced in this section will be compared in the numerical analysis, but only the indolent strategy I, the promising strategy $\mathscr{P}$ and the probabilistic strategy $\mathscr{P}r$ will be considered in the theoretical analysis.

4.1 Indolent strategy

The indolent strategy I is the strategy that never voluntarily restarts. For I there are mainly three properties of interest used in this thesis.

- The probability to finish a win-conditioned round is $\mathrm{P}_I(\neg R \mid G_W) = 1$.
- For Rivest$(N, \epsilon;\, p)$ and GCBC$(N, \epsilon, r;\, p)$, the length of a round is always N.
- As I never restarts, its distribution of heads at the end of a round can be used to bound the distribution for other strategies. This connection makes analyzing some properties of the optimal strategy possible by just analyzing I, which will be an essential part of the proofs later on in Section 5.3.2.

4.2 Promising strategy

The promising strategy is a new and original strategy constructed in order to play Rivest$(N, \epsilon;\, p)$, $\epsilon > 0$. One motivation behind the strategy is the idea to split the game into several intervals of length δ with δ chosen in such a way that with a fixed probability the number of heads tossed inside any of these intervals will be $\geq (p+\epsilon)\delta$, and so from CLT it follows that $\delta = \mathcal{O}(1/\epsilon^2)$.

With this construction of intervals it is possible to make the strategy behave a lot like a geometric sum. This can be done by doing restarts at the end of an interval n_i if the number of heads $< (p + \epsilon/2)n_i$. On loss-conditioned rounds it is expected to only play $\mathcal{O}(1)$ intervals, and on win-conditioned rounds it will finish the round with probability bounded away from zero. The formal definition of the strategy is the following

Definition 4.1 (The promising strategy $\mathscr{P}$ for Rivest$(N, \epsilon;\, p)$ and GCBC$(N, \epsilon, r;\, p)$)**.** $\mathscr{P}$ takes two parameters, $\alpha > 0$ and $\beta \in (0, 1)$.

If $\epsilon \leq 0$ then $\mathcal{P}$ is completely indolent and will never try to restart. If $\epsilon > 0$ then it will place equidistant checkpoints with distance $\delta := \lceil \alpha/\epsilon^2 \rceil$. So checkpoints are placed at $n_i := i\delta, i = 1, 2, \ldots, \lfloor (N-1)/\delta \rfloor$. It will restart at checkpoint i iff $H_{n_i} < (p + \beta\epsilon)n_i$.

The choice of α and β are rather arbitrary, one reasonable choice is $\alpha = 2$ and $\beta = 0.5$ which we will use going forward.

The basic idea of $\mathcal{P}$ is to place equidistant checkpoints with distance δ and only allow restarts at these checkpoints. This limitation forces the strategy to wait before taking an action, which will give the strategy significant information to base its decision on. The main drawback is that this forces rounds to be at least δ long, even in cases where early restarts could be beneficial. This means that δ needs to be deliberately chosen to take both effects into account. It turns out that δ close to $1/\epsilon^2$ is a good choice.

Theorem 4.1 (Promising strategy is constant-factor from optimal)**.** *For the games GCBC(N, ϵ, r; p) and Rivest(N, ϵ; p), if the parameters α and β are chosen such that*

$$\frac{\ln 2}{2} < (1-\beta)^2\alpha,$$

then $\mathrm{E}_{\mathcal{P}}(T)$ is at most a constant factor from $\mathrm{E}^*(T)$ (constant w.r.t. N, ϵ and r).

Remark. The constraint on α and β comes from Theorem 4.3. Other than fulfilling this constraint, the choice of α and β is rather arbitrary and will only affect numerical constants in the derived bounds. One reasonable choice is $\alpha = 2$ and $\beta = 0.5$, which will be used as an example whenever constants are numerically evaluated.

Theorem 4.1 is the main result of this thesis and will proven in Chapter 5.

4.2.1 Properties

In the master theorem for restart games, Theorem 2.4, there are three important properties of a strategy. The expected length of a loss-conditioned round $\mathrm{E}_{\mathcal{P}}(t \mid G_L)$, the probability of finishing a win-conditioned round $\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W)$, and the expected length of a win-conditioned restart round $\mathrm{E}_{\mathcal{P}}(t \mid G_W, R)$.

For the promising strategy it is obvious that the length of any round is at least $\min(N, \delta)$. In the case of a loss-conditioned round it is also possible to bound the length of a round from above.

Theorem 4.2 (Upper bound on $\mathrm{E}_{\mathcal{P}}(t \mid G_L)$)**.** *In both GCBC($N, \epsilon, r$; p) and Rivest(N, ϵ; p), $\epsilon > 0$, the expected length of a loss-conditioned round*

$$\mathrm{E}_{\mathcal{P}}(t \mid G_L) \leq \frac{\alpha+1}{\epsilon^2}\frac{1}{1-\mathrm{e}^{-2\beta^2\alpha}}$$
$$\{\alpha = 2, \beta = 0.5\} = \frac{1}{\epsilon^2}4.746...$$

Another important property is $\mathrm{P}_{\mathcal{S}}(\neg R \mid G_W)$. The promising strategy plays a substantial amount of time before restarting and because of this is able to keep this probability finite, i.e. if it is possible to win, it will win with probability bounded away from zero.

Theorem 4.3 (Lower bound on $\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W)$)**.** *In both GCBC($N, \epsilon, r$; p) and Rivest(N, ϵ; p), the probability to finish a win-conditioned round*

$$\begin{aligned} \mathrm{P}_{\mathcal{P}}(\neg R \mid G_W) &\geq 1 - \frac{1}{\mathrm{e}^{2(1-\beta)^2\alpha} - 1} \\ \{\alpha = 2, \beta = 0.5\} &= 0.418... \end{aligned}$$

Remark. The bound is only usable if

$$\frac{\ln 2}{2} < (1-\beta)^2\alpha,$$

otherwise the bound on the probability is trivial.

The only property left to analyze is $\mathrm{E}_{\mathcal{P}}(t \mid G_W, R)$. As a consequence of Theorem 4.3 the promising strategy will only play a finite expected number of win-conditioned restart rounds. This in turn means that the trivial bound $\mathrm{E}_{\mathcal{P}}(t \mid G_W, R) \leq N$ is sufficient to do constant-factor from optimal analysis.

Proof of Theorem 4.2 for GCBC(N, ϵ, r; p). One way of thinking of $\mathcal{P}$ is that at every checkpoint there is a buy-in cost to continue playing. This cost is δ except after the last checkpoint where it is $\leq \delta$. This means that the length of a round

$$\begin{aligned} t &\leq \delta(1 + \#\text{buy-ins}) \\ \left\{\delta = \left\lceil \frac{\alpha}{\epsilon^2} \right\rceil, \epsilon < 1\right\} &\leq \frac{\alpha+1}{\epsilon^2}(1 + \#\text{buy-ins}). \end{aligned}$$

A buy-in at checkpoint i requires that you both clear all the checkpoints before i and have enough heads at n_i. To get an upper bound on $\mathrm{E}_{\mathcal{P}}(\#\text{buy-ins} \mid G_L)$ the first requirement can be dropped to get

$$\begin{aligned} \mathrm{E}_{\mathcal{P}}(\#\text{buy-ins} \mid G_L) &\leq \sum_i \mathrm{P}(H_{n_i} \geq (p+\beta\epsilon)n_i \mid H_{n_i} \sim \text{Binomial}(n_i, p)) \\ \{\text{Hoeffding's inequality (A.1)}\} &\leq \sum_i \mathrm{e}^{-2\beta^2\epsilon^2 n_i} \\ \left\{n_i = i\left\lceil \frac{\alpha}{\epsilon^2} \right\rceil \geq i\frac{\alpha}{\epsilon^2}\right\} &\leq \sum_i \mathrm{e}^{-2\beta^2\alpha i}. \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{E}_{\mathcal{P}}(t \mid G_L) &\leq \delta(1 + \mathrm{E}_{\mathcal{P}}(\#\text{buy-ins} \mid G_L)) \\ &\leq \frac{\alpha+1}{\epsilon^2}\left(1 + \sum_i \mathrm{e}^{-2\beta^2\alpha i}\right) \\ &\leq \frac{\alpha+1}{\epsilon^2}\sum_{i=0}^{\infty} \mathrm{e}^{-2\beta^2\alpha i} \\ &= \frac{\alpha+1}{\epsilon^2}\frac{1}{1 - \mathrm{e}^{-2\beta^2\alpha}}. \end{aligned}$$

□

Proof of Theorem 4.2 for Rivest(N, ϵ; p), $\epsilon > 0$. Just note that

$$\mathrm{E}_{\mathcal{P}}(t \mid G_L) \leq \mathrm{E}_{\mathcal{P}}(t).$$

This is because G_L only includes those games where $H_N < K$, while not conditioning on loss will also include those where $H_N \geq K$, Rounds with $H_N \geq K$ will have more heads in general which will make the lengths of the rounds longer.

$\mathrm{E}_{\mathcal{P}}(t)$ is just the expected length of a round played using $\mathcal{P}$ with a p-coin, which is exactly what the previous proof covers. $\square$

Proof of Theorem 4.3 for GCBC(N, ϵ, r; p). It is possible to bound $\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W)$ from below by bounding $\mathrm{P}_{\mathcal{P}}(R \mid G_W)$ from above. $\mathrm{P}_{\mathcal{P}}(R \mid G_W)$ can be interpreted as the probability of doing a restart at any of the possible checkpoints,

$$\mathrm{P}_{\mathcal{P}}(R \mid G_W) = \sum_i \mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W).$$

For the restart to happen at checkpoint i it is required that you clear all checkpoints $< i$ and have too few heads at n_i. To get an upper bound on $\mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W)$ the first requirement can be dropped to get

$$\begin{aligned}
\mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W) &\leq \mathrm{P}(H_{n_i} < (p+\beta\epsilon)n_i \mid H_{n_i} \sim \text{Binomial}(n_i, p+\epsilon)) \\
&= \mathrm{P}(H_{n_i} < ((p+\epsilon) - (1-\beta)\epsilon)n_i \mid H_{n_i} \sim \text{Binomial}(n_i, p+\epsilon)) \\
\{\text{Hoeffding (A.2)}\} &\leq \mathrm{e}^{-2(1-\beta)^2\epsilon^2 n_i} \\
\left\{n_i = i\left\lceil\frac{\alpha}{\epsilon^2}\right\rceil \geq i\frac{\alpha}{\epsilon^2}\right\} &\leq \mathrm{e}^{-2(1-\beta)^2\alpha i}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W) &= 1 - \mathrm{P}_{\mathcal{P}}(R \mid G_W) \\
&= 1 - \sum_i \mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W) \\
&\geq 1 - \sum_i \mathrm{e}^{-2(1-\beta)^2\alpha i} \\
&\geq 1 - \sum_{i=1}^{\infty} \mathrm{e}^{-2(1-\beta)^2\alpha i} \\
&= 1 - \frac{1}{\mathrm{e}^{2(1-\beta)^2\alpha} - 1}.
\end{aligned}$$

$\square$

Proof of Theorem 4.3 for Rivest(N, ϵ; p). For Rivest there are two cases, either $\epsilon \leq 0$ or $\epsilon > 0$. If $\epsilon \leq 0$ then $\mathcal{P}$ is completely indolent and therefore $\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W) = 1$.

If $\epsilon > 0$ then the proof is almost the same as for GCBC, the only difference is the method used to show the bound on $\mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W)$. From the same argument as before

$$\mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W) \leq \mathrm{P}(H_{n_i} < (p+\beta\epsilon)n_i \mid G_W).$$

G_W cointains many different games, all with $H_N \geq K$, i.e. the number of heads at the end of the game must result in a win. Now note that the probability on the RHS is maximized when $H_N = K$ which will make H_n hypergeometrically distributed. So, letting $H_{n_i} \sim \mathrm{HyperGeom}(K, N, n_i)$, we have

$$\begin{aligned} \mathrm{P}_{\mathcal{P}}(\text{restart at } n_i \mid G_W) &\leq \mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i) \\ &= \mathrm{P}(H_{n_i} < ((p + \epsilon) - (1 - \beta)\epsilon)n_i) \\ \{\text{Hoeffding (A.4)}\} &\leq \mathrm{e}^{-2(1-\beta)^2\epsilon^2 n_i}. \end{aligned}$$

This is the same exact bound as for GCBC. ☐

4.3 Probabilistic strategy

Unlike the promising strategy, the probabilistic strategy is very general and can be played on any restart game.

Definition 4.2 (Probabilistic strategy $\mathcal{Pr}$ for any Restart(G))**.** $\mathcal{Pr}$ takes in one parameter, $\lambda \in (0, 1)$. It restarts at a state iff the probability of winning the base game playing from that state is $< \lambda \mathrm{P}(G_W)$.

It turns out that this strategy behaves similarly to the promising strategy, to the extent that Theorem 4.1 can be used to prove that the probabilistic strategy is constant-factor from optimal.

Theorem 4.4 (Probabilistic strategy is constant-factor from optimal)**.** *For GCBC$(N, \epsilon, r;\, p)$ and Rivest$(N, \epsilon;\, p)$, the probabilistic strategy $\mathcal{Pr}$ is for any choice of λ constant-factor from optimal, i.e. $\mathrm{E}_{\mathcal{Pr}}(T)$ is at most a constant factor from $\mathrm{E}^*(T)$ (constant w.r.t. N, ϵ and r; but not w.r.t. p and λ).*

The first similar behavior is that both strategies only restart on a fraction of the win-conditioned rounds. For the promising strategy this is stated in Theorem 4.3, and a similar theorem can be stated for the probabilistic strategy.

Theorem 4.5 (Upper bound on $\mathrm{P}(R \mid G_W)$ for the probabilistic strategy)**.** *For any restart game Restart(G) such that $\mathrm{P}(G_W) > 0$ it holds that*

$$\mathrm{P}_{\mathcal{Pr}}(R \mid G_W) \leq \lambda.$$

Proof of Theorem 4.5. Note that for $\mathcal{Pr}$

$$\mathrm{P}_{\mathcal{Pr}}(G_W \mid R) \leq \lambda \mathrm{P}(G_W). \tag{4.1}$$

This is because the probabilistic strategy only restarts if the probability of the round being a winning round is lower than λ times what it was initially. The condition for restarting does not depend on the future which makes the time of restart a stopping time, so (4.1) follows from the strong Markov property.

Combining inequality (4.1) with Bayes' theorem it follows that

$$\begin{aligned}
\mathrm{P}_{\mathcal{Pr}}(R \mid G_W) &= \frac{\mathrm{P}_{\mathcal{Pr}}(G_W \mid R)}{\mathrm{P}(G_W)} \mathrm{P}_{\mathcal{Pr}}(R) \\
&\leq \lambda \mathrm{P}_{\mathcal{Pr}}(R) \\
&\leq \lambda.
\end{aligned}$$

□

The other similarity between $\mathcal{P}$ and $\mathcal{Pr}$ is that the bound in Theorem 4.2 can also be used to bound the probabilistic strategy. This is because that for appropriate choices of α and β, the probabilistic strategy is more aggressive than the promising strategy.

Theorem 4.6 (Probabilistic is more aggressive than promising)**.** *Valid for GCBC*$(N, \epsilon, r \geq 1/2; p)$ *and Rivest*$(N, \epsilon; p)$. *Let $\mathcal{Pr}$ be the probabilistic strategy with parameter λ, and $\mathcal{P}$ be the promising strategy with parameters α and β such that*

$$4\mathrm{e}^{-2(1-\beta)^2\alpha} \leq \lambda.$$

If $\mathcal{P}$ restarts at a state (n, H_n) then $\mathcal{Pr}$ also restarts at (n, H_n).

Proof of Theorem 4.6. To show that $\mathcal{P}$ restarting implies that $\mathcal{Pr}$ restarts, it is necessary to show that when $\mathcal{P}$ restarts, the conditional probability of winning is $< \lambda \mathrm{P}(G_W)$.

If $\epsilon \leq 0$ then promising strategy never restarts, so the statement in the theorem is trivially true. If $\epsilon > 0$ then as seen in the proof of Theorem 4.3

$$\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i \mid G_W) \leq \mathrm{e}^{-2(1-\beta)^2\alpha}$$

holds for both games. From Bayes' theorem it follows that

$$\begin{aligned}
\mathrm{P}(G_W \mid H_{n_i} < (p + \beta\epsilon)n_i) &= \frac{\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i \mid G_W)}{\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i)} \mathrm{P}(G_W) \\
&\leq \frac{\mathrm{e}^{-2(1-\beta)^2\alpha}}{\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i)} \mathrm{P}(G_W).
\end{aligned}$$

The denominator needs to be bounded from below in different ways depending on the game.

For Rivest it holds that

$$\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i) \geq \frac{1}{2}$$

as the bound $(p + \beta\epsilon)n_i$ is above the median of H_{n_i}.

For GCBC it holds that

$$\begin{aligned}
\mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i) &\geq \mathrm{P}(H_{n_i} < (p + \beta\epsilon)n_i \mid G_L)\mathrm{P}(G_L) \\
&\geq \frac{1}{4}
\end{aligned}$$

as the probability of playing with a bad coin $r = \mathrm{P}(G_L)$ is assumed to be at least $1/2$, and $(p + \beta\epsilon)n_i$ is above the median of H_{n_i} when playing with a bad coin.

So for both Rivest and GCBC we have that

$$\begin{aligned} \mathrm{P}(G_W \mid H_{n_i} < (p+\beta\epsilon)n_i) &\le \frac{\mathrm{e}^{-2(1-\beta)^2\alpha}}{\mathrm{P}(H_{n_i} < (p+\beta\epsilon)n_i)} \mathrm{P}(G_W) \\ &\le 4\mathrm{e}^{-2(1-\beta)^2\alpha}\mathrm{P}(G_W). \end{aligned}$$

Now using the assumption $4\mathrm{e}^{-2(1-\beta)^2\alpha} < \lambda$ gives

$$\mathrm{P}(G_W \mid H_{n_i} < (p+\beta\epsilon)n_i) < \lambda \mathrm{P}(G_W).$$

This means that the probabilistic strategy would have restarted if the promising strategy had restarted. $\square$

Now, having proved that the probabilistic strategy is more aggressive than the promising strategy, it is possible to prove Theorem 4.4 using Theorem 4.1.

Proof of Theorem 4.4. Note that for GCBC$(N, \epsilon, r;\, p)$ with $r < \frac{1}{2}$, the probabilistic strategy is constant-factor from optimal since

- $r < \frac{1}{2}$ implies at least half of the rounds are win-conditioned rounds
- Theorem 4.5 implies that the expected number of win-conditioned rounds played by $\mathcal{Pr}$ is bounded by a constant.

So we can assume $r \ge 1/2$ in the future, which is necessary in order to be able to use Theorem 4.6.

What remains is analyzing the terms in the master theorem for restart games, Theorem 2.4, and showing that $\mathrm{E}_{\mathcal{Pr}}(T)$ is bounded from above by a constant times $\mathrm{E}_{\mathcal{P}}(T)$. The three terms describe the expected cost of loss-conditioned rounds, the expected cost of restarted win-conditioned rounds, and the expected cost of the winning round.

Let X denote the sum of the last two terms. Note that for GCBC and Rivest, independent of strategy, $X \ge N$. For $\mathcal{Pr}$, Theorem 4.5 implies that X is bounded from above by a constant times N. This means that $\mathcal{Pr}$ is no more than a constant factor from minimizing X, so X can be ignored.

Only the first term remains, and the only part of it that is strategy dependent is the factor

$$\frac{\mathrm{E}(t \mid G_L)}{\mathrm{P}(\neg R \mid G_W)}.$$

Theorem 4.6 says that $\mathcal{Pr}$ is more aggressive than $\mathcal{P}$, so the numerator $\mathrm{E}_{\mathcal{Pr}}(t \mid G_L) \le \mathrm{E}_{\mathcal{P}}(t \mid G_L)$. For $\mathcal{Pr}$, Theorem 4.5 implies that the denominator is bounded away from zero. This means that the ratio (and as a consequence the entire term) cannot be more than a constant factor larger for $\mathcal{Pr}$ than for $\mathcal{P}$.

So $\mathcal{Pr}$ is constant-factor from $\mathcal{P}$. Since $\mathcal{P}$ is constant-factor from optimal according to Theorem 4.1, it follows that $\mathcal{Pr}$ is constant-factor from optimal. $\square$

4.4 Properties of an optimal strategy

There are some properties that any optimal DM restart strategy must have playing Rivest and GCBC. One obvious property is that the optimal strategy must restart immediately if the strategy is not able to win the current round no matter the outcome. Another more subtle property is that if it restarts in state (n, H_n) must also restart in $(n, H_n - 1)$.

Definition 4.3 (Consistent strategy)**.** A DM restart strategy $\mathcal{S}$ for Rivest$(N, \epsilon; p)$ or GCBC$(N, \epsilon, r; p)$ is said to be *consistent* if for each $n < N$ there exists a *cutoff point* $c_n \in \{0, \ldots, n\}$, such that $\mathcal{S}$ restarts at state (n, H_n) iff $H_n < c_n$.

The indolent strategy is the most trivial example of a consistent strategy. Note that out of all consistent strategies, indolent is the one that keeps the worst rounds. Every other consistent strategy will try to filter out the bad rounds. So given that n coins has been tossed, out of all consistent strategies, indolent minimizes the number of heads up to that point. Formally this can be stated as the following.

Theorem 4.7 (Indolent has the fewest number of heads)**.** *Valid for Rivest*$(N, \epsilon; p)$ *and GCBC*$(N, \epsilon, r; p)$*. Let* $n \in \{0, \ldots, N\}$*,* $k \in \{0, \ldots, n\}$ *and* $\mathcal{S}$ *be any consistent strategy. Then*

$$\mathrm{P}_{\mathcal{S}}(H_n \geq k \mid \neg R) \geq \mathrm{P}_{I}(H_n \geq k \mid \neg R).$$

Proof of Theorem 4.7. While the theorem is intuitively correct, it is possible to formally show it using the Fortuin–Kasteleyn–Ginibre (FKG) inequality [6]. The inequality states that if A and B are two *increasing events* then they are positively correlated, i.e.

$$\mathrm{P}(A \cap B) \geq \mathrm{P}(A)\mathrm{P}(B),$$

or equivalently

$$\mathrm{P}(A \mid B) \geq \mathrm{P}(A).$$

Note that the statement in the theorem is equivalent to

$$\mathrm{P}_{\mathcal{S}}(H_n \geq k \mid G_W, \neg R) \geq \mathrm{P}(H_n \geq k \mid G_W).$$

which is exactly

$$\mathrm{P}(A \mid B, G_W) \geq \mathrm{P}(A \mid G_W),$$

where A is the event that $H_n \geq k$ and B is the event that $\mathcal{S}$ does not restart. So, to apply FKG all that is necessary to show is that A and B are increasing events.

That A and B are increasing events follows directly from the definition in [6] and that $\mathcal{S}$ is consistent. A is an increasing event since if you take an outcome in A and replace a tail with a head, it will still lie in A. The same also holds for B. □

Theorem 4.7 is a fundamental inequality for all consistent strategies playing Rivest, and as such connects any optimal DM restart strategy to the indolent strategy. This will be used in the proof of Theorem 4.1 to allow certain bounds on the indolent strategy to be used on any optimal DM restart strategy.

4.5 Other strategies

There are three extra strategies considered in the numerical analysis.

Optimist
: Only restarts if there is 0 probability of winning

Promising ($\boldsymbol{\delta = 1}$)
: A degenerate case of the promising strategy with distance 1 between checkpoints

5

Optimality of promising

In this section we will prove Theorem 4.1, the theorem stating that the promising strategy $\mathcal{P}$ is constant-factor from optimal playing either GCBC$(N, \epsilon, r;\, p)$ or Rivest$(N, \epsilon;\, p)$. For simplicity we will prove the statement for $\mathcal{P}$ with parameters $\alpha = 2$ and $\beta = 0.5$, but the exact same proof works for any α and β such that

$$\frac{\ln 2}{2} < (1-\beta)^2 \alpha$$

and will only result in a different numerical constant.

To simplify notation, anytime c or c_i is used it refers to a positive constant (constant w.r.t. N, ϵ and r; but not w.r.t. p).

Definition 5.1 (Constant-factor from optimal)**.** A DM strategy $\mathcal{S}$ is said to be constant-factor from optimal in either game if

$$\mathrm{E}_{\mathcal{S}}(T) \leq c\,\mathrm{E}^*(T).$$

The method used to show that a strategy is constant-factor from optimal is based on the following theorem

Theorem 5.1. *Valid for strategy $\mathcal{S} = \mathcal{P}$ or $\mathcal{S} = \mathcal{I}$. Given GCBC$(N, \epsilon, r;\, p)$ or Rivest$(N, \epsilon;\, p)$, if for some restriction of the (N, ϵ) parameter space there exists a DM restart strategy $\mathcal{S}'$ such that*

$$\mathrm{E}_{\mathcal{S}}(t \mid G_L) \leq c_1 \mathrm{E}_{\mathcal{S}'}(t \mid G_L),$$

then for that restriction there exists a c_2 (dependent on c_1) such that

$$\mathrm{E}_{\mathcal{S}}(T) \leq c_2 \mathrm{E}_{\mathcal{S}'}(T).$$

Proof of Theorem 5.1. The theorem follows from analysis of the parameters in the master theorem for restart games, Theorem 2.4.

Note that the indolent strategy $\mathcal{I}$ has $\mathrm{P}_{\mathcal{I}}(\neg R \mid G_W) = 1$ and in the case of the promising strategy $\mathcal{P}$, Theorem 4.3 states that $\mathrm{P}_{\mathcal{P}}(\neg R \mid G_W)$ is bounded from below by a positive constant. Essentially this means that both strategies are a constant-factor away from maximizing $\mathrm{P}(\neg R \mid G_W)$, so no other strategy could improve this parameter by more than a constant factor.

The expression for $\mathrm{E}_{\mathcal{S}}(T)$ in Theorem 2.4 contains three terms. The third term $\mathrm{E}_{\mathcal{S}}(t \mid \neg R) = N$ in the case of GCBC and Rivest. The second term for S is $\leq cN$ as

$\mathrm{P}_{\mathcal{S}}(\neg R \mid G_W)$ is bounded away from zero and trivially $\mathrm{E}_{\mathcal{S}}(t \mid G_W, R) \leq N$. So the only way for c_2 to not exist, i.e. $\mathrm{E}_{\mathcal{S}'}(T) \ll \mathrm{E}_{\mathcal{S}}(T)$, is for the first term for S' to be much smaller than the first term for S, which can only happen if $\mathrm{E}_{\mathcal{S}'}(t \mid G_L) \ll \mathrm{E}_{\mathcal{S}}(t \mid G_L)$. This would contradict the existence of c_1. $\square$

Note that this theorem allows for comparison between strategies. For example because $\mathrm{E}_I(t \mid G_L) = N \geq \mathrm{E}_{\mathcal{P}}(t \mid G_L)$ it follows from Theorem 5.1 that

$$\mathrm{E}_{\mathcal{P}}(T) \leq c\mathrm{E}_I(T).$$

One implication from this is that if indolent is constant-factor from optimal for some subset of the games, then the same is true for promising. A generalized version of this result is stated in the following corollary.

Corollary 5.2 (Corollary of Theorem 5.1)**.** *Valid for strategy $\mathcal{S} = \mathcal{P}$ or $\mathcal{S} = I$. Given GCBC($N, \epsilon, r;\, p$) or Rivest($N, \epsilon;\, p$), if for some restriction of the (N, ϵ) parameter space $\mathcal{S}'$ is constant-factor from optimal and $\mathrm{E}_{\mathcal{S}}(t \mid G_L) \leq c\,\mathrm{E}_{\mathcal{S}'}(t \mid G_L)$ then $\mathcal{S}$ is also constant-factor from optimal.*

This corollary is the most important tool for constant-factor from optimal analysis. The idea is to show that $\mathcal{P}$ is constant-factor from optimal by showing that there exists some constant-factor from optimal strategy $\mathcal{S}'$ such that $\mathrm{E}_{\mathcal{S}}(t \mid G_L) \leq c\,\mathrm{E}_{\mathcal{S}'}(t \mid G_L)$. It then follows from Corollary 5.2 that $\mathcal{P}$ is also constant-factor from optimal. This means that most of the proof of Theorem 4.1 will be about showing the existence of $\mathcal{S}'$, a strategy that is constant-factor from optimal while also playing loss-conditioned rounds of roughly the same length as $\mathcal{P}$.

5.1 Regions

The proof of $\mathcal{P}$ being constant-factor from optimal for GCBC($N, \epsilon, r;\, p$) and Rivest($N, \epsilon;\, p$) is different depending on N and ϵ. For $\epsilon > 0$ there are 3 regions

1. $\frac{1-p}{4} < \epsilon$
2. $\frac{1-p}{4} \geq \epsilon$, $N < \frac{4}{\epsilon^2}$
3. $\frac{1-p}{4} \geq \epsilon$, $N \geq \frac{4}{\epsilon^2}$

The first two regions deal with boundary cases. Region 1 is the case where $p + \epsilon \approx 1$ and region 2 is the case where N is small in comparison to $\frac{1}{\epsilon^2}$. Note that the constants (the fours in the enumeration) used are not entirely arbitrary. They are chosen so that bounds on probabilities come out correct. The constants could be fine-tuned to make the final constant factor (stated in Theorem 5.11) be a lot smaller, but here the constants are taken to be integer to make the proofs nicer.

Theorem 5.3 (Valid for GCBC and Rivest, region 1)**.** *Promising is constant-factor from optimal.*

Theorem 5.4 (Valid for GCBC and Rivest, region 2)**.** *Indolent is constant-factor from optimal.*

Theorem 5.5 (Valid for GCBC and Rivest, region 3)**.** *Promising is constant-factor from optimal.*

Using these three theorems it is possible to prove Theorem 4.1. So the focus from this point onwards will be to prove these three theorems, containing six statements. It will be done in four proofs, two of them will be mostly trivial and will be done in this section. The other two will need more advanced techniques and will be proven in the next section.

Proof of the optimality theorem, Theorem 4.1. The three regions covers almost all possible choices of N and ϵ. The only exception is $\epsilon \leq 0$ in the case of Rivest, but that case is trivial as $\mathcal{P}$ will be indolent by definition and from inequality (3.1) it follows that the expected number of rounds in the game is at most 2. So for $\epsilon \leq 0$ $\mathcal{P}$ is clearly constant-factor from optimal, so we only have to take into account $\epsilon > 0$.

Combining Theorem 5.4 and Corollary 5.2 (with $\mathcal{S} = \mathcal{P}$ and $\mathcal{S}' = I$) implies $\mathcal{P}$ is constant-factor from optimal in region 2, which together with Theorems 5.3 and 5.5 implies $\mathcal{P}$ is constant-factor from optimal. ☐

Proof of Theorem 5.3, GCBC and Rivest, region 1. In region 1 $\frac{1-p}{4} < \epsilon$ so Theorem 4.2 gives that

$$\mathrm{E}_{\mathcal{P}}(t \mid G_L) \leq \frac{4^2}{(1-p)^2} 4.746...,$$

which is a constant. Together with Corollary 5.2 (with $\mathcal{S} = \mathcal{P}$ and $\mathcal{S}'$ being an optimal strategy) it follows that $\mathcal{P}$ is constant-factor from optimal. ☐

The following is a proof of Theorem 5.4 in the case of Rivest. The corresponding proof for GCBC will be done in the next section.

Proof of Theorem 5.4, Rivest, region 2. For Rivest$(N, \epsilon; p)$

$$\mathrm{P}(G_W) = \mathrm{P}(H_N \geq (p+\epsilon)N), \quad \text{where } H_N \sim \mathrm{Binomial}(N, p),$$

which can be bounded independent of ϵ as

$$\begin{aligned} \mathrm{P}(H_N \geq (p+\epsilon)N) &= \mathrm{P}\left(\frac{H_N - pN}{\sqrt{Np(1-p)}} \geq \frac{\epsilon\sqrt{N}}{\sqrt{p(1-p)}}\right) \\ \left\{\text{Region 2} \implies N < \frac{4}{\epsilon^2}\right\} &\geq \mathrm{P}\left(\frac{H_N - pN}{\sqrt{Np(1-p)}} \geq \frac{2}{\sqrt{p(1-p)}}\right). \end{aligned}$$

Note that for large enough N this is bounded from below by a positive constant as CLT yields

$$\mathrm{P}\left(\frac{H_N - pN}{\sqrt{Np(1-p)}} \geq \frac{2}{\sqrt{p(1-p)}}\right) \xrightarrow{N\to\infty} 1 - \Phi\left(\frac{2}{\sqrt{p(1-p)}}\right) > 0.$$

For finite N there are only finitely many choices of (N, K), and by definition of Rivest $\mathrm{P}(G_W)$ is always > 0. Taken together, this means that for the entirety of region 2

$$\mathrm{P}(G_W) \geq c.$$

This in turn implies that in region 2 $\mathcal{I}$ will finish the game in a finite expected number of rounds, independent of N and ϵ, and is thus constant factor from optimal. $\square$

5.2 Box of forced play

It is left to show that indolent is constant-factor from optimal for region 2 for GCBC, and that promising is constant-factor from optimal in region 3 for GCBC and Rivest. In order to do these proofs we will need to construct a constant-factor from optimal strategy that plays loss-conditioned rounds of similar length as the promising strategy. The idea is to enforce play for some subset of the states, and then using

- Theorem 2.7 to show there exists a constant-factor from optimal strategy that abides by these rules.
- Kolmogorov's inequality to show that the enforcement of play will lead to playing rounds of similar length as the promising strategy.

The subset of states chosen to enforce play on is the following box.

Definition 5.2 (Box of forced play)**.** For GCBC$(N, \epsilon, r; p)$ and Rivest$(N, \epsilon; p)$, $\epsilon > 0$, define the *box of forced play* to be the set of states (n, H_n) such that

$$n \leq \frac{1}{2\epsilon^2} \quad \text{and} \quad H_n - p\,n \geq -\frac{1}{2\epsilon}.$$

Lemma 5.6 (Strategy forced to play in box)**.** *Valid for region* 2 *and* 3 *in GCBC and region* 3 *in Rivest. There exists a constant-factor from optimal strategy that plays for all states in the box.*

In the rest of this section we will use Lemma 5.6 to prove all remaining optimality theorems. The lemma itself will be proven in the next section. To help break down the arguments, we will first state and prove the following two propositions, and then use them in order to prove the remaining optimality theorems.

Proposition 5.7. *Valid for region* 2 *in GCBC*$(N, \epsilon, r; p)$*. There exists a constant-factor from optimal strategy* $\mathcal{S}$ *and a constant* c *such that*

$$\mathrm{E}_{\mathcal{S}}(t \mid G_L) \geq c\,N.$$

Proposition 5.8. *Valid for region* 3 *in GCBC*$(N, \epsilon, r; p)$ *and region* 3 *in Rivest*$(N, \epsilon; p)$*. There exists a constant-factor from optimal strategy* $\mathcal{S}$ *and a constant* c *such that*

$$\mathrm{E}_{\mathcal{S}}(t \mid G_L) \geq \frac{c}{\epsilon^2}.$$

Definition 5.3 (Stuck in box)**.** Following the rule of playing inside the box, for a round define *stuck in box* as the event that

$$H_n - p\,n \geq -\frac{1}{2\epsilon}, \ \forall (n, H_n) \in \{(k, H_k)\}_{k=0}^{t} : n \leq \frac{1}{2\epsilon^2}.$$

Proof of Proposition 5.8, GCBC, region 3. Lemma 5.6 states that there exists a constant-factor from optimal strategy $\mathcal{S}$ that plays everywhere inside the box of forced play. Note that $N > \frac{1}{2\epsilon^2}$ in region 3 so being stuck in box means playing at least $\frac{1}{2\epsilon^2}$ times. So to show that $\mathrm{E}_{\mathcal{S}}(t \mid G_L) \geq \frac{c}{\epsilon^2}$ all that is needed is to show that $\mathrm{P}(\text{stuck in box} \mid G_L)$ is bounded away from zero.

Using Kolmogorov's inequality (A.5) it is possible to show that the probability of the complementary event, $1 - \mathrm{P}(\text{stuck in box} \mid G_L)$, is

$$\begin{aligned}
\mathrm{P}\left(\min_{n \leq \frac{1}{2\epsilon^2}} (H_n - p\,n) < -\frac{1}{2\epsilon}\right) &\leq 4\epsilon^2 \frac{1}{2\epsilon^2} p(1-p) \\
&= 2p(1-p) \\
&\leq \frac{1}{2},
\end{aligned}$$

so $\mathrm{P}(\text{stuck in box} \mid G_L) \geq \frac{1}{2}$. ∎

Proof of Proposition 5.8, Rivest, region 3. Lemma 5.6 states that there exists a constant-factor from optimal strategy $\mathcal{S}$ that plays everywhere inside the box of forced play. The goal is to show that $\mathrm{P}(\text{stuck in box} \mid G_L)$ is bounded away from zero, which would imply that $\mathrm{E}_{\mathcal{S}}(t \mid G_L) \geq \frac{c}{\epsilon^2}$ for some constant c.

Using Bayes' formula it follows that

$$\begin{aligned}
\mathrm{P}(\text{stuck in box} \mid G_L) &= \frac{\mathrm{P}(\text{stuck in box}) - \mathrm{P}(\text{stuck in box} \mid G_W)\mathrm{P}(G_W)}{\mathrm{P}(G_L)} \\
&\geq \mathrm{P}(\text{stuck in box}) - \mathrm{P}(G_W)
\end{aligned}$$

The same argument as in the proof of Proposition 5.8 for GCBC can also be applied to Rivest (the analysis only deals with p-coins). Thus we have that $\mathrm{P}(\text{stuck in box}) \geq \frac{1}{2}$.

Now note that in region 3, letting $H_N \sim \text{Binomial}(N, p)$,

$$\begin{aligned}
\mathrm{P}(G_W) &= \mathrm{P}(H_N \geq (p+\epsilon)N) \\
\{\text{Hoeffding's inequality (A.1)}\} &\leq \mathrm{e}^{-2\epsilon^2 N} \\
\{\text{Region 3} \implies N \geq \frac{4}{\epsilon^2}\} &\leq \mathrm{e}^{-8}.
\end{aligned}$$

In conclusion

$$\mathrm{P}(\text{stuck in box} \mid G_L) \geq \frac{1}{2} - \mathrm{e}^{-8},$$

which is bounded away from zero. ∎

Proof of Proposition 5.7, GCBC, region 2. Lemma 5.6 states that there exists a constant-factor from optimal strategy $\mathcal{S}$ that plays everywhere inside the box of forced play. Note that $\frac{N}{8} \leq \frac{1}{2\epsilon^2}$ in region 2 so being stuck in box means playing at least $\frac{N}{8}$ times. So if it is shown that $\mathrm{P}(\text{stuck in box} \mid G_L)$ is bounded away from zero, then it follows that $\mathrm{E}_{\mathcal{S}}(t \mid G_L) \geq cN$ for some constant c.

Note that region 2 and region 3 are very similar for a fixed ϵ. The difference is that region 2 contains small N ($N < \frac{4}{\epsilon^2}$) and region 3 contains large N ($N \geq \frac{4}{\epsilon^2}$).

It was shown in the proof of Proposition 5.8 for GCBC in region 3 (large N) that $\mathrm{P}(\text{stuck in box} \mid G_L) \geq \frac{1}{2}$. For fixed ϵ, $\mathrm{P}(\text{stuck in box} \mid G_L)$ is a decreasing function as a function of N, as it is more probable to get stuck in a narrower box. So the same bound on $\mathrm{P}(\text{stuck in box} \mid G_L)$ is also valid in region 2 (small N). $\square$

Combining these propositions with Corollary 5.2 it is possible to prove the remaining optimality theorems.

Proof of Theorem 5.4, GCBC, region 2. Proposition 5.7 states that there exists a constant-factor from optimal strategy $\mathcal{S}'$ such that $\mathrm{E}_{\mathcal{S}'}(t \mid G_L) \geq c\,N$. Since $\mathrm{E}_{I}(t \mid G_L) \leq N$ it follows from Corollary 5.2 (with $\mathcal{S} = I$ and $\mathcal{S}' = \mathcal{S}'$) that I must also be constant-factor from optimal. $\square$

Proof of Theorem 5.5, GCBC and Rivest, region 3. Proposition 5.8 states that there exists a constant-factor from optimal strategy $\mathcal{S}'$ such that $\mathrm{E}_{\mathcal{S}'}(t \mid G_L) \geq \frac{c}{\epsilon^2}$. Since one of the properties of $\mathcal{P}$, given by Theorem 4.2, is that $\mathrm{E}_{\mathcal{P}}(t \mid G_L) \leq \frac{1}{\epsilon^2}4.746...$ it follows from Corollary 5.2 (with $\mathcal{S} = \mathcal{P}$ and $\mathcal{S}' = \mathcal{S}'$) that $\mathcal{P}$ must also be constant-factor from optimal. $\square$

5.3 Grades inside box

This section aims to prove the only remaining result, Lemma 5.6. First note that Lemma 5.6 can directly be proven by combining Theorem 2.7 with the following lemma.

Lemma 5.9 (States inside the box of forced play are not too bad)**.** *Valid for region* 2 *and* 3 *in GCBC and region* 3 *in Rivest. The grade of states inside the box of forced play is at most* $c\,\mathrm{E}^*(T)$.

To prove Lemma 5.9 we need to analyze the grade of states inside the box in the different games. It happens to be that it is more natural to show this statement for a bigger box.

Definition 5.4 (The extended box)**.** For $\mathrm{GCBC}(N, \epsilon, r;\, p)$ and $\mathrm{Rivest}(N, \epsilon;\, p)$, $\epsilon > 0$, define the *extended box* to be the set of states (n, H_n) such that

$$n \leq \frac{1}{2\epsilon^2} \quad \text{and} \quad H_n - (p + \epsilon)\, n \geq -\frac{1}{\epsilon}.$$

Lemma 5.10 (States inside the extended box are not too bad)**.** *Valid for region* 2 *and* 3 *in GCBC and region* 3 *in Rivest. The grade of states inside the extended box is at most* $c\,\mathrm{E}^*(T)$.

Proof of Lemma 5.9. Simply note that the box of forced play fits inside the extended box as the two boxes have the same width $n \leq \frac{1}{2\epsilon^2}$ and for the big box

$$H_n - np \geq -\frac{1}{\epsilon} + \epsilon n$$

and for the small box

$$H_n - np \geq -\frac{1}{2\epsilon}.$$

Now note that because $n \leq \frac{1}{2\epsilon^2}$

$$-\frac{1}{\epsilon} + \epsilon n \leq -\frac{1}{2\epsilon},$$

so the height constraint for the box of forced play is more strict than the constraint for the extended box. □

The rest of this section will only be to prove Lemma 5.10 as this is the only result yet to be proven. The proof for the two games are very different, GCBC (in region 2 and 3) will be basic and relatively short, while the proof for Rivest (in region 3) will require a lot of calculations and bounds.

5.3.1 GCBC

The proof of Lemma 5.10 for GCBC will be based on the idea that when you are inside the extended box, the conditional probability of having a bad coin given history will not be much higher than what it was initially. This in turn will imply that the grade of every state inside the extended box are also not much higher than for the initial state.

Proof of Lemma 5.10 for region 2 and 3 for GCBC($N, \epsilon, r; p$). The grade of a state (n, H_n) is defined as playing a modified version of the game with (n, H_n) as the initial state. Note that this is in itself a GCBC game but with a smaller N and different $r = \mathrm{P}(G_L)$ (the probability of being given a bad coin). Clearly N being smaller just makes the game easier, so the focus will be on how $\mathrm{P}(G_L)$ changes given history.

It turns out that considering the ratio $\frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)}$ is more natural than considering $\mathrm{P}(G_L)$ directly. Using Bayes' theorem it is possible to bound how this ratio changes from knowing that the first n tosses resulted in $H_n = k$ heads.

$$\begin{aligned}
\frac{\mathrm{P}(G_L \mid H_n = k)}{\mathrm{P}(G_W \mid H_n = k)} &= \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \frac{\mathrm{P}(H_n = k \mid G_L)}{\mathrm{P}(H_n = k \mid G_W)} \\
&= \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k} (p+\epsilon)^k (1-p-\epsilon)^{n-k}} \\
&= \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \left(\frac{p}{p+\epsilon}\right)^k \left(\frac{1-p}{1-p-\epsilon}\right)^{n-k} \\
&= \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \left(1 - \frac{\epsilon}{p+\epsilon}\right)^k \left(1 + \frac{\epsilon}{1-p-\epsilon}\right)^{n-k} \\
\{1 + x \leq \mathrm{e}^x\} &\leq \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(-\frac{\epsilon}{p+\epsilon} k + \frac{\epsilon}{1-p-\epsilon}(n-k)\right) \\
&= \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(-\frac{\epsilon}{(p+\epsilon)(1-p-\epsilon)}(k - n(p+\epsilon))\right)
\end{aligned}$$

Assuming the state (n, k) is inside the extended box in region 2 or 3, this ratio can be bounded further as

$$\begin{aligned}
\frac{\mathrm{P}(G_L \mid H_n = k)}{\mathrm{P}(G_W \mid H_n = k)} &\leq \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(-\frac{\epsilon}{(p+\epsilon)(1-p-\epsilon)}(k - n(p+\epsilon))\right) \\
\left\{k - n(p+\epsilon) \geq -\frac{1}{\epsilon}\right\} &\leq \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(\frac{1}{(p+\epsilon)(1-p-\epsilon)}\right) \\
\left\{p + \epsilon \geq p, \frac{1-p}{4} \geq \epsilon\right\} &\leq \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(\frac{4}{3} \frac{1}{p(1-p)}\right).
\end{aligned}$$

The conclusion is that inside the extended box

$$\frac{\mathrm{P}(G_L \mid H_n)}{\mathrm{P}(G_W \mid H_n)} \leq \frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)} \exp\left(\frac{4}{3} \frac{1}{p(1-p)}\right).$$

This in turn implies that for any state (n, H_n) inside the box

$$\gamma(n, H_n) \leq \gamma(0,0) \exp\left(\frac{4}{3} \frac{1}{p(1-p)}\right).$$

This is because the ratio $\frac{\mathrm{P}(G_L)}{\mathrm{P}(G_W)}$ is the only parameter in the master theorem for restart games, Theorem 2.4, that could increase had you used the same strategy playing the modified game, and that increase is bounded by the factor $\exp\left(\frac{4}{3} \frac{1}{p(1-p)}\right)$. $\square$

5.3.2 Rivest

This section will prove the only remaining unproven result, Lemma 5.10 for Rivest, by using a bound on the indolent strategy. This bound on indolent will be proven in the next section.

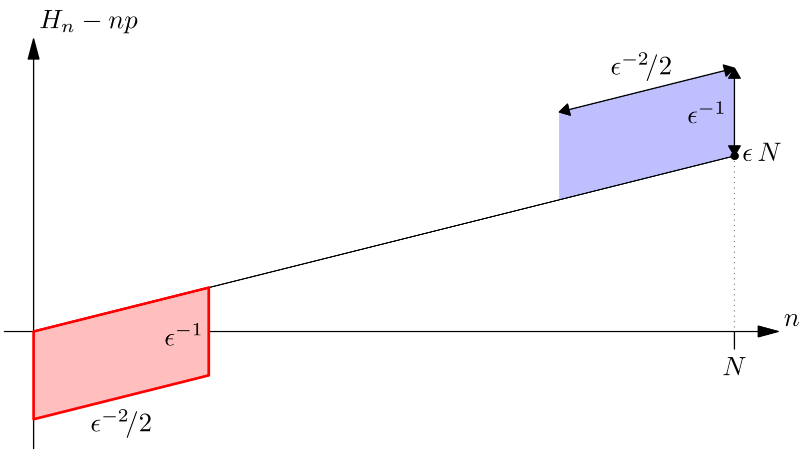


Figure 5.1: Translation of the lower part of the extended box. The lower part of the extended box is drawn in red in the bottom-left, and the translated box is drawn in blue in the top-right. Note that winning from the bottom-right state in the extended box is the same as starting at the origin and reaching the top-left corner of the translated box (or higher).

The idea of the proof of Lemma 5.10 is that winning a round from playing in any state inside the extended box is the same thing as playing Rivest but having a shorter game and needing more heads. Consider translating the lower part of the extended box from the origin to the end of the game, see Figure 5.1. If when playing from the origin you pass above ($\geq$) the translated box, then you would have won playing from anywhere inside the extended box.

First we will state a theorem about this for the indolent strategy and then show that the same holds for optimal strategies as well.

Theorem 5.11 (Indolent wins earlier and with a margin)**.** *Valid in region 3 for Rivest$(N, \epsilon; p)$. For the indolent strategy I and n such that $N - \frac{1}{2\epsilon^2} \leq n \leq N$ it holds that*

$$\mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \,\middle|\, \neg R\right) \geq \left(1 - \mathrm{e}^{-\frac{9}{16}}\right) \frac{1}{3} \mathrm{e}^{-\frac{9.36}{p(1-p)}}.$$

There are two reasons why this theorem is only stated for the indolent strategy. One is that it is possible to show it by just doing analysis on a random walk without involving any restarts. The second reason is that the indolent strategy is the extreme case of a consistent strategy (Definition 4.3), and by Theorem 4.7 the bound will hold for any consistent strategy. This means that Theorem 5.11 combined with Theorem 4.7 can be used to show a lower bound on the grades inside the extended box.

Proof of Lemma 5.10 in region 3 *for Rivest*$(N, \epsilon; p)$*.* Let $\mathcal{S}^\star$ be an optimal DM strategy in Rivest. W.l.o.g. $\mathcal{S}^\star$ can be assumed to be consistent. $\mathcal{S}^\star$ being consistent means that according to Theorem 5.11 and Theorem 4.7

$$\mathrm{P}_{\mathcal{S}^\star}\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \,\middle|\, \neg R\right) \geq \left(1 - \mathrm{e}^{-\frac{9}{16}}\right) \frac{1}{3} \mathrm{e}^{-\frac{9.36}{p(1-p)}}.$$

where $N - \frac{1}{2\epsilon^2} \leq n \leq N$. This means that if $\mathcal{S}^\star$ doesn't restart, then it will pass above the translated box (as described in Figure 5.1) with probability bounded from zero.

The implications from this is that if you play $\mathcal{S}^\star$ from any state (n, H_n) inside the extended box as if that state was the origin state, the probability of winning a round will be at least $\left(1 - \mathrm{e}^{-\frac{9}{16}}\right)\frac{1}{3}\mathrm{e}^{-\frac{9.36}{p(1-p)}}\mathrm{P}_{\mathcal{S}^\star}(\neg R)$. Moreover, the expected duration from playing $\mathcal{S}^\star$ from (n, H_n) can be used to upper bound $\gamma(n, H_n)$,

$$\gamma(n, H_n) \leq \left(1 - \mathrm{e}^{-\frac{9}{16}}\right)^{-1} 3\, \mathrm{e}^{\frac{9.36}{p(1-p)}} \gamma(0, 0).$$

□

5.4 Bounds on indolent for Rivest

This section aims to prove the only remaining result, Theorem 5.11. The calculations used to show the bound given in the theorem are relatively heavy, but follows a rather simple idea.

The idea is to analyze how quickly the binomial distribution decays. From knowing this it is possible to show that the indolent strategy will, when it finishes a round, not only have exactly the necessary numbers of heads but also, with probability bounded away from zero, have an excess of heads (to be concretized in Lemma 5.13). The last step is to show that having this margin at the end of the game also implies that the strategy will, with probability bounded away from zero, have an excess amount of heads even earlier in the game, proving Theorem 5.11.

The first question is, for how long is the tail of the binomial distribution "constant"?

Lemma 5.12 (Property of the binomial distribution)**.** *Let $p \in (0, 1)$ and $\epsilon \in \left(0, \frac{1-p}{4}\right]$. Let n and k be integers such that $n(p+\epsilon)$ is integer, $n \geq \frac{4}{\epsilon^2}$ and $0 \leq k - n(p+\epsilon) \leq \frac{3}{\epsilon}$. Then*

$$1 \leq \frac{f(n(p+\epsilon), n, p)}{f(k, n, p)} \leq \mathrm{e}^{\frac{9.36}{p(1-p)}}$$

where $f(k, n, p)$ is the probability mass function of the binomial distribution.

So the probability mass function $f(k, n, p)$ for the binomial distribution can be bounded from below by a constant times $f(n(p+\epsilon), n, p)$ when $n(p+\epsilon) \leq k \leq n(p+\epsilon) + \frac{3}{\epsilon}$ and $n \geq \frac{4}{\epsilon^2}$.

Proof of Lemma 5.12. Note that for the binomial distribution

$$\frac{f(k, n, p)}{f(k+1, n, p)} = \frac{(k+1)(1-p)}{(n-k)p}.$$

Given the constraints in the theorem it follows that

$$n\epsilon \geq \frac{4}{\epsilon} \geq 4,$$

which in turn yields the following bounds

$$
\begin{aligned}
k &\le n(p+\epsilon) + \frac{3}{\epsilon} \\
&\le n\Big(p + \frac{7}{4}\epsilon\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
k+1 &\le n\Big(p + \frac{7}{4}\epsilon\Big) + 1 \\
&\le n(p+2\epsilon).
\end{aligned}
$$

So

$$
\begin{aligned}
\frac{f(k,n,p)}{f(k+1,n,p)} &= \frac{(k+1)(1-p)}{(n-k)p} \\
&\le \frac{n(p+2\epsilon)(1-p)}{n(1-p-\frac{7}{4}\epsilon)p} \\
&= \Big(1 + 2\frac{\epsilon}{p}\Big)\Big(1 + \frac{7}{4}\frac{\epsilon}{1-p-\frac{7}{4}\epsilon}\Big) \\
\{1+x \le \mathrm{e}^x\} &\le \exp\Big(2\frac{\epsilon}{p} + \frac{7}{4}\frac{\epsilon}{1-p-\frac{7}{4}\epsilon}\Big) \\
\Big\{\epsilon \le \frac{1-p}{4}\Big\} &\le \exp\Big(2\frac{\epsilon}{p} + \frac{28}{9}\frac{\epsilon}{1-p}\Big) \\
&= \exp\Big(\epsilon\frac{10p+18}{9p(1-p)}\Big) \\
\{p \le 1\} &\le \exp\Big(\epsilon\frac{28}{9}\frac{1}{p(1-p)}\Big) \\
&\le \exp\Big(\epsilon\frac{3.12}{p(1-p)}\Big).
\end{aligned}
$$

From this we get that for $k \ge n(p+\epsilon)$

$$
\begin{aligned}
\frac{f(n(p+\epsilon),n,p)}{f(k,n,p)} &= \prod_{i=n(p+\epsilon)}^{k-1} \frac{f(i,n,p)}{f(i+1,n,p)} \\
\Big\{\frac{f(i,n,p)}{f(i+1,n,p)} \le \mathrm{e}^{\epsilon\frac{3.12}{p(1-p)}}\Big\} &\le \Big(\mathrm{e}^{\epsilon\frac{3.12}{p(1-p)}}\Big)^{k-n(p+\epsilon)} \\
\Big\{k \le n(p+\epsilon) + \frac{3}{\epsilon}\Big\} &\le \Big(\mathrm{e}^{\epsilon\frac{3.12}{p(1-p)}}\Big)^{3/\epsilon} \\
&= \mathrm{e}^{\frac{9.36}{p(1-p)}}.
\end{aligned}
$$

This shows the inequality on the right hand side. To show ≥ 1 just note that the mean of the binomial distribution $np \le n(p+\epsilon) \le k$, and the Binomial is decreasing after the mean, so

$$
\frac{f(n(p+\epsilon),n,p)}{f(k,n,p)} \ge 1.
$$

$\square$

Lemma 5.12 is a tool that describes the decay rate of the tail of the binomial distribution. With this lemma, it is possible to show how much margin the indolent strategy can reasonably be expected to have when it wins, i.e. how many extra heads it has when it finishes a round.

Lemma 5.13 (Winning with a margin)**.** *Valid for the indolent strategy I playing Rivest*$(N, \epsilon;\, p)$*. If* $0 < \epsilon \leq \frac{1-p}{4}$ *and* $N \geq \frac{4}{\epsilon^2}$ *then*

$$\frac{\mathrm{P}_I\left(N(p+\epsilon) + 2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N\right)}{\mathrm{P}_I(N(p+\epsilon) \leq H_N)} \geq \frac{1}{3}\mathrm{e}^{-\frac{9.36}{p(1-p)}}.$$

Proof of Lemma 5.13. The lemma can be proven using the bound in Lemma 5.12. First rewrite the expression in order to be able to use the bound,

$$\begin{aligned}
&\frac{\mathrm{P}_I\left(N(p+\epsilon) + 2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N\right)}{\mathrm{P}_I(N(p+\epsilon) \leq H_N)} \\
&= \frac{\mathrm{P}_I\left(2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon)\right)}{\mathrm{P}_I\left(0\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon)\right)} \\
&= \frac{\mathrm{P}_I\left(2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right) + \mathrm{P}_I\left(3\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon)\right)}{\mathrm{P}_I\left(0\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right) + \mathrm{P}_I\left(3\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon)\right)} \\
\left\{\frac{a}{b} \leq 1 \Rightarrow \frac{a+c}{b+c} \geq \frac{a}{b}\right\} &\geq \frac{\mathrm{P}_I\left(2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right)}{\mathrm{P}_I\left(0\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right)}.
\end{aligned}$$

What remains is bounding the probability in the numerator from below, and the probability in the denominator from above. As these are integrals, they can be trivially bounded from below by the interval width times the minimum, and from above by the interval width times the maximum. Note that because the binomial distribution is decreasing after its mean Np we have that

$$\mathrm{P}_I\left(0\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right) \leq 3\left\lfloor\frac{1}{\epsilon}\right\rfloor f(N(p+\epsilon), N, p),$$

where f is the probability mass function of the binomial distribution, and in a similar manner it follows from Lemma 5.12 that

$$\mathrm{P}_I\left(2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right) \geq \left\lfloor\frac{1}{\epsilon}\right\rfloor \mathrm{e}^{-\frac{9.36}{p(1-p)}} f(N(p+\epsilon), N, p).$$

So

$$\begin{aligned}
&\frac{\mathrm{P}_I\left(2\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right)}{\mathrm{P}_I\left(0\left\lfloor\frac{1}{\epsilon}\right\rfloor \leq H_N - N(p+\epsilon) < 3\left\lfloor\frac{1}{\epsilon}\right\rfloor\right)} \\
&\geq \frac{\left\lfloor\frac{1}{\epsilon}\right\rfloor \mathrm{e}^{-\frac{9.36}{p(1-p)}} f(N(p+\epsilon), N, p).}{3\left\lfloor\frac{1}{\epsilon}\right\rfloor f(N(p+\epsilon), N, p),} \\
&= \frac{1}{3}\mathrm{e}^{-\frac{9.36}{p(1-p)}}.
\end{aligned}$$

$\square$

With Lemma 5.13 proven there is only one single proof left in this entire chapter, the proof of Theorem 5.11.

Proof of Theorem 5.11. This theorem essentially says that if you finish a round, then with probability bounded away from zero you will have had more than enough heads earlier in the game. Lemma 5.13 already states that if you finish a round then with probability bounded away from zero you will win with an excess of heads, so Theorem 5.11 essentially generalizes Lemma 5.13.

First note that if $N = n$ then the inequality reduces to a weaker version of Lemma 5.13, so the inequality in that case simply follows from Lemma 5.13. So we can assume that $n \neq N$ going forward.

The main idea of the proof of Theorem 5.11 is to split the probability on th LHS of the inequality

$$\mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; \neg R\right)$$

into two factors in such a way that one factor can be bounded using Lemma 5.13, and the other can be bounded using Hoeffding's inequality (A.3). Starting out with the probability we get

$$\begin{aligned}
&\mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; \neg R\right) \\
&= \mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon)\right) \\
&\geq \mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \wedge H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon)\right) \\
&= \mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right) \times \\
&\quad \mathrm{P}_I\left(H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon)\right) \\
&= \mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right) \times \\
&\quad \frac{\mathrm{P}_I\left(H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right)}{\mathrm{P}_I(H_N \geq N(p+\epsilon))}.
\end{aligned}$$

Lemma 5.13 implies that the ratio can be bounded from below by $\frac{1}{3}\mathrm{e}^{-\frac{9.36}{p(1-p)}}$. The only remaining probability left to bound is

$$\begin{aligned}
&\mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N \geq N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right) \\
&\geq \mathrm{P}_I\left(H_n \geq n(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N = N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right) \\
&= \mathrm{P}_I\left(H_N - H_n \leq (N-n)(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N = N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right) \\
&= 1 - \mathrm{P}_I\left(H_N - H_n > (N-n)(p+\epsilon) + \left\lfloor \frac{1}{\epsilon} \right\rfloor \;\middle|\; H_N = N(p+\epsilon) + 2\left\lfloor \frac{1}{\epsilon} \right\rfloor\right).
\end{aligned}$$

To simplify notation, let

$$Q := \mathrm{P}_I\Big(H_N - H_n > (N-n)(p+\epsilon) + \Big\lfloor \frac{1}{\epsilon} \Big\rfloor \,\Big|\, H_N = N(p+\epsilon) + 2\Big\lfloor \frac{1}{\epsilon} \Big\rfloor\Big).$$

Note that $(H_N - H_n | H_N) \sim \mathrm{HyperGeom}(N(p+\epsilon) + 2\lfloor \frac{1}{\epsilon} \rfloor, N, N-n)$. The goal is to use Hoeffding's inequality (A.3) to bound Q from above. In order to apply the Hoeffding bound to Q we will first rewrite Q in such a way that it is natural to apply the inequality, namely

$$Q = \mathrm{P}(X > (q+t)(N-n))$$

where $X \sim \mathrm{HyperGeom}(N(p+\epsilon) + 2\lfloor \frac{1}{\epsilon} \rfloor, N, N-n)$, $q = (p+\epsilon) + \frac{2}{N}\lfloor \frac{1}{\epsilon} \rfloor$ and $t = \lfloor \frac{1}{\epsilon} \rfloor \big(\frac{1}{N-n} - \frac{2}{N}\big)$. Since

$$\begin{aligned} t &= \Big\lfloor \frac{1}{\epsilon} \Big\rfloor \Big(\frac{1}{N-n} - \frac{2}{N}\Big) \\ \Big\{\frac{1}{N-n} \geq 2\epsilon^2, \frac{2}{N} \leq \frac{\epsilon^2}{2}\Big\} &\geq \Big\lfloor \frac{1}{\epsilon} \Big\rfloor \frac{3}{4} \frac{1}{N-n}, \end{aligned}$$

which is positive, we can apply the Hoeffding bound (A.3) to get

$$\begin{aligned} Q &\leq \mathrm{e}^{-2t^2(N-n)} \\ \Big\{t \geq \Big\lfloor \frac{1}{\epsilon} \Big\rfloor \frac{3}{4} \frac{1}{N-n}\Big\} &\leq \exp\Big(-2\Big\lfloor \frac{1}{\epsilon} \Big\rfloor^2 \frac{9}{16}\Big(\frac{1}{N-n}\Big)^2 (N-n)\Big) \\ &= \exp\Big(-\frac{9}{8}\Big\lfloor \frac{1}{\epsilon} \Big\rfloor^2 \frac{1}{N-n}\Big) \\ \Big\{\frac{1}{N-n} \geq 2\epsilon^2\Big\} &\leq \exp\Big(-\frac{9}{4}\Big\lfloor \frac{1}{\epsilon} \Big\rfloor^2 \epsilon^2\Big) \\ \Big\{\Big\lfloor \frac{1}{\epsilon} \Big\rfloor^2 \epsilon^2 \geq \frac{1}{4}\Big\} &\leq \mathrm{e}^{-\frac{9}{16}}. \end{aligned}$$

The conclusion is that

$$\mathrm{P}_I\Big(H_n \geq n(p+\epsilon) + \Big\lfloor \frac{1}{\epsilon} \Big\rfloor \,\Big|\, \neg R\Big) \geq \big(1 - \mathrm{e}^{-\frac{9}{16}}\big)\frac{1}{3}\mathrm{e}^{-\frac{9.36}{p(1-p)}}.$$

□

6

Numerical analysis

The quitting version of the Rivest coin tossing game is useful not only for theoretical analysis but also for numerical computation. Here we will present methods for computing an optimal strategy for any given instance of the game, followed by some results of such computations in Section 6.3.

We will consider an instance of Rivest$(N, K; p)$ defined in Section 3.1. It it natural to relabel the state space for computational uses. Let (n, k) be the state with n coins left to toss and k heads left until reaching K heads.

6.1 Computing an optimal strategy

One step in computing an optimal restart strategy is computing an optimal strategy for the quitting version of the game.

For brevity let $T_q(n, k)$ be $\mathrm{E}^*(C_q)$ for the Rivest quitting game with initial state (n, k) with quitting cost q. When analyzing a quitting game on a graph it is often natural to have the expected cost T_q defined in terms of T_q in other reachable states. For the quitting version of the Rivest coin game the relation for T_q is particularly nice because the graph describing the game is a directed acyclic graph (DAG) and very regular in structure, see Figure 3.1.

Defining a recurrence for T_q is straightforward since the graph is a DAG. At any given state two choices exist: quit with some quitting cost q, or toss a coin and end up at one of two new states where we assume T_q is known. The possible outcomes can be represented graphically in Figure 6.1, and summarized as a formula we have

$$T_q(n, k) = \min\{q,\ 1 + p\,T_q(n-1, k-1) + (1-p)\,T_q(n-1, k)\}.$$

with initial values

$$\begin{aligned} T_q(0, k) &= 0 \quad \text{if } k \leq 0 \\ T_q(0, k) &= q \quad \text{if } k > 0 \end{aligned}$$

corresponding to winning and losing states with no coins left to toss.

6.1.1 Dynamic programming

Naive computation of $T_q(N, K)$ leads to an exponential number of function evaluations, $O(2^N)$. This can be improved by use of dynamic programming, which is a very broad technique applicable to problems that compose into overlapping subproblems.

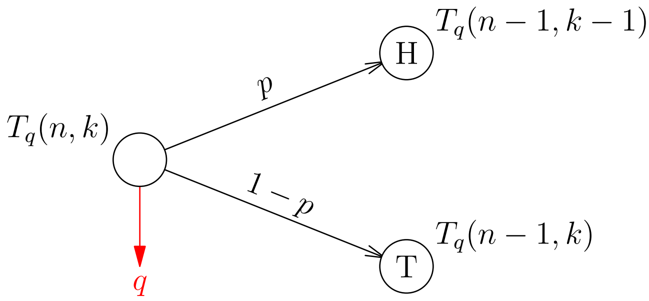


Figure 6.1: Diagram of the two possible choices at state (n, k) which determines $T_q(n, k)$. Either quit at cost q or continue playing and get one plus the weighted sum of T_q in the two reachable states.

In essence it involves identifying common subproblems, computing them only once and storing the results (memoization). In this case evaluating function calls in increasing order of n (working backwards from the last coin toss) works and brings the number of evaluation down to $\mathcal{O}(NK)$.

As a consequence of Theorem 2.9, if the quitting cost equals the grade of a state in Rivest then the table computed during the dynamic programming process gives an optimal DM restart strategy for playing from that state.

6.1.2 Finding the correct quitting cost

In order to apply Corollary 2.12 one needs to know $\mathrm{E}^*(T) = \gamma(N, K)$. One alternative definition of grade given in Corollary 2.10 says that $\gamma(N, K)$ is the largest q such that quitting at (N, K) is optimal, i.e. largest q such that $T_q(N, K) = q$. This means that $\mathrm{E}^*(T)$ can be found with binary search. To be able to binary search a lower and an upper bound on $\mathrm{E}^*(T)$ is needed. Obtaining a lower bound is no issue (trivially $\mathrm{E}^*(T) \geq N$). Obtaining an upper bounds takes some more finesse (and will yield a better lower bound in the process).

Finding bounds through repeated squaring and adjustments A rough upper and lower bound can be computed through repeated squaring. See Algorithm 1.

Algorithm 1 Find bound through repeated squaring

Postcondition: $l \leq \gamma < r$

```
l ← 1
r ← 2
while T_r(N, K) = r do
    l ← r
    r ← r^2
return [l, r)
```

The number of evaluations of $T_q(N, K)$ needed is the smallest x such that $2^{2^x} > \gamma$, hence $x = \lceil \log_2 \log_2 \gamma \rceil$. So it finds an upper bound quickly. The bound on γ can be improved by adjusting with earlier terms in the series of squares. Which is equivalent to a binary search in the exponent of 2^n. See Algorithm 2.

The number of evaluations for this improvement step is the same as for the bounds finding.

Algorithm 2 Binary search in exponent

Precondition: l, r such that $l \leq \gamma < r$ and $l^2 = r = 2^{2^x}$ for some integer x
Postcondition: $r = 2l$

for f in $\{\sqrt{l}, \dots, 2^8, 2^4, 2^2, 2^1\}$ **do**
  if $T_{lf}(N, K) = lf$ **then**
    $l \leftarrow lf$
  else
    $r \leftarrow lf$
return $[l, r)$

Binary search Now all that remains is to do a binary search between the found l and r to approach γ. Either continue the binary search until some wanted precision (relative error) is reached, or continue searching until the strategy derived from the dynamic programming table is the same for the upper and the lower bound.

When binary searching to a given precision (relative error) the number of iterations needed is $\mathcal{O}(\#\text{significant digits})$.

It is worth noting that we while it is technically possible to continue binary searching the exponent and never falling back on the standard binary search, this would make q an irrational number, and since the code makes use of exact fractions this is not feasible.

Total running time If arithmetic operations could have been assumed to take $\mathcal{O}(1)$ time then the complexity would be

$$\mathcal{O}(NK(\log\log\gamma + \#\text{significant digits}))$$

but this is not the case since we cannot use floating point numbers for such large numbers, and have to resort to big integer arithmetic (exact fractions). The cost could be bounded from above by looking at what the worst cost of an arithmetic operation on fractions could be for a given N and p and multiplying the complexity by that.

6.2 Implementation details

The simulation code was written in C++ making use of the GNU Multiple Precision Arithmetic Library (GMP). This was done to keep answers exact, since the numbers involved far exceed the range and precision of floating point numbers. The actual implementation is linked in Appendix B.2.

··· indolent
··· optimistic
— promising (step $2/\epsilon^2$)
— promising (step $0.5/\epsilon^2$)
-- probabilistic
-- promising (step 1)
— optimal

Figure 6.2: Common legend for all the plots.

6.3 Numerical results

When doing runs for data collections a set of parameters of interest was recorded for the strategies described in Chapter 4. Specifically, the strategies evaluated were

optimal
the computed optimal strategy

promising ($\boldsymbol{\delta = 2/\epsilon^2}$)
promising strategy with distance $2/\epsilon^2$ between checkpoints

promising ($\boldsymbol{\delta = 0.5/\epsilon^2}$)
promising strategy with distance $0.5/\epsilon^2$ between checkpoints

probabilistic ($\boldsymbol{\lambda = 1/2}$)
restarts if probability of winning at a given state is $< \mathrm{P}(G_W)/2$

promising ($\boldsymbol{\delta = 1}$)
a degenerate version of the promising strategy with distance 1 between checkpoints

indolent
never restarts except when forced to because of failing a round

optimist
only restarts if there is 0 probability of winning

where for all promising strategies $\beta = 1/2$. The parameters of particular interest were

$\mathbf{E}(\boldsymbol{t \mid R})$
Expected time of a round, given that there is a restart

$\mathbf{E}(\boldsymbol{T})$
Expected duration of the game

$\mathbf{P}(\boldsymbol{\neg R \mid G_W})$
Probability of finishing a win-conditioned base game

$\frac{\mathbf{E}(t \mid R)}{\mathbf{P}(\neg R \mid G_W)}$ Closely related to the important fraction discussed in Section 2.2.2

Technical details about how the data generation process was performed can be found in Appendix B.1. In the following sections a subset of particularly interesting plots are shown, remaining plots are linked in Appendix B.2. The common legend for all the plots can be found in Figure 6.2.

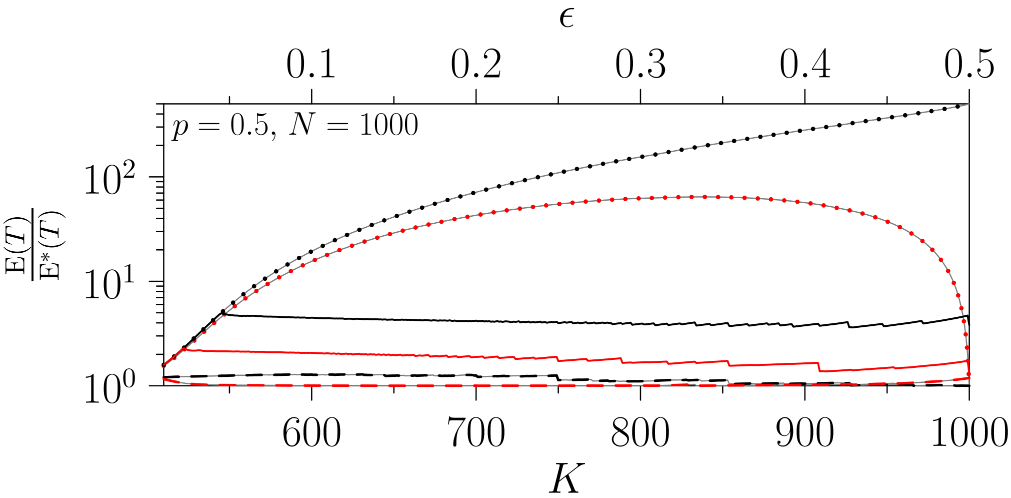


Figure 6.3: Expected duration of the game relative to optimal. Note that the promising strategies with $\mathcal{O}(1/\epsilon^2)$-spaced checkpoints flatten out to be a constant factor from optimal as soon as the checkpoints start occurring, which confirms the behavior proven in Chapter 5. Also note that the two dashed lines (promising with step 1 in red and probabilistic in black) which aren't constrained by the spaced out checkpoints have considerably better constant factors.

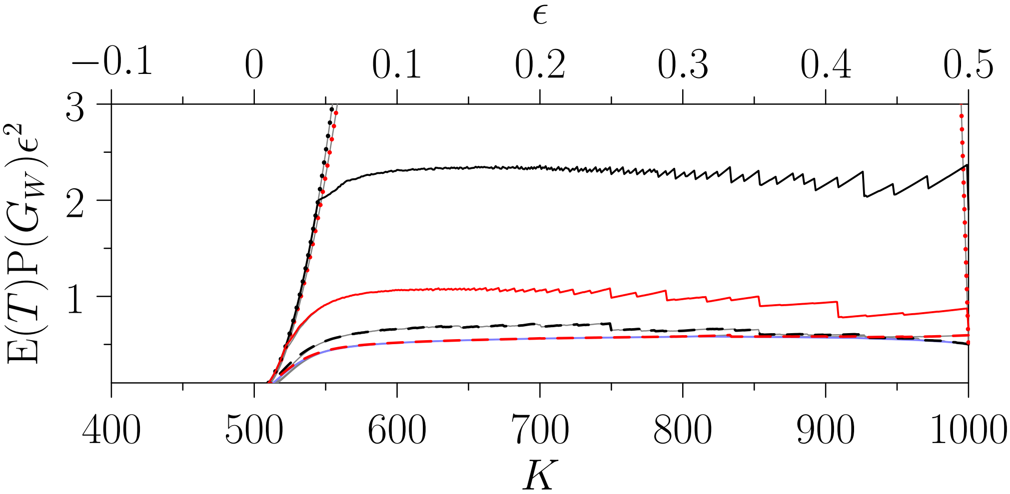


Figure 6.4: Normalized duration plot, cropped to contain the non-trivial strategies. Note that all non-trivial strategies are almost perfectly constant in this plot, despite $\mathrm{E}(T)$ being incredibly large. Also note that both solid lines (promising strategies) are rather close to optimal (within a factor of roughly 2 and 4). It should also be noted that the degenerate promising with step 1 is extremely close to optimal except for the most extreme K where probabilistic instead closely follows the optimal strategy.

6.3.1 Expected total time

The expected duration of a game exhibits the behaviors expected from Chapter 5, which can be seen in Figure 6.3. Since $\mathrm{E}(T)$ grows very large it is not feasible to plot it for all strategies in an absolute manner (not relative to optimal). This can however be handled by plotting $\mathrm{E}(T)\mathrm{P}(G_W)\epsilon^2$ like in Figure 6.4, which normalizes the growth of $\mathrm{E}(T)$ and removes the $1/\epsilon^2$ scaling.

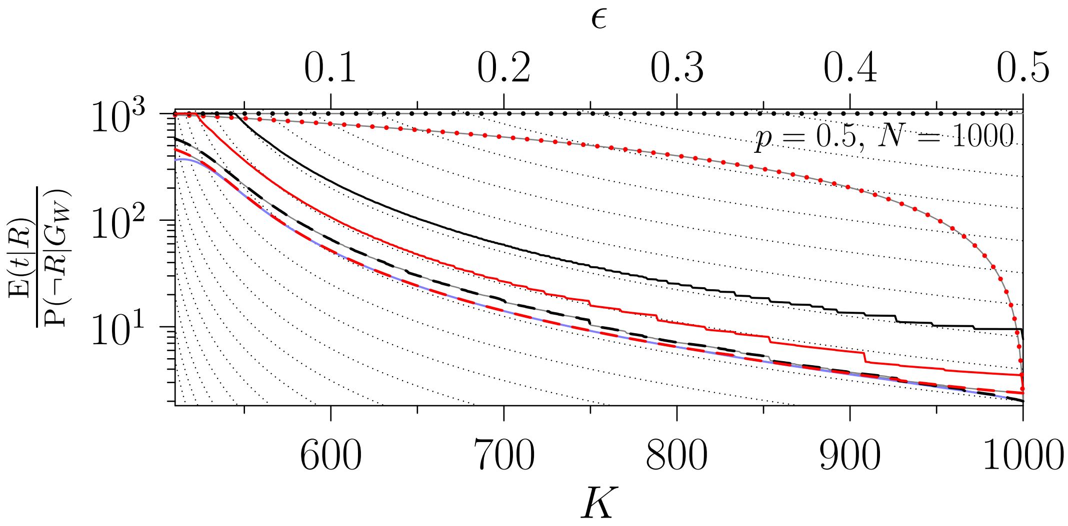


Figure 6.5: Plot showing the important ratio as a function of K and ϵ. The dotted lines are curves of the form $2^a/\epsilon^2$, $a \in \mathbb{Z}$. Note that all of the non-trivial strategies studied follow the curves closely for not too small ϵ.

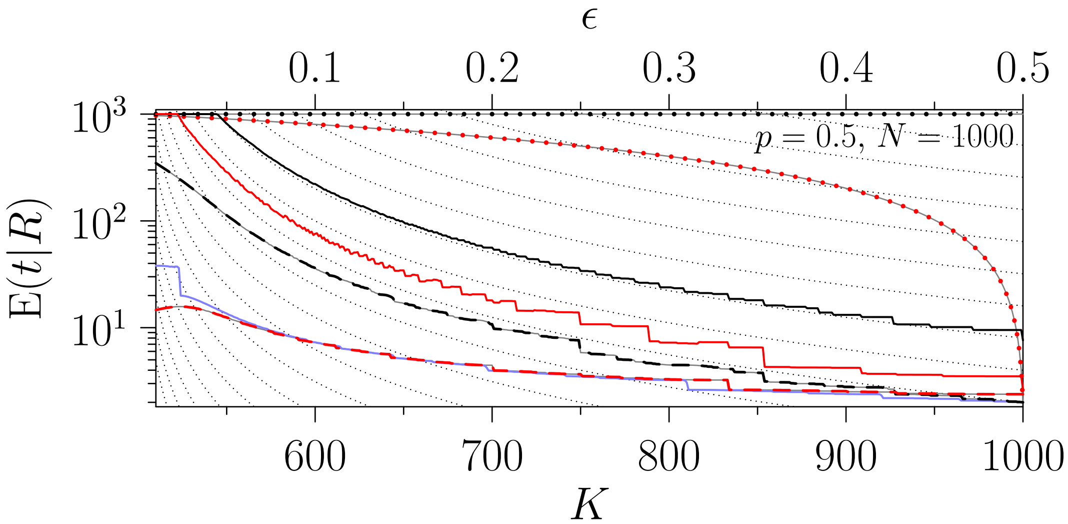


Figure 6.6: Plot showing the expected duration of a restarting round as a function of K and ϵ. The dotted lines are curves of the form $2^a/\epsilon^2$, $a \in \mathbb{Z}$. Note that both promising strategies with checkpoints spaced out on the order of $1/\epsilon^2$ (solid black and red) follows the dotted curves closely as expected. Probabilistic (dashed black) also seems to follow this scaling roughly. Promising with step 1 and optimal does not follow these lines, meaning $E(t \mid R)$ is not proportional to $1/\epsilon^2$. This in turn implies that their $P(\neg R \mid G_W)$ must have an ϵ dependence to maintain the $1/\epsilon^2$ scaling seen in Figure 6.5. This also mean they are much more aggressive.

6.3.2 Strategy properties

From the theoretical analysis it is clear that $1/\epsilon^2$ scaling is a key feature of an optimal (or close to optimal) strategy. It is interesting to investigate if this is apparent in practice, and how it shows up. For this there are two main results to consider: the ratio from Section 2.2.2 in Figure 6.5 and the expected round duration of a restarting round in Figure 6.6. Together these two plots highlights some differences between

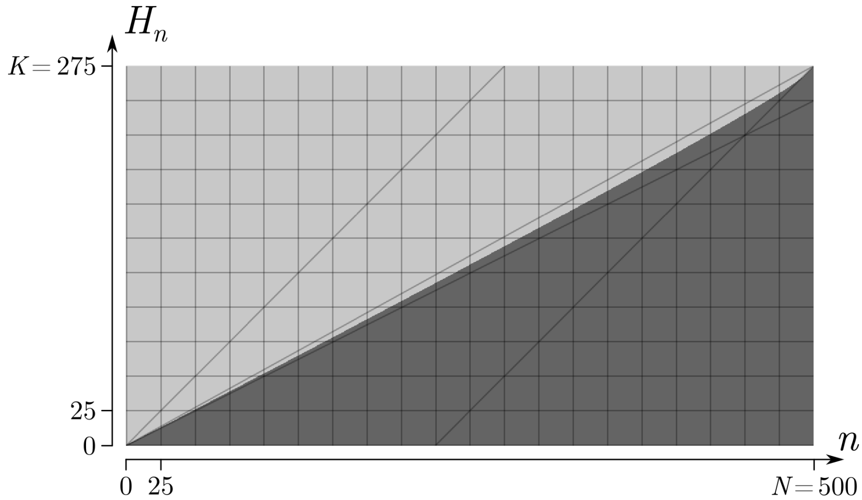


Figure 6.7: An optimal strategy for the case $N = 500$, $K = 275$ and $p = 0.5$, with the initial state in the bottom left. The light gray area corresponds to play states and the dark gray area corresponds to restart states. The grid mark multiples of 25, the 45 degree lines marks the boundary of unreachable and guaranteed loss states respectively. The two other diagonal lines mark the expected number of heads (related to p) and the number of heads needed to be on pace to finish the game (related to $p + \epsilon$). Note that the boundary between play states and restart states lies between the p and $p + \epsilon$ lines, and that could be rather well approximated with a line until the very end, which matches the behavior noted in Figure 6.4.

the strategies.

In Figure 6.5 *all* of the non-trivial strategies seem to excibit $\frac{1}{\epsilon^2}$ scaling. This is to be expected since Chapter 5 shows that the promising strategy is constant-factor from optimal and hence scales like optimal strategies.

However, the result seen in Figure 6.6 which covers only the numerator of the ratio shows major differences. In the case of the promising strategy $\mathrm{P}(\neg R \mid G_W)$ is essentially kept fixed, so $\mathrm{E}(t \mid R)$ has $\frac{1}{\epsilon^2}$ scaling. This is also roughly the case of probabilistic. The other non-trivial algorithms, including an optimal strategy, does *not* exhibit this scaling, but are instead a lot more aggressive in their restarts (which forces $\mathrm{P}(\neg R \mid G_W)$ to be lower as well).

6.3.3 An optimal strategy

From the dynamic programming table the optimal strategy can be derived. One example of which states it plays in is shown in figure Figure 6.7.

7

Discussion

We've proven that the two strategies, the promising strategy $\mathcal{P}$ and the probabilistic strategy $\mathcal{Pr}$ are constant-factor from optimal in the two games GCBC$(N, \epsilon; p)$ and Rivest$(N, \epsilon; p)$, with the constant only allowed to depend on the probability of heads p, see Theorems 4.1 and 4.4. Both of these strategies can be generalized to other restart games, so their properties are of particular interest as they might be shared by other constant-factor from optimal strategies in restart games similar to Rivest.

Notable properties of restart strategies The main property that both strategies share is that when playing GCBC and Rivest the probability of not restarting on a win-conditioned round $\mathrm{P}(\neg R \mid G_W)$ is bounded away from zero, see Theorems 4.3 and 4.5. This means that in other restart games there might exist constant-factor from optimal strategies that also keeps $\mathrm{P}(\neg R \mid G_W)$ bounded away from zero. In particular the probabilistic strategy is a good candidate for this since Theorem 4.5 states that $\mathrm{P}(\neg R \mid G_W)$ is bounded away from zero in any restart game.

There is also the question of how long you should play before considering to restart. For example if the first coin tossed was a tail, should you restart? From the numerical analysis, see for example Figure 6.5, it is clear that the optimal strategy does early restarts. However this is not the case for the promising strategy, which by definition will always wait α/ϵ^2 tosses before even considering restarting, so early restarts is not necessary in order to be constant-factor from optimal.

Scaling Another use of knowing what strategies are constant from optimal is that it is possible to express the minimum expected duration of the game up to a constant factor. This can be done by combining Theorems 2.4 and 4.1 to 4.3. The conclusion is that for a non-trivial instance of Rivest$(N, \epsilon; p)$ and GCBC$(N, \epsilon; p)$, i.e. when $\mathrm{P}(G_W) \ll 1$ and $N > \frac{1}{\epsilon^2}$,

$$\mathrm{E}^*(T) \propto \frac{1}{\mathrm{P}(G_W)\epsilon^2}.$$

From the numerical results in Figure 6.4 it is apparent that this is accurate. The numerical result also shows that the constants involved are much smaller than what was proven theoretically. Interestingly, the way of accomplishing this scaling is very different between the strategies. Promising has the optimal scaling, but is not nearly as aggressive as the strategies that outperform it. The probabilistic strategy plays more aggressive than promising, but still not as aggressive as the optimal strategy.

Constant factor for the promising strategy For the Rivest coin game the derived upper bound on the constant factor for the promising strategy is very loose, at best $> 10^{17}$, and much worse for extreme values of p (tends to ∞ as $p \to 0$ or

$p \to 1$). This is in part because we have valued simplicity in the proof over trying to optimize the bound, and in part because the problem itself and the methods used involve exponentials which can easily grow out of hand. Note that this does not necessarily mean that the strategy itself is bad. In the cases studied numerically the actual factor for promising is about 2–4, see Figures 6.3 and 6.4, which is much smaller than the proven bound.

Possible future work One idea of a future subject for investigation could be to try to improve these theoretical bounds as we see that they are much smaller in the numerical analysis that what was shown in the theoretical analysis. Shorter and/or simpler arguments for the bounds and optimality would also be nice.

One significant reason for the very loose bounds is probably that the dependence on p has been neglected, so this is a prime subject for further investigation. We intentionally fixed the parameter p in this thesis to make things simpler and to able to use Hoeffding's inequality, see Appendix A.1, which is known to be extremely loose for p close to 0 or 1 [7]. So the analysis for varying p would require the use of another bound.

Another reason for considering a varying p is to allow for something like fixing Np and letting N tend to infinity. This would make the binomial distribution in Rivest tend to a Poisson distribution. So for extreme values of p Rivest will essentially be a completely different game. In stark contrast, the limit studied in this thesis reduces to Brownian motion and normal distributions.

A last candidate for future work is to consider the serendipitous mishap that is the degenerate promising strategy with $\delta = 1$. The strategy came about as a mistake when doing numerical simulations, but turned out to perform very well (close to optimal). Looking at Figure 6.4; which include the promising strategy with $\delta = 2/\epsilon^2$, $\delta = 0.5/\epsilon^2$ and $\delta = 1$; it seems that the strategy performs increasingly well for smaller δ, getting very close the optimal strategy. However, in Figure 6.6 it is apparent that this strategy does not share the properties of the typical promising strategy and utilizes much more aggressive restarts and as a consequence sacrifices $\mathrm{P}(\neg R \mid G_W)$. It could even be the case that $\mathrm{P}(\neg R \mid G_W)$ could become arbitrarily close to zero for some choice of game parameters.

Generalizability of our results Recall that the Rivest coin game was introduced as a simplified model of other restart games. Rivest introduced the game to model one step of Dixon's factorization algorithm. The Rivest coin game is of course a good model for this, however the p involved might need to be small. This means that our proofs for optimality does not tell much of this situation, as the bounds gets very loose.

Optimization of clinical trials A discipline where optimal stopping is common is in optimization of clinical trials [8][9]. A game similar to GCBC could be used to model this kind of scenario. A good coin corresponds to an effective drug, and a bad coin corresponds to a placebo. Tossing a coin corresponds to trying the drug on a subject. Head means a positive response and tail means a negative response. Restarting means abandoning the current medicine and developing a new one (potentially with some cost). The goal is to minimize the number of test subjects needed (for cost or humanitarian reasons) to find a good drug. For the sake of simplicity, let's assume that it is mandated that a drug go through N tests before being able to be considered a good drug.

Without a restart cost, this game is exactly GCBC and we have shown constant-factor from optimal strategies for this game. Adding a restart cost will make not discarding good medicines more important. Our strategies already keep a large fraction of all good medicines, so they will only get better compared to optimal. So for this kind of simple model of clinical trials our models, our results should be applicable. But in practice, these kinds of models would be solved numerically similar to the work we have done.

Speedrunning A motivation for us considering the coin game initially was as a toy model of speedrunning, completing some game as quickly as possible. Play the Rivest coin game. Heads are interpreted as gaining some unit of time compared to some baseline. Tails are interpreted as losing some unit of time compared to some baseline. The goal is to at the end of the game beat the baseline by some amount, i.e. setting a record.

There are some major issues with this model. Speedruns are in general not time homogeneous, but Rivest is. Speedruns typically contains both numerous small time gains/losses and some rare but very large gains/losses. Modeling this as Rivest can only capture the small frequent gains/losses. Possibly, if Np would be fixed in Rivest then it might be possible to model the rare large gains/losses but not the frequent small ones.

8
Conclusions

By introducing restart games and a framework for analyzing them, we have managed to conceive two strategies, the *promising strategy* and the *probabilistic strategy*, that we (for a fixed probability of heads p) proved to be constant-factor from optimal for the Rivest coin game and the related good coin bad coin game. This shows that the approach of analyzing based on win-conditioned and loss-conditioned rounds, which has been the basis for the framework and the main motivation for the promising strategy, has been very successful.

The most interesting and useful parts of the analysis framework is

- Theorem 2.4 which decomposes the expected duration of a restart game conditioned on win/loss.
- Theorem 2.6 which provides necessary and sufficient conditions for a strategy to be optimal.
- Theorem 2.7 which shows the effects of forcing a strategy to play in some states.
- Theorem 2.11 which shows the close connection between restart games and quitting games (optimal stopping problems).

Theorems 2.4 and 2.7 forms the basis for the analysis of optimality in Chapter 5. Theorem 2.11 is the key to making the numerical analysis possible.

Overall we think the thesis managed to yield good results in that is has proven scaling properties of the optimal strategy of the Rivest coin game which (as far as we know) has not been shown earlier. The numerical results also gives interesting insights, and suggests some other strategies that could be investigated for optimality in future works. Another good candidate for future works is to try to analyze the p dependence that has been neglected in this thesis.

Bibliography

[1] I. Dumitriu, P. Tetali, and P. Winkler, "On playing golf with two balls", *SIAM J. Discrete Math.*, vol. 16, pp. 604–615, Jul. 2003. DOI: 10.1137/S0895480102408341.

[2] M. Luby, A. Sinclair, and D. Zuckerman, "Optimal speedup of las vegas algorithms", in *[1993] The 2nd Israel Symposium on Theory and Computing Systems*, IEEE. DOI: 10.1109/istcs.1993.253477.

[3] S. Janson and Y. Peres, "Hitting times for random walks with restarts", *SIAM Journal on Discrete Mathematics*, vol. 26, no. 2, pp. 537–547, Jan. 2012. DOI: 10.1137/100796352.

[4] G. F. Bachelis and F. J. Massey, "A coin tossing problem of R. L. Rivest", in *Functional-Analytic Methods for Partial Differential Equations*, H. Fujita, T. Ikebe, and S. T. Kuroda, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 36–52, ISBN: 978-3-540-46818-9.

[5] I. Hu and S. S. Venkatesh, "On the minimum expected duration of a coin tossing game", *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 581–593, Mar. 1993, ISSN: 0018-9448. DOI: 10.1109/18.212287.

[6] J. Steif, *Lecture notes: Percolations – Lecture 2*, http://www.math.chalmers.se/~steif/perclecture2.pdf, 2005.

[7] J. Domke, "Statistical Machine Learning: Learning Theory", Lecture notes https://people.cs.umass.edu/~domke/courses/sml2010/10theory.pdf, 2010.

[8] N. Stallard, "Sample size determination for phase ii clinical trials based on bayesian decision theory", *Biometrics*, vol. 54, no. 1, pp. 279–294, 1998, ISSN: 0006341X, 15410420.

[9] S. Jobjörnsson, "Hur designas en medicisk studie på bästa sätt?", Popular science talk at Vetenskapsfestivalen, Apr. 2018.

[10] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963. DOI: 10.1080/01621459.1963.10500830.

A

Well-known bounds

A.1 Hoeffding's inequality

The following bounds are special cases of Hoeffding's inequality [10].

Theorem A.1 (Tail bounds on binomial distribution). *Let* $X \sim \text{Binomial}(n, p)$*, then for* $t > 0$

$$\mathrm{P}(X \geq (p+t)n) \leq \mathrm{e}^{-2t^2 n} \tag{A.1}$$

$$\mathrm{P}(X \leq (p-t)n) \leq \mathrm{e}^{-2t^2 n} \tag{A.2}$$

Theorem A.2 (Tail bounds on hypergeometric distribution). *Let* $X \sim \text{HyperGeom}(K, N, n)$ *and let* $q := \frac{K}{N}$*, then for* $t > 0$

$$\mathrm{P}(X \geq (q+t)n) \leq \mathrm{e}^{-2t^2 n} \tag{A.3}$$

$$\mathrm{P}(X \leq (q-t)n) \leq \mathrm{e}^{-2t^2 n} \tag{A.4}$$

A.2 Kolmogorov's inequality

Theorem A.3 (Tail bound on random walk). *Let* $X_1, \ldots, X_n$ *be independent random variables with expected value* $\mathrm{E}(X_k) = 0$ *and variance* $Var(X_k) < \infty$ *for* $k = 1, \ldots, n$*. Then, for each* $\lambda > 0$*,*

$$\mathrm{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2} Var(S_n),$$

where $S_k = X_1 + \ldots + X_k$*.*

Corollary A.4 (Corollary of Theorem A.3). *In particular if* $X_k = Y_k - \mathrm{E}(Y_k)$ *where* $Y_k \sim Bernoulli(p)$*, then*

$$\mathrm{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2} \sqrt{np(1-p)}.$$

From this it is also possible to construct a single sided bound

$$\mathrm{P}\left(\min_{1 \leq k \leq n} S_k \leq -\lambda\right) \leq \frac{1}{\lambda^2} np(1-p). \tag{A.5}$$

B
Numerical extras

B.1 Data generation

In the case of fixed N and p (the line plots) K was simply swept over all possible k, $k = 1, \dots, N$. The parameter space swept in N and p was

$$(N, p) \in \{100, 500, 1000\} \times \{0.4, 0.45, 0.5, 0.55, 0.6\}.$$

In the case of fixed p only (here $p = 1/2$), n was swept exponentially spaced (linear in the exponent) between l and r and ϵ was swept linearly in $[-\frac{p}{4}, 1-p]$. The number of sampled points in n and k was chosen deliberately on the form $2^m + 1$, which causes the sampled points for m to be a subset of the sampled points for $m + 1$. So m can be swept upwards to generate an increasingly fine grained result, filling in details.

B.2 Extra resources

A more comprehensive set of plots, as well as the code used for simulation, can be found at bitbucket.org/algmyr/optimal-restart-extra-content.