Master thesis

ANALYSIS OF A SIMPLE METHOD TO APPROXIMATE THE EXPECTED STATE OF A MARKOV CHAIN

By
Johannes Schneider

Supervisor: Prof. Osamu Watanabe

Tokyo Institute of Technology

May 2004

Table of Contents

List of Figures

Abstract

Watanabe, Sawai and Takahashi [9] proposed an approximation method of the expected state for homogeneous Markov chains, named as *pseudo expectation*. It is based on simple probabilistic recurrence formulas. The aim of this work is to give a bound on the approximation error of the *pseudo expectation*.

Two bounding techniques are explained. Both are given as recurrence formulas.

The first one uses statistical properties of the process. It requires at least a bound on the 2nd moment, e.g. the variance.

The second one, the so called *linearisation error*, which captures the non-linearity of the process, is expressed in terms of the function f used to iteratively calculate the pseudo expectation.

The use of these techniques is demonstrated by an example, giving an explicit error formula (in closed form) for the considered process.

Acknowledgements

I am deeply indebted to my supervisor at the Tokyo Institute of Technology, Professor Osamu Watanabe, for his tremendous support for preparing this thesis and for his direction towards my research. I am also very grateful to my supervisor at the Swiss Federal Institute of Technology Zurich (ETHZ), Professor Emo Welzl, for giving me the great chance to go to Japan.

Apart from that, Mr. Niikura was always a good partner on working on the given research problem. Moreover, I would like to say thank you to all of the members of Prof. Watanabe's laboratory for cheering me up on the tough days of my research. Especially I am obliged to thank my tutor and friend Mr. Yamamoto.

Of course, I am grateful to my parents for their patience and *love*. Without them this work would never have come into existence (literally).

Tokyo, Johannes Schneider
May 6th, 2004

Chapter 1

Introduction

Markov chains or Markov processes have numerous applications in almost all fields of engineering. Very often, one is interested in the average behavior of a Markov process over time. The study of Markov processes is an established area, and for analyzing them, various techniques have been developed. But if its state space is huge, conventional techniques for analyzing the average behavior of the Markov process are in most cases not useful due to their computational complexity.

In some situations, the process follows a simple rule, though this does not necessarily mean that the state space is small. For analyzing some of such a kind of Markov processes, which often can be seen as a model of a randomized algorithm, Watanabe (see, e.g., [9]) has proposed to use a simple probabilistic recurrence formula — a *pseudo expectation*. Watanabe, Sawai, and Takahashi [9] showed, through computer experiments, that this pseudo expectation approximates well some Markov process for analyzing the behavior of some randomized algorithm. Later stochastic properties of this pseudo expectation have been investigated by Takahashi and Niikura [8]. Unfortunately, though, no theoretical result has been shown to bound the difference between the real and the pseudo expectation, although some investigations have been reported recently ([6], [7]).

The goal of this thesis is to estimate and bound the error of the pseudo expectation. For this goal, we propose two approaches.

(1) Stochastic Approach:

We consider the distribution of states at each step. We assume/ estimate a certain concentration around the pseudo expectation, thereby deriving an error bound.

(2) Mechanical Approach:

We estimate the approximation error by investigating the function used to define the pseudo expectation.

For each of these approaches, we propose methods for analyzing the approximation error. Using these methods, we indeed obtain "reasonable" bounds for a simple example.

1.1 A simple example

In order to illustrate our goal concretely, we introduce here a simple Markov process - *1-dimensional process.*

1. Player A and B have $A_n \geq 0$ and $B_n \geq 0$ balls at time n respectively.
2. If A (B) wins, A (B) gives $\min\{d, A_n\}$ ($\min\{d, B_n\}$) balls to B (A).
3. The probability of the event A wins depends on A_n. Let N be the total number of balls. Then

$$\mathbf{prob}(A_n) \quad = \quad \frac{w \cdot A_n}{w \cdot A_n + B_n}$$

This process will be used throughout this thesis as a basic example. More details are given in section 2.4.

The rule for this process is very simple. Nevertheless, it does not mean that the state space is small. In fact, the size of the state space, N in this example, could be very large, which makes the analysis hard. For example, the size of the transition matrix is N^2. Because of this it is not so easy to compute the real expectation $\mathbb{E}[X_n]$. (Note: Even if the transition matrix is large, since it is sparse, one may be able to use some clever algorithm to compute $\mathbb{E}[X_n]$. But it seems hard to reduce the computational complexity significantly more than $O(n \cdot N)$.)

1.2 Motivation: Analysis of randomized algorithms

Simple Markov processes (as the one shown above in 1.1) can be seen as a model of randomized algorithms on random instances. In fact, Watanabe ([9] and [7]) proposed to use Markov processes to analyze the performance/properties of randomized algorithms. In order to illustrate this approach and give a motivation for our investigation, we give one example:

Consider the following variation of 3-SAT, one of the famous NP-complete problems.

Definition 1.2.1. (Random 3-SAT (5 Occurrence) Problem) An instance is a Boolean formula F on n Boolean variables of the 3-conjunctive normal form. That is, F is of the form

$$F = C_1 \wedge C_2 \wedge \cdots \wedge C_m$$

where each C_i consists of 3 literals. (A literal is a variable or its negation.) Assume that every variable appears 5 times in the formula.

The obvious task is to find one satisfying assignment for F (if it exists). In a random scenario F is given uniformly at random from all possible 3-CNF formulas (see definition 1.2.1).

We then consider the following simple randomized local search algorithm:

```
LocalSearch3SAT(input F);
begin
    a ← (0,...,0);
    repeat the following MAXS times {
        if (F(a) = 1) output a and halt;
        for each variable x_i, compute its penalty(*) pen_i;
        n_0 ← number of penalty 0 variables;
        n_1 ← number of penalty 1 variables;
        ...
        n_5 ← number of penalty 5 variables;
        select x_i with

                prob. = W_i/(W_1 * n_1 + ... + W_5 * n_5);

        flip the assignment for x_i;
    }
    output "fail" and halt;
end
```

(*) pen_i = the number of unsatisfied clauses (under the current assignment) having the variable x_i.

It would be more standard to select a variable to flip from those with the highest penalty. But by choosing $W_0 = 0$, $W_1 < W_2 < \cdots < W_5$, the algorithm executes almost the same way as the standard one.

Now let us approximate the behavior of the algorithm by using a simple Markov process. A state is a tuple of six numbers $(n_0, n_1, ..., n_5)$, which expresses the number of variables of each penalty. Thus, we have $n_0 + n_1 + \cdots + n_5 = n$. As in the algorithm, we select, at each step, k, $0 \leq k \leq 5$, with probability $p(k) = W_k \cdot n_k/T$, where $T = W_0 n_0 + W_1 n_1 + \cdots + W_5 n_5$. Suppose that $k = 4$ is selected. Then we need to decrement n_4 by one, and increment n_1 by one, simulating the flip of some variable x_i with penalty 4. Some more updates are necessary because there are some variables whose penalty is changed by flipping x_i (and clauses having x_i get satisfied/unsatisfied). There are exactly $10 = 2 \times 5$ such variables (assuming no repetition). In the algorithm, those 10 variables are fixed from F. But for the Markov process approximation, we select these 10 variables *randomly* (more precisely, randomly proportional to $(n_0, ..., n_5)$). The one step transition of our Markov process is determined in this way. (Here we omit writing down all the details.)

Of course, this Markov process is just an approximation of the real execution, one may expect that it is close to the real one, in particular, if the time bound $MAXS$ is not so large and not so many variables are flipped more than once during the execution. Thus, we may hope that some of the characteristics of random executions of the algorithm on random instances can be investigated by using this Markov process.

For analyzing the behavior of the algorithm, we would like to see, for example, how n_0 increases on average. Our pseudo expectation can be useful for such an analysis. Note that the state space is $O(n^5)$, which is quite large even for small n. Therefore, the conventional Markov chain analysis is hard to use. On the other hand, we can estimate the expected value of n_0 at t-th step by

$$\mathbb{E}_{n_0}[n_0^{(t)}] \approx (f_{n_0})^{t-1}(\,(n_0^{(0)}, ..., n_5^{(0)})\,),$$

where $\overrightarrow{f}$ is a function for computing the average state change at each step, $f_{n_0}^{t-1}(\cdots)$ is the first component of the state obtained from the initial state $(n_0^{(0)}, ..., n_5^{(0)})$ by applying $\overrightarrow{f}$ for $t-1$ times. Furthermore, by studying $(f_{n_0})^t$ (e.g., computing its first derivative), we can investigate how n_0 grows on average. This is the advantage of using the pseudo average.

1.3 Summary of results

The goal of this thesis is to estimate and bound the error of pseudo expectations. For this goal, we obtained the following main results.

Two general bounding techniques are explained. Both are given as recurrence formulas.

The first one, given in chapter 4, uses statistical properties of the Markov chain, e.g. requires at least a bound on the 2nd moment of the stochastic process.

The second one, the so called *linearisation error*, which captures the non-linearity of the process, is expressed in terms of the function f, which is used to iteratively calculate the pseudo expectation.

Apart from that a technique for computing a lower and upper bound for the convergence speed of the pseudo expectation is given (see section 3.1). For the considered example process, a tight bound for the steady state error has been obtained (see section 3.2).

Chapter 2

Preliminaries

Some definitions needed for the technical discussion done in the following chapters are given.

First of all, some notations used in this thesis are explained in section 2.1. The definitions of Markov chains in 1 and 2 dimensions are stated in section 2.2.1 and 2.2.2.

Besides, the notation of the pseudo expectation is explained in more detail - again in 1 and 2 dimensions (section 2.3.1 and 2.3.2 respectively).

Finally, a concrete example of a one dimensional stochastic process is given in section 2.4 and illustrated by some numerical instances in 2.4.1.

2.1 Notation

To allow a fast and smooth way of reading, as few non-standard abbreviations and new definitions as possible were used.

The interval $[c, d]$ is defined for both $c > d$ and $d \leq c$. In case $c > d$, just set $[c, d] := [d, c]$.

The interval $]c, d[$ stands for $[c, d]$, but the points c and d are excluded from the interval.

In the expression

$$(A, B)^T = \begin{pmatrix} A \\ B \end{pmatrix}$$

the letter T denotes the transpose.

$$(A, B) = (X, Y) \Leftrightarrow A = X \text{ and } B = Y$$

2.2 General Model of the process

Only homogeneous Markov chains with discrete time, finite state space and fixed inital state are considered. For a short definition of these terms for one dimension see section 2.2.1 or for a more detailed description Norris [1]. Section 2.2.2 gives the definitions in 2 dimensions.

2.2.1 1 Dimension

The construction of a Markov chain requires two basic ingredients, namely a transition matrix and an initial distribution. To begin with, consider the definition of the transition matrix. Assume a finite set $S = \{1, \ldots, m\}$ of states. Assign to each pair $(i, j) \in S^2$ of states a real number p_{ij} such that the properties

$$p_{ij} \geq 0 \ \forall (i, j) \in S^2$$
$$\sum_{j \in S} p_{ij} = 1 \ \forall i \in S$$

are satisfied and define the transition matrix P by

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

Definition 2.2.1. (Markov chain) The sequence of random variables$(X_n)_{n \in \mathbb{N}_0}$ with $X_i \in S$ is called a homogeneous Markov chain with discrete time and inital state X_0, state space S, and transition matrix P, if for every $n \in \mathbb{N}_0$ the condition

$$\mathbf{prob}[X_{n+1} = j | X_0 = i_0, \ldots, X_n = i_n] = \mathbf{prob}[X_{n+1} = j | X_n = i_n] = p_{i_n j}$$

is satisfied for all $(i_0, \ldots, i_n, j) \in S^{n+2}$, for which

$$\mathbf{prob}[X_0 = i_0, \ldots, X_n = i_n] > 0$$

The first identity in definition (2.2.1), which is also called “Markov property”, defines the “memory” or “order” of the chain. In this case, the order equals one since the transition probabilities are entirely determined by the preceding state. The restriction to order-one chains is no serious limitation since processes with arbitrary finite memory s can be interpreted as order-one Markov chains on the product space S^s. The second identity in the above definition is called homogeneity condition. It assures that the transition probabilities do not vary with the time n.

2.2.2 2 Dimensions

The concepts are the same as for the 1 dimensional case, thus only the definition is given.

Definition 2.2.2. (Markov Chain) The sequence of random variables (A_n, B_n) with $n \in \mathbb{N}_0$ and $(A_n, B_n) \in S_A \times S_B$ is called a homogeneous Markov chain with discrete time and inital state (A_0, B_0), state space $S = S_A \times S_B$, and transition matrix P, if for every $n \in \mathbb{N}_0$ the condition

$$\begin{aligned}&\mathbf{prob}[(A_{n+1}, B_{n+1}) = j|(A_0, B_0) = i_0, \ldots, (A_n, B_n) = i_n]\\&= \mathbf{prob}[(A_{n+1}, B_{n+1}) = j|(A_n, B_n) = i_n] = p_{i_n j}\end{aligned}$$

is satisfied for all $(i_0, \ldots, i_n, j) \in S^{n+2}$, for which $\mathbf{prob}[(A_0, B_0) = i_0, \ldots, (A_n, B_n) = i_n] > 0$.

2.3 Pseudo expectation

The notion of the pseudo expectation (and some basic properties of it) as well as the error between the real and the pseudo expectation are concretized in 1 (see 2.3.1) and 2 dimensions (see 2.3.2).

2.3.1 1 Variable

Given a Markov chain as defined in (2.2.1) and let f be a function such that

$$\mathbb{E}[X_{n+1}|X_n] = f(X_n), a.s. \tag{2.3.1}$$

then the pseudo expectation f^n is just the iterated application of f for the constant inital state X_0, e.g. $\underbrace{f(f(\ldots f}_{n-times}(X_0)\ldots))$.

The steady state expectation is approximated by the fix point of the function f. More precisely, if $0 \leq f'(x) < 1$ for all $x \in I_S$, where

$$I_S := [\min_s S, \max_s S]$$

then f is contract and has only one fix point b_f, given by the solution of the equation:

$$f(b_f) = b_f \tag{2.3.2}$$

This statement is well-known in literature (see 3.1.1).

The error of the approximation is simply given by

$$err(n, a) := \mathbb{E}[X_n|X_0 = a] - f^n(a) \tag{2.3.3}$$

Note that, the error after the first step is zero due to the definition of f (2.3.1).

If f is linear

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1}|X_n]] = \mathbb{E}[f(X_n)] = f(\mathbb{E}[X_n])$$

and therefore for all n:

$$\mathbb{E}[X_n] = f^n(X_0) \tag{2.3.4}$$

But for a non-linear f in general $\mathbb{E}[X_n] \neq f^n(X_0)$.

2.3.2 2 Variables

The definition of the pseudo expectation for a Markov chain as defined in (2.2.2) is a straight forward extension of the case, where the pseudo expectation depends on 1 variable.

In higher dimensions the expectation becomes a vector of functions e.g.

$$\overrightarrow{\mathbb{E}}[A, B] = (\mathbb{E}_A[A, B], \mathbb{E}_B[A, B])^T$$

Let $\overrightarrow{f} = (f_A(A,B), f_B(A,B))^T$ be a function such that

$$\overrightarrow{f}(A_i, B_i) = \overrightarrow{\mathbb{E}}[A_{i+1}, B_{i+1}|A_i, B_i] \tag{2.3.5}$$

then the pseudo expectation $\overrightarrow{f^n}$ is given by:

$$\overrightarrow{f^n}(A,B) = (f_A^n(A,B), f_B^n(A,B))^T$$

with f_A^n defined as

$$f_A^n(A,B) = f_A^{n-1}(f(A,B)^T)$$

f_B^n is analogous.

Define the interval I_{S_A} as $[\min_{x \in S_A} x, \max_{x \in S_A} x]$ and I_{S_B} analogously and also

$$I_S := I_{S_A} \times I_{S_B}$$

Again, the steady state expectation is approximated by the fix point of the function $\overrightarrow{f}$. More precisely, if $0 \leq \frac{\partial f(a,b)}{\partial a} < 1$ and $0 \leq \frac{\partial f(a,b)}{\partial b} < 1$ for all $(a,b) \in I_{S_A} \times I_{S_B}$, then $\overrightarrow{f}$ is contract and has only one fix point $\overrightarrow{b^f} = (b_A^f, b_B^f)^T$, given by the solution of the equation:

$$\overrightarrow{f}(\overrightarrow{b_f}^T) = \overrightarrow{b_f} \tag{2.3.6}$$

This statement is well-known in literature (see 3.1.1).

The error $\overrightarrow{err}(n,a,b) = (err_A(n,a,b), err_B(n,a,b))^T$ of the approximation is simply given by

$$\overrightarrow{err}(n,a,b) := \overrightarrow{\mathbb{E}}[A_n, B_n|(A_0, B_0) = (a,b)] - \overrightarrow{f^n}(a,b) \tag{2.3.7}$$

As for the one dimensional case, the error after the first step is zero due to definition (see 2.3.5).

If the function f_A and f_B are linear then the approximation is exact (for a proof see (4.3.1)).

2.4 Example process

To illustrate the techniques for the error analysis, the following game is used throughout the document:

1. Player A and B have $A_n \geq 0$ and $B_n \geq 0$ balls at time n respectively.
2. If A (B) wins, A (B) gives $\min\{d, A_n\}$ ($\min\{d, B_n\}$) balls to B (A). (Let the total number of balls N divided by d be an integer).

   For the rest of the document d is implicitly assumed to be 1, if it is omitted.
3. The probability of the event A wins depends on A_n. Then

$$\forall n, A_n + B_n = N \tag{2.4.1}$$

   Let $\mathbf{prob}(A_n)$ and $\mathbf{prob}(B_n)$ be the probability of the event A wins and B wins respectively. Then

$$\forall n, \mathbf{prob}(A_n) + \mathbf{prob}(B_n) = 1$$

$$\mathbb{E}\left[\begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} \middle| \begin{pmatrix} A_n \\ B_n \end{pmatrix}\right] = \begin{pmatrix} A_n \\ B_n \end{pmatrix} + d \cdot \begin{pmatrix} -\mathbf{prob}(A_n) + \mathbf{prob}(B_n) \\ \mathbf{prob}(A_n) - \mathbf{prob}(B_n) \end{pmatrix}$$

   The winning probability for A_n is given by

$$\begin{aligned} \mathbf{prob}(A_n) &= \frac{w \cdot A_n}{w \cdot A_n + B_n} \\ &= 1 - \frac{w \cdot A_n}{(w-1) \cdot A_n + N} \qquad \text{due to (2.4.1)} \end{aligned}$$

   and for B_n by

$$\mathbf{prob}(B_n) = \frac{B_n}{w \cdot A_n + B_n}$$

where the weight w is a positive constant such that $1 < w < \frac{N}{2 \cdot d}$. Due to (2.4.1) and because of the linearity of expectation we have

$$\forall n, \mathbb{E}[A_n] + \mathbb{E}[B_n] = N$$

For that reason, it is sufficient to consider $\mathbb{E}[A_n]$ to obtain the state of the game. The function f becomes:

$$\begin{aligned} f(A_n) &= \mathbb{E}[A_{n+1} | A_n]] \\ &= d \cdot (\mathbf{prob}(B_n) - \mathbf{prob}(A_n)) \\ &= d \cdot (1 - \frac{2 \cdot w \cdot A_n}{(w-1) \cdot A_n + N}) \end{aligned}$$

Remarks:

1. For $w = 1$ the function f is linear in A_n:

$$f(A_n) = (1 - \frac{2 \cdot d}{N}) \cdot A_n + d$$

   Thus the approximation is correct due to (2.3.4).

2. The range $0 < w < 1$ is symmetric in the sense that B_n instead of A_n is weighted by $\widetilde{w} = \frac{1}{w} \in]1, \infty[$, e.g. the probabilities of A and B wins are given by:

$$\begin{aligned} \mathbf{prob}(A_n) &= \frac{w \cdot A_n}{w \cdot A_n + B_n} = \frac{A_n}{\widetilde{w} \cdot B_n + A_n} \\ \mathbf{prob}(B_n) &= \frac{B_n}{w \cdot A_n + B_n} = \frac{\widetilde{w} \cdot B_n}{A_n + \widetilde{w} \cdot B_n} \end{aligned}$$

For the considered interval of w, the function f is contract, e.g. $0 \leq f'(x) < 1$ for any $x \in [0, N]$. Besides, the first derivative f' and the second f'' are also continuous for $x \in [0, N]$ as can be seen below:

$$f'(x) = d \cdot (1 - \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot x)^2}) < 1 \tag{2.4.2}$$

and

$$f''(x) = \frac{4 \cdot d \cdot w \cdot (w-1) \cdot N}{(N + (w-1) \cdot x)^3}$$

Since f is contract the fix point b_f can be calculated using (2.3.2):

$$\begin{aligned} f(b_f) &= b_f \\ \Leftrightarrow b_f &= \frac{N}{w+1} \end{aligned} \tag{2.4.3}$$

2.4.1 Illustration

To get a better feeling of the behavior of the error and the process, some numerical instances and figures are discussed for the example process.

Numerical results are given for the error, the variance of the process and the behavior of the probability distribution over time. This is done for different weights and starting states. The number of balls N is fixed to 40, since for big N and big number n of steps the computations take too long. (In each step 2 matrixes have to be multiplied, which gives a running time of $O(N^{2.376} \cdot n)$).

To begin with the case $A_0 = N = 40$ for $w = 2$ and $w = 10$ is considered. The pseudo and real expectation are quite close (figure 2.1 below). By looking at figure 2.2, it can be seen that the error has a maximum. The same also holds for the variance (see figure 2.3). To capture this non-monotone behavior, especially bound the maximum tightly, can be regarded as the "ultimate" goal.

The figures 2.4 and 2.5 for the probability distribution reveal that the maximum probability as a function of n, e.g. $\max_{a \in [0,N]} \mathbf{prob}(A_n)$ has a minimum for some n, which is roughly $\approx N$. Besides for a fixed step n the probabilities decay very fast when moving away from the most likely state.

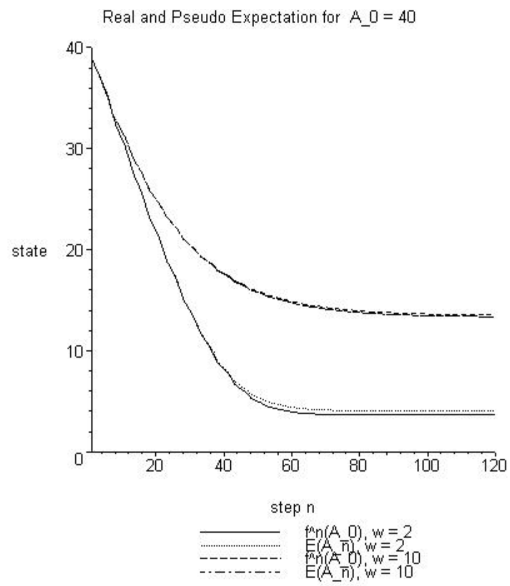

Figure 2.1: Pseudo and real expectation for $A_0 = N$

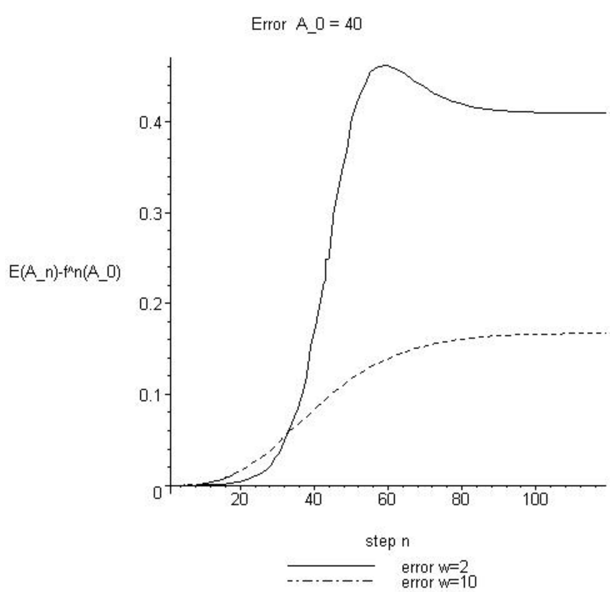

Figure 2.2: Error for $A_0 = N$

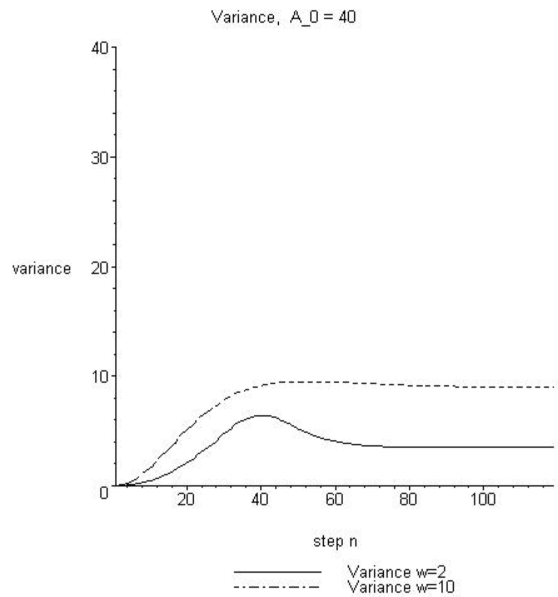

Figure 2.3: Variance for $A_0 = N$

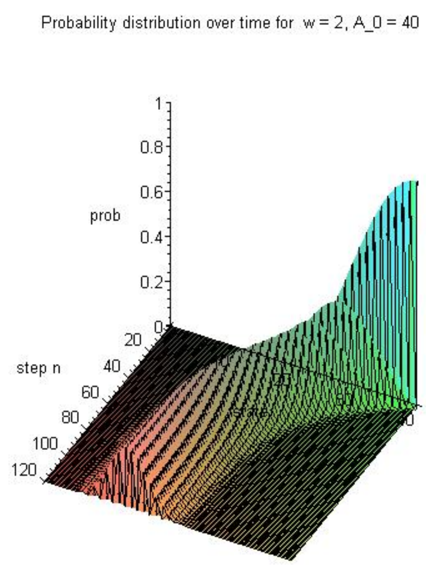

Figure 2.4: Probability distribution of states for $A_0 = N$ and $w = 2$

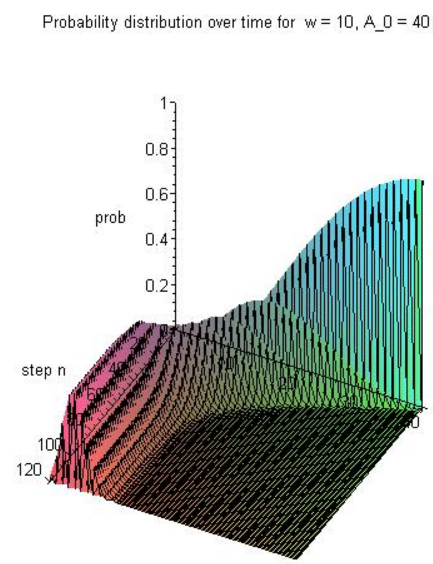

Figure 2.5: Probability distribution of states for $A_0 = N$ and $w = 10$

For the sake of completeness, the same plots but for $A_0 = 0$ and $w \in \{2, 10\}$ are also shown (figure 2.6, 2.7 and 2.8). For this case the maxima for the error and the variance seem to be very small or non-existent.

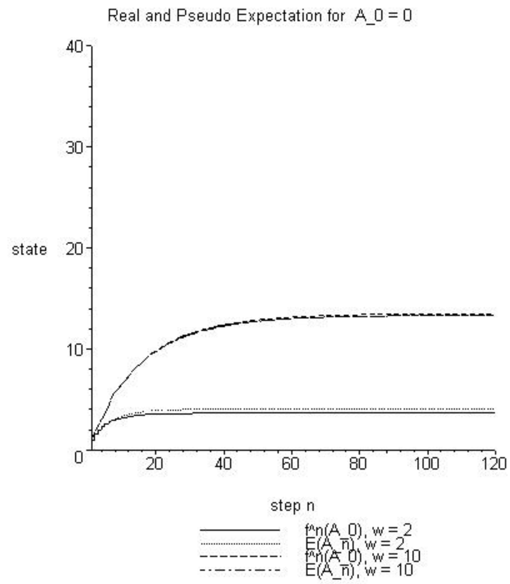

Figure 2.6: Pseudo and real expectation for $A_0 = 0$

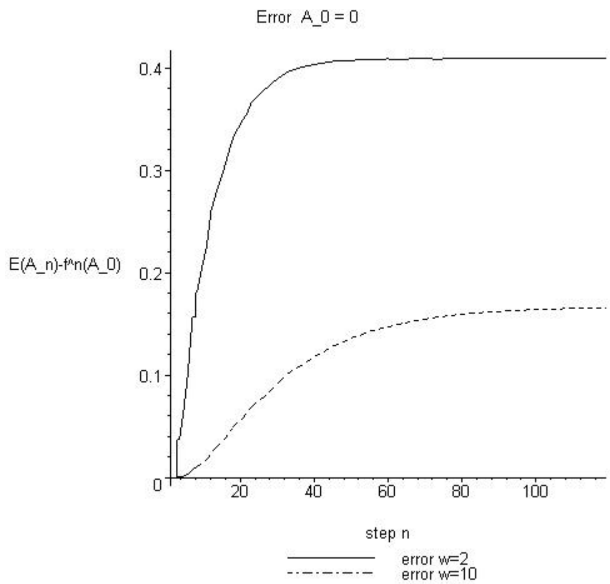

Figure 2.7: Error for $A_0 = 0$

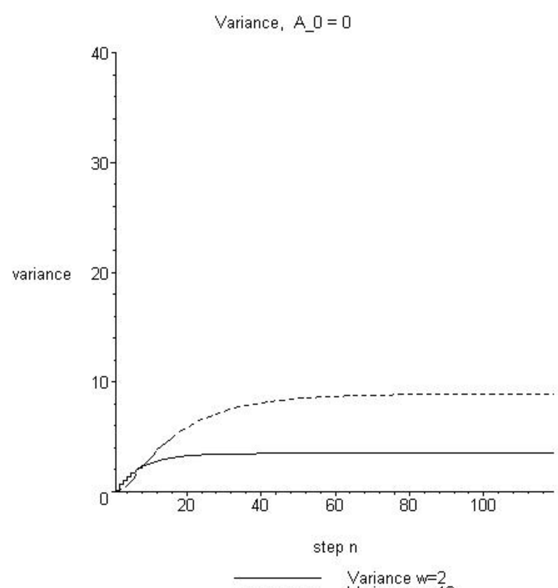

Figure 2.8: Variance for $A_0 = 0$

Chapter 3

Some properties of the pseudo expectation

In this chapter several characteristics of the pseudo expectation in 1 variable are derived. But apart from the convergence speed of the pseudo expectation in section 3.1, these properties (namely "steady state distribution and error for steady state" in 3.2, "relation between the state with maximum probability and the fix point" in 3.3) are more or less specific for the considered example (defined in 2.4).

3.1 Bound on the convergence speed of the pseudo expectation

Two general techniques are applied to the example (see section 2.4). Using the Banach fix point theorem 3.1.1 will give a first bound on the convergence speed of the pseudo expectation. Later on, it will be improved (section 3.1.2), using integrals.

In fact, the number of steps until the distance to the fix point is less than 1 is investigated, since it takes ∞ many steps to actually reach the fix point.

3.1.1 Application of Banach fix point theorem

The Banach fixed point theorem 3.1.1 is well known in literature. It also gives a bound on the speed of convergence for a contraction mapping such as the function f defining the pseudo expectation.

Theorem 3.1.1. *(Banach Fixed Point theorem) Let X be a non-empty complete metric space. Let $G : X \to X$ be a contraction mapping on X , i.e, there exists a real number $q < 1$ such that $\|G(x) - G(y)\| \leq q\|x - y\|$ for all $x, y \in X$. Then the map G admits one and only one fixed point $\tilde{x} \in X$ (this means $G(\tilde{x}) = \tilde{x}$). Furthermore, this fixed point can be found as follows: start with an arbitrary element $x^{(0)}$ (and define a sequence by $x^n = G(x^{(n-1)})$ for*

$n = 1, 2, 3, \ldots$ *This sequence converges, and its limit is* $\tilde{x}$. *The following inequality describes the speed of convergence:*

$$\|\tilde{x} - x^{(n)})\| \leq \frac{q^n}{1-q}\|x^{(1)} - x^{(0)}\|$$

(For a proof, see for instance, Agarwal [5])

Lemma 3.1.2. *Let* $A_0, A_1, \ldots, A_n$ *be a Markov process as defined in section 2.4 and also* $\frac{N}{2} > w > 1$, *then for the pseudo expectation with fix point* b_f *holds:*

If $A_0 > b_f$ *then after*

$$n = \frac{w \cdot N}{2} \cdot \ln(\frac{w \cdot N}{2})$$

steps and if $A_0 \leq b_f$ *then after*

$$n = \frac{2 \cdot N}{w} \cdot \ln(\frac{2 \cdot N}{w})$$

steps

$$|f^n(A_0) - b_f| \leq 1$$

Proof. Applying theorem 3.1.1 for the case $f^i(A_0) \in [b_f, N]$, yields for q:

$$\begin{aligned} q &\leq \max_{\in[b_f,N]} f'(x) \\ &= f'(N) \\ &= 1 - \frac{2}{w \cdot N} \end{aligned} \tag{3.1.1}$$

Since at most one ball can be taken per step, the difference between any two steps is at most 1. In particular for the starting state and the state after the first step holds:

$$\|x^{(1)} - x^{(0)}\| \leq 1$$

Using this and the previous inequality (3.1.1) for q for $n = \frac{w \cdot N}{2} \cdot \ln(\frac{w \cdot N}{2})$ steps gives:

$$\begin{aligned} \frac{q^n}{1-q}\|x^{(1)} - x^{(0)}\| &\leq \frac{(1 - \frac{2}{w \cdot N})^{\frac{w \cdot N}{2} \cdot \ln(\frac{w \cdot N}{2})}}{\frac{2}{w \cdot N}} \\ &\leq \frac{e^{-\ln(\frac{w \cdot N}{2})}}{\frac{2}{w \cdot N}} \\ &\leq 1 \end{aligned}$$

For $f^i(A_0) \in [0, b_f]$, the constant q can be bound in the same way:

$$
\begin{aligned}
q &\le \max_{\in[0,b_f]} f'(x) \\
&= f'(b_f) \\
&= 1 - \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot \frac{N}{w+1})^2} \\
&= 1 - \frac{2 \cdot w}{N \cdot (1 + \frac{w-1}{w+1})^2} \\
&= 1 - \frac{(w+1)^2}{2 \cdot w \cdot N} \\
&\ge 1 - \frac{w}{2 \cdot N} \qquad (3.1.2)
\end{aligned}
$$

The difference between two steps is - as before - bounded by 1:

$$\|x^{(1)} - x^{(0)}\| \le 1$$

Combining that with the inequality (3.1.2) for $n = \frac{2 \cdot N}{w} \cdot \ln(\frac{2 \cdot N}{w})$ steps yields:

$$
\begin{aligned}
\frac{q^n}{1-q}\|x^{(1)} - x^{(0)}\| &\le \frac{(1 - \frac{w}{2 \cdot N})^{\frac{2 \cdot N}{w} \cdot \ln(\frac{2 \cdot N}{w})}}{\frac{w}{2 \cdot N}} \\
&\le \frac{e^{-\ln(\frac{2 \cdot N}{w})}}{\frac{w}{2 \cdot N}} \\
&\le 1
\end{aligned}
$$

Thus, lemma 3.1.2 is proven. □

3.1.2 Improved bound

Next, a different technique is used to bound the number of steps to get close to the fix point. The idea is to consider the number of steps, which are needed to get from state a to $a \pm 1$ and then sum over all a from the starting point to the fix point ± 1. This technique requires apart from the continuity of f that $|f(a) - a|$ decreases monotonically, when a gets closer to the fix point.

Theorem 3.1.3. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.4 and also $\frac{N}{2} > w > 1$, then for the pseudo expectation with fix point b_f holds:*

If $A_0 < b_f$ then after

$$n = \frac{2 \cdot N}{w} \cdot \ln(\frac{N}{w})$$

steps and if $A_0 \ge b_f$ then after

$$n = \frac{2 \cdot N}{w} \cdot (\ln(N) + \frac{w}{2})$$

steps

$$|f^n(A_0) - b_f| \le 1$$

Proof. Consider $a \in [b_f, N]$ first.

Denote the number of steps to go from state a to $a \pm 1$ by $st_{a\pm1}$. Then for $st_{a\pm1}$ must hold:

$$f^{st_{a\pm1}}(a) \leq a \pm 1$$

Since $f(a) - a$ decreases with decreasing a for $a \in [b_f, N]$, it follows

$$st_{a-1} \leq \frac{1}{f(a-1) - (a-1)}$$

(For more details, see also the remark 3.1.1) Thus the total number of steps st_{tot} to get from N to $b_f + 1$ becomes

$$\begin{aligned}
st_{tot} &= \sum_{a=b_f+2}^{N} st_{a-1} \\
&\leq \sum_{a=b_f+2}^{N} \frac{1}{f(a-1)-(a-1)} \\
&\leq \int_{a=b_f+1}^{N-1} \frac{1}{f(a)-a} \text{ Since } f(a)-a \text{ decreases with decreasing } a \text{ for } a \in]b_f, N] \\
&\leq \int_{a=b_f+1}^{N-1} \frac{1}{1-\frac{2\cdot w\cdot a}{N+(w-1)\cdot a}} \\
&\leq \int_{a=b_f+1}^{N} \frac{N+(w-1)\cdot a}{N-(w+1)\cdot a} \\
&\leq N \cdot \int_{a=b_f+1}^{N} \frac{1}{N-(w+1)\cdot a} \cdot da + (w-1) \cdot \int_{a=b_f+1}^{N} \frac{a}{N-(w+1)\cdot a} \cdot da \quad (3.1.3)
\end{aligned}$$

The integrals can easily be solved by substituting $u = N - (w+1) \cdot a$:

$$a = \frac{N-u}{w+1}$$

and

$$\frac{du}{da} = -(w+1)$$

Thus

$$\begin{aligned}
\int_{a=b_f+1}^{N} \frac{1}{N-(w+1)\cdot a} \cdot da &= \frac{-1}{w+1} \int_{u=-(w+1)}^{-w\cdot N} \frac{1}{u} \cdot du \\
&= \frac{1}{w+1} \cdot \ln(\frac{w \cdot N}{w+1}) \\
&\geq \frac{\ln(N)}{w}
\end{aligned}$$

and also

$$\begin{aligned}\int_{a=b_f+1}^{N}\frac{a}{N-(w+1)\cdot a}\cdot da &= \frac{-1}{w+1}\cdot\int_{u=-(w+1)}^{-w\cdot N}\frac{\frac{N-u}{w+1}}{u}\cdot du\\ &= \frac{-1}{(w+1)^2}\cdot\int_{u=-(w+1)}^{-w\cdot N}(\frac{N}{u}-1)\cdot du\\ &= \frac{1}{(w+1)^2}\cdot\Big(N\cdot\ln(N)+w\cdot N-(w+1)\Big)\\ &\leq \frac{N\cdot\ln(N)}{w^2}+\frac{N}{w}\end{aligned}$$

Plugging the results for the integrals in the previous formula (3.1.3) gives:

$$\begin{aligned}st_{tot} &\leq N\cdot\int_{a=b_f+1}^{N}\frac{1}{N-(w+1)\cdot a}\cdot da+(w-1)\cdot\int_{a=b_f+1}^{N}\frac{a}{N-(w+1)\cdot a}\cdot da\\ &\leq N\cdot\frac{\ln(N)}{w}+(w-1)\cdot\Big(\frac{N\cdot\ln(N)}{w^2}+\frac{N}{w}\Big)\\ &\leq \frac{2\cdot N}{w}\cdot(\ln(N)+\frac{w}{2})\end{aligned}$$

The case $a\in[0,b_f]$ is treated in the same fashion.
The basic calculations will be given:

$$\begin{aligned}st_{tot} &= \sum_{a=0}^{b_f-2} st_{a+1}\\ &\leq \sum_{a=0}^{b_f-2}\frac{1}{f(a+1)-(a+1)}\\ &\leq \int_{a=1}^{b_f-1}\frac{1}{f(a)-a}\ \text{Since } f(a)-a \text{ decreases with increasing } a \text{ for } a\in[0,b_f[\\ &\leq \int_{a=0}^{b_f-1}\frac{1}{1-\frac{2\cdot w\cdot a}{N+(w-1)\cdot a}}\\ &\leq N\cdot\int_{a=0}^{b_f-1}\frac{1}{N-(w+1)\cdot a}\cdot da\\ &\quad +(w-1)\cdot\int_{a=0}^{b_f-1}\frac{a}{N-(w+1)\cdot a}\cdot da \qquad (3.1.4)\end{aligned}$$

Again the integrals can be solved by the same substitution and will be bounded as follows:

$$\begin{aligned}\int_{a=0}^{b_f-1}\frac{1}{N-(w+1)\cdot a}\cdot da &= \frac{-1}{w+1}\int_{u=N}^{w+1}\frac{1}{u}\cdot du\\ &= \frac{1}{w+1}\ln(\frac{N}{w+1})\\ &\geq \frac{\ln(\frac{N}{w})}{w}\end{aligned}$$

and also

$$\begin{aligned}
\int_{a=0}^{b_f-1} \frac{a}{N-(w+1)\cdot a}\cdot da &= \frac{-1}{w+1}\cdot\int_{u=-(w+1)}^{-w\cdot N} \frac{\frac{N-u}{w+1}}{u}\cdot du \\
&= \frac{-1}{(w+1)^2}\cdot\int_{u=N}^{w+1} (\frac{N}{u}-1)\cdot du \\
&\leq \frac{1}{(w+1)^2}\cdot\left(N\cdot\ln(\frac{N}{w})+w+1-N\right) \\
&\leq \frac{N\cdot\ln(\frac{N}{w})}{w^2}
\end{aligned}$$

Analogously, plugging these results for the integrals in the previous formula (3.1.4) gives:

$$\begin{aligned}
st_{tot} &\leq N\cdot\int_{a=0}^{b_f-1}\frac{1}{N-(w+1)\cdot a}\cdot da+(w-1)\cdot\int_{a=0}^{b_f-1}\frac{a}{N-(w+1)\cdot a}\cdot da \\
&\leq N\cdot\frac{\ln(\frac{N}{w})}{w}+(w-1)\cdot\frac{N\cdot\ln(\frac{N}{w})}{w^2} \\
&\leq \frac{2\cdot N}{w}\cdot\ln(\frac{N}{w})
\end{aligned}$$

This completes the proof of theorem 3.1.3. □

Remark 3.1.1. A lower bound can also be calculated by considering $\frac{1}{|f^{-1}(a)-a|}$ instead of $\frac{1}{|f(a)-a|}$. This can best be seen by looking at figure 3.1. In this figure area II A_{II}, meaning the area in the rectangle, gives the true number of steps needed to get from a to $f(a)$, which is obviously 1:

$$A_{II}=\frac{1}{|f(a)-a|}\cdot|f(a)-a|=1$$

The hatched area shows the error for the upper bound e.g. the difference between the integral $\int_{x=a}^{f(a)}\frac{1}{|f(x)-x|}$ and the correct value A_{II}:

$$\int_{x=a}^{f(a)}\frac{1}{|f(x)-x|}dx-A_{II}$$

The doted area gives the error for the newly introduced lower bound, e.g. the difference between the integral $\int_{x=a}^{f(a)}\frac{1}{|f^{-1}(x)-x|}$ and the correct value A_{II}.

$$\int_{x=a}^{f(a)}\frac{1}{|f^{-1}(x)-x|}\cdot dx-A_{II}=\int_{y=f^{-1}(a)}^{a}\frac{1}{|f(y)-y|}\cdot f'(y)\cdot dy-A_{II}$$

where the substitution of $x=f(y)$ and consequently $\frac{dx}{dy}=f'(y)$ leads to the second expression. Thus a lower bound for the total number of steps to get from N to $b_f\pm 1$ can be obtained.

Theorem 3.1.4. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.4 and also $\frac{N}{2} > w > 1$, then for the pseudo expectation with fix point b_f holds:*

If $A_0 < b_f$ then after at least

$$n = \frac{N}{2 \cdot w} \cdot 2 \cdot \ln(\frac{w \cdot N}{3 \cdot (w+1)}) - \ln(\frac{N}{w}) \qquad w \geq 3$$

steps and if $A_0 \geq b_f$ then after at least

$$n = \frac{N}{2 \cdot w} \cdot \left(w - 1 + \ln(\frac{N}{4}) \right) - \frac{3}{2} - \ln(1 + \frac{N}{w})$$

steps

$$|f^n(A_0) - b_f| \leq 1$$

Proof. Again the case $A_0 \geq b_f$ will be investigated first.

As for the upper bound, denote the number of steps to go from state a to $a \pm 1$ by $st_{a\pm1}$. Then for $st_{a\pm1}$ to be a lower bound must hold:

$$f^{st_{a\pm1}}(a) \geq a \pm 1$$

Due to the monotone behavior of f this can be written as: (see also beginning of this section together with 3.1)

$$st_{a-1} \geq \int_{x=a}^{a-1} \frac{1}{|f^{-1}(x) - x|} dx$$

Thus the total number of steps st_{tot} to get from N to $b_f + 1$ becomes

$$\begin{aligned}
st_{tot} &\geq \int_{x=N}^{bf+2} \frac{1}{|f^{-1}(x) - x|} dx \\
&= \int_{y=f^{-1}(N)}^{f^{-1}(bf+2)} \frac{1}{|f(y) - y|} \cdot f'(y) dy \\
&\geq \int_{y=N}^{bf+3} \frac{1}{|f(y) - y|} \cdot f'(y) dy \\
&= \int_{y=N}^{bf+3} \frac{1}{|f(y) - y|} dy + \int_{y=N}^{bf+3} \frac{1}{|f(y) - y|} \cdot (f'(y) - 1) dy \\
&= \int_{y=N}^{bf+3} \frac{N + (w-1) \cdot y}{N - (w+1) \cdot y} dy \\
&\quad - \int_{y=N}^{bf+3} \frac{N + (w-1) \cdot y}{N - (w+1) \cdot y} \cdot \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot y)^2} dy \\
&= \int_{y=N}^{bf+3} \frac{N + (w-1) \cdot y}{N - (w+1) \cdot y} dy \\
&\quad - \int_{y=N}^{bf+3} \frac{2 \cdot w \cdot N}{(N - (w+1) \cdot y) \cdot (N + (w-1) \cdot y)} dy \qquad (3.1.5)
\end{aligned}$$

The first integral in 3.1.5 is almost the same as before in 3.1.3 apart from the integration limits.

Using the same technique (namely substitution) again, we have:

$$
\begin{aligned}
st_{tot} \quad &\geq \quad \frac{w \cdot N \cdot \left(w - 1 + \ln(\frac{w \cdot N}{3 \cdot (w+1)}) \right) - 3 \cdot (w^2 - 1)}{(w+1)^2} \\
&\quad + \ln\left(\frac{3 \cdot (w+1)^2}{2 \cdot w \cdot N + 3 \cdot (w^2 - 1)} \right) \\
&\geq \quad \frac{N}{2 \cdot w} \cdot \left(w - 1 + \ln(\frac{N}{4}) \right) - \frac{3}{2} - \ln(1 + \frac{N}{w}) \qquad w \geq 3 \qquad (3.1.6)
\end{aligned}
$$

The case where the starting state $A_0 \leq b_f$ is analyzed in the same manner:

$$
\begin{aligned}
st_{tot} \quad &\geq \quad \int_{x=0}^{bf-2} \frac{1}{|f^{-1}(x) - x|} dx \\
&= \quad \int_{y=f^{-1}(0)}^{f^{-1}(bf-2)} \frac{1}{|f(y) - y|} \cdot f'(y) dy \\
&\geq \quad \int_{y=0}^{bf-3} \frac{1}{|f(y) - y|} \cdot f'(y) dy
\end{aligned}
$$

Thus we get:

$$
\begin{aligned}
st_{tot} \quad &\geq \quad \frac{w \cdot N \cdot \left(2 \cdot \ln(\frac{w \cdot N}{3 \cdot (w+1)}) - 1 \right) + N + 3 \cdot (w^2 - 1)}{(w+1)^2} \\
&\quad + \ln\left(\frac{3 \cdot (w+1)^2}{2 \cdot w \cdot N - 3 \cdot (w^2 - 1)} \right) \\
&\geq \quad \frac{N}{2 \cdot w} \cdot 2 \cdot \ln(\frac{w \cdot N}{3 \cdot (w+1)}) - \ln(\frac{N}{w}) \qquad w \geq 3 \qquad (3.1.7)
\end{aligned}
$$

□

To obtain a tighter bound, one could use a function $c(a)$ which satisfies the condition:

$$
c(a) \cdot \int_{x=a}^{f(a)} \frac{1}{|f(x) - x|} dx = 1
$$

and then calculate the number of steps from d to e by

$$
st_{totap} = \int_{x=d}^{e} |c(x) \cdot \frac{1}{f(x) - x}| dx
$$

which gives

$$
st_{tot} - 1 < st_{totap} < st_{tot} + 1
$$

Since the integrals for the example are not so easy to handle, this approach is not demonstrated.

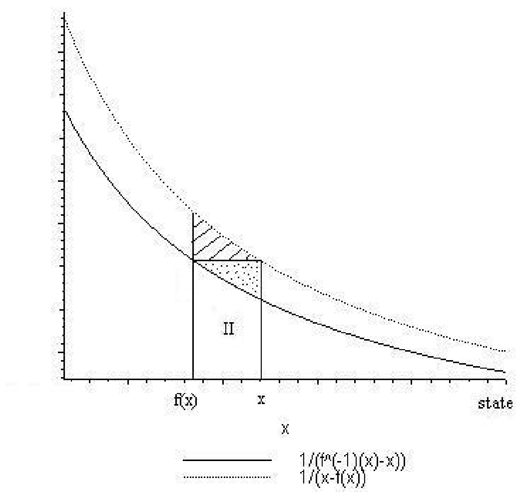

Figure 3.1: Pseudo and real expectation for $A_0 = N$

3.2 Steady state distribution

This chapter consists of some general remarks about the steady state distribution in section 3.2.1 and a deduction of it for the example process in 3.2.2 (and moreover an error bound for the steady state is given).

3.2.1 General

In general the steady state distribution $\mathbf{p_s}$ of a Markov chain with transition matrix P has to satisfy the following condition:

$$P \cdot \mathbf{p_s} = \mathbf{p_s} \tag{3.2.1}$$

For $\mathbf{p_s}$ to be a distribution, it must hold that

$$\sum_{j=0}^{N} \mathbf{p_s}(A = j) = 1 \tag{3.2.2}$$

$$\forall j \in [0, N]\ \mathbf{p_s}(A = j) \geq 0 \tag{3.2.3}$$

For any application, where the number of states is a parameter t (e.g. our example process described in section 2.4), these conditions give a system of linear equations, where the number of equations also depends on this parameter and (usually) cannot be easily solved without the knowledge of t.

Remark: As an alternative to looking at linear equations, in some cases an eigenvalue analysis of P might be preferable (for details see Mehdi [3], p.100 and following).

Obviously, once the steady state distribution is obtained, the expectation can be calculated. The error for the steady state is then just the difference between the fix point b_f and the true expectation.

3.2.2 Example

For the considered example (see section 2.4), in general the expectation of the process does not converge to a single value, but changes between two converging series. Note that given the process is in some state i, then in the next state the probability of this state will always be 0, since the Markov process leaves this state with probability 1. Mathematically speaking, $\mathbf{prob}(B_n = i) > 0 \Rightarrow \mathbf{prob}(B_{n+1} = i) = 0$. The Markov process has period 2. For that reason, the limit of P^n, where P denotes the transition matrix, does not exist.

But if one looks at the Markov process at steps with an even and odd number separately and considers $\lim_{n\to\infty} \mathbf{prob}(P^{2\cdot n})$ and respectively $\lim_{n\to\infty} \mathbf{prob}(P^{2\cdot n+1})$, then these limits exist.

Theorem 3.2.1. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.4, let $N = N_0 \cdot d$ with $N_0 \in \mathbb{N}$, let A_0 also be a multiple of d and denote $\mathbf{prob}(A_n)$ as the probability distribution of the state at step n, then the steady state distribution $\mathbf{p_s}$ is*

$$\mathbf{p_s}(A = j) = \begin{cases} \frac{1}{2} \frac{w^{N_0-i-1} \cdot (N_0 + i \cdot (w-1))}{N_0 \cdot (1+w)^{N_0-1}} \cdot \binom{N_0}{i} & \text{if } j = i \cdot d,\ i \in \mathbf{N} \\ 0 & \text{otherwise} \end{cases}$$

and furthermore

$$\lim_{n\to\infty} \mathbf{prob}(A_{2n} = j) = \begin{cases} 2 \cdot \mathbf{p_s}(A = j) & \textit{if } A_0 \textit{ even} \\ 0 & \textit{otherwise} \end{cases} \tag{3.2.4}$$

$$\lim_{n\to\infty} \mathbf{prob}(A_{2n+1} = j) = \begin{cases} 2 \cdot \mathbf{p_s}(A = j) & \textit{if } A_0 \textit{ odd} \\ 0 & \textit{otherwise} \end{cases}$$

Remarks:

1. For $w = 1$ the steady state distribution becomes the binomial distribution:

$$\mathbf{p_s}(A = j) = \binom{N}{j} \cdot 2^{-N} \tag{3.2.5}$$

2. Since the proof given below, is more or less just a verification of the conditions (e.g. 3.2.1, 3.2.2 and 3.2.3), required for $\mathbf{p_s}$ to hold, the following question remains: How can one deduce such a formula? In fact it was obtained by considering several examples with fixed number of states (e.g. balls N) and solving the linear system given by the previously mentioned conditions.

Proof. At first, it will be shown that $\mathbf{p_s}(A = j)$, which is assumed to be written as a row vector, is indeed the steady state distribution of the transition matrix P. More precisely, $\mathbf{p_s}$ satisfies the conditions for being a steady state distribution (3.2.1, 3.2.2 and 3.2.3).

In general the transition matrix P looks like:

$$P = \begin{pmatrix} & p_{0,d} & & & & \\ & & \ddots & 0 & & \\ & & & \ddots & & \\ p_{d,0} & & & & & \\ & \ddots & 0 & \ddots & & \\ & & \ddots & & p_{N-d,N} & \\ & & & \ddots & & \\ & 0 & & p_{N,N-d} & & \end{pmatrix}$$

The steady state equations (3.2.1) expressed in terms of $\mathbf{p_s}(A = i)$ become:

$$\mathbf{p_s}(A = i) = \begin{cases} \mathbf{p_s}(A = d) \cdot \mathbf{prob}(A = 0 | A = d) & \text{if } i = 0 \\ \mathbf{p_s}(A = N - d) \cdot \mathbf{prob}(A = N | A = N - d) & \text{if } i = N \\ \mathbf{p_s}(A = i - d) \cdot \mathbf{prob}(A = i | A = i - d) & \\ +\mathbf{p_s}(A = i + d) \cdot \mathbf{prob}(A = i | A = i + d) & \text{if } 0 < i < N \text{ and } \frac{i}{d} \in \mathbb{N}_0 \\ 0 & \text{otherwise} \end{cases}$$

To begin with, look at the first equation for $i = 0$. Evaluating the right hand side using the expression for $\mathbf{p_s}$ with $i = d$ and the transition probability from $i = d$ to $i = 0$ and for the left hand side $\mathbf{p_s}$ with $i = 0$, yields equality between the two sides:

$$\begin{aligned}\frac{w^{N_0-1}\cdot N_0}{N_0\cdot(1+w)^{N_0-1}} &= \frac{w}{N_0+w-1}\cdot\frac{w^{N_0-2}\cdot(N_0+w-1)}{N_0\cdot(1+w)^{N_0-1}}\cdot N_0\\ \Leftrightarrow 1 &= 1\end{aligned}$$

Similarly for $i = N$, one obtains:

$$\begin{aligned}\frac{w^{-1}\cdot(N_0+N_0\cdot(w-1))}{N_0\cdot(1+w)^{N_0-1}} &= \frac{1}{w\cdot(N_0-1)+1}\cdot\frac{(N_0+(N_0-1)\cdot(w-1))}{N_0\cdot(1+w)^{N_0-1}}\cdot N_0\\ \Leftrightarrow \frac{1}{N_0\cdot(1+w)^{N_0-1}} &= \frac{1}{N_0\cdot(1+w)^{N_0-1}}\end{aligned}$$

In case $0 < i \cdot d < N$ with $i \in \mathbb{N}_0$, the equation are of the form (already multiplied by $(1+w)^{N_0-1}$):

$$\begin{aligned}&\frac{1}{2}\frac{w^{N_0-i-1}\cdot(N_0+i\cdot(w-1))}{N_0}\cdot\binom{N_0}{i}\\ &= \frac{w\cdot(i+1)}{w\cdot(i+1)+N-i-1}\cdot\frac{1}{2}\frac{w^{N_0-i-2}\cdot(N_0+(i+1)\cdot(w-1))}{N_0}\cdot\binom{N_0}{i+1}\\ &+\frac{N-i+1}{w\cdot(i-1)+N-i+1}\cdot\frac{1}{2}\frac{w^{N_0-i}\cdot(N_0+(i-1)\cdot(w-1))}{N_0}\cdot\binom{N_0}{i-1}\end{aligned}$$

The right hand side can be transformed into the left hand side:

$$\begin{aligned}&\frac{w\cdot(i+1)}{w\cdot(i+1)+N_0-i-1}\cdot\frac{1}{2}\frac{w^{N_0-i-2}\cdot(N_0+(i+1)\cdot(w-1))}{N_0}\cdot\binom{N_0}{i+1}\\ &+\frac{N_0-i+1}{w\cdot(i-1)+N_0-i+1}\cdot\frac{1}{2}\frac{w^{N_0-i}\cdot(N_0+(i-1)\cdot(w-1))}{N_0}\cdot\binom{N_0}{i-1}\\ &=\frac{1}{2}\frac{w^{N_0-i-1}}{N_0}\cdot\Bigg(\frac{(i+1)\cdot(N_0+(i+1)\cdot(w-1))}{w\cdot(i+1)+N_0-i-1}\cdot\binom{N_0}{i+1}\\ &+\frac{(N-i+1)\cdot(w\cdot(N_0+(i-1)\cdot(w-1)))}{w\cdot(i-1)+N_0-i+1}\cdot\binom{N_0}{i-1}\Bigg)\\ &=\frac{1}{2}\frac{w^{N_0-i-1}}{N_0}\cdot\binom{N_0}{i}\cdot\Bigg(\frac{N_0+(i+1)\cdot(w-1)}{w\cdot(i+1)+N_0-i-1}\cdot(N_0-i)\\ &+\frac{w\cdot i\cdot(N_0+(i-1)\cdot(w-1))}{(w\cdot(i-1)+N_0-i+1)}\Bigg)\\ &=\frac{1}{2}\frac{w^{N_0-i-1}}{N_0}\cdot\binom{N_0}{i}\cdot\Big((N_0-i)+w\cdot i\Big)\\ &=\frac{1}{2}\frac{w^{N_0-i-1}\cdot(N_0+i\cdot(w-1))}{N_0}\cdot\binom{N_0}{i}\end{aligned}$$

Next it will be shown that $\mathbf{p_s}$ is a distribution. The first condition (3.2.3) that all probabilities are at least 0 follows directly from the definition of $\mathbf{p_s}$. It remains to verify that the sum over all probabilities equals 1 (3.2.2):

$$\begin{aligned}
\sum_{j=0}^{N} \mathbf{p_s}(A=j) &= \sum_{i=0}^{N_0} \frac{1}{2} \frac{w^{N_0-i-1} \cdot (N_0 + i \cdot (w-1))}{N_0 \cdot (1+w)^{N_0-1}} \cdot \binom{N_0}{i} \\
&= \frac{1}{2 \cdot N_0 \cdot (1+w)^{N_0-1}} \cdot \Bigg(\frac{N_0}{w} \cdot \underbrace{\sum_{i=0}^{N_0} w^{N_0-i} \cdot \binom{N_0}{i}}_{(1+w)^{N_0}} \\
&\quad + \frac{w-1}{w} \cdot \underbrace{\sum_{i=0}^{N_0} w^{N_0-i} \cdot i \cdot \binom{N_0}{i}}_{(1+w)^{N_0-1} \cdot N_0} \Bigg) \\
&= \frac{1}{2 \cdot w \cdot (1+w)^{N_0-1}} \cdot \left((1+w)^{N_0} + (w-1) \cdot (1+w)^{N_0-1} \right) \\
&= \frac{(w+1)+(w-1)}{2 \cdot w} = 1
\end{aligned}$$

(A good reference for handling binomial coefficients is Zeilberger [4].)

Thus the last step, is to show that 3.2.4 holds. Intuitively this holds since the sum of the steady state distribution for odd and even states equals $\frac{1}{2}$ for each case (the calculations are omitted). Thus normalization gives the factor 2 occuring in 3.2.4.

More formally, this can be derived in the following way: The Markov chain is irreducible and positive recurrent, since the state space is finite and all states communicate. Using Theorem VIII (p.107, Berger [2]), with $r(x, y) = (x - y) \bmod (2 \cdot d)$ justifies (3.2.4) and concludes the last step of the proof of theorem (3.2.1).

□

Error for steady state

First the steady state expectation is obtained (lemma 3.2.2) and then the steady state error is calculated (see 3.2.3).

Lemma 3.2.2. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.4, let $N = N_0 \cdot d$ with $N_0 \in \mathbb{N}$, then the expectation of the process $\mathbb{E}[A_n]$ for large n ($n \to \infty$) is bounded by*

$$\left| \lim_{n\to\infty} \mathbb{E}[A_n] - \frac{d \cdot (2 \cdot N_0 + w - 1)}{2 \cdot (w+1)} \right| \leq \frac{d}{2}$$

Proof. Since the expectation might oscillate between two values, first the arithmetic mean of the oscillating expectation is calculated :

$$\begin{aligned}
\lim_{k\to\infty}\frac{\mathbb{E}[A_k]+\mathbb{E}[A_{k+1}]}{2} &= \sum_{i=0}^{N}\mathbf{p_s}(A=i)\cdot i \\
&= \sum_{i=0}^{N_0}\frac{1}{2}\frac{w^{N_0-i-1}\cdot(N_0+i\cdot(w-1))}{N_0\cdot(1+w)^{N_0-1}}\cdot\binom{N_0}{i}\cdot i\cdot d \\
&= \frac{1}{2\cdot N_0\cdot(1+w)^{N_0-1}}\cdot\Bigg(\frac{N_0}{w}\cdot\underbrace{\sum_{i=0}^{N_0}w^{N_0-i}\cdot i\cdot\binom{N_0}{i}}_{(1+w)^{N_0-1}\cdot N_0} \\
&\quad +\frac{w-1}{w}\cdot\underbrace{\sum_{i=0}^{N_0}w^{N_0-i}\cdot i^2\cdot\binom{N_0}{i}}_{(1+w)^{N_0-2}\cdot N_0\cdot(N_0+w))}\Bigg) \\
&= \frac{d}{2\cdot w}\cdot\left(N_0+\frac{w-1}{w+1}\cdot(N_0+w)\right) \\
&= \frac{d\cdot(2\cdot N_0+w-1)}{2\cdot(w+1)} \qquad (3.2.6)
\end{aligned}$$

The limit of the maximum difference between the oscillating expectations $\lim_{k\to\infty}|\mathbb{E}[A_k]-\mathbb{E}[A_{k+1}]|$ is less or equal to d, which will be shown next. Assume N_0 odd:

$$\begin{aligned}
&\lim_{k\to\infty}\Big|\mathbb{E}[A_k]-\mathbb{E}[A_{k+1}]\Big| \\
&= \lim_{k\to\infty}\Bigg|\sum_{i=0}^{\lfloor\frac{N_0}{2}\rfloor}(\mathbf{prob}(A_k=2\cdot i)\cdot 2\cdot i-\mathbf{prob}(A_{k+1}=2\cdot i+1)\cdot(2\cdot i+1))\Bigg| \\
&= \lim_{k\to\infty}\Bigg|\sum_{i=0}^{\lfloor\frac{N_0}{2}\rfloor}\Big(\mathbf{prob}(A_k=2\cdot i)\cdot 2\cdot i-\mathbf{prob}(A_k=2\cdot i)\cdot\underbrace{\mathbb{E}[A_{k+1}|A_k=2\cdot i]}_{f(2\cdot i)}\Big)\Bigg| \\
&= \lim_{k\to\infty}\Bigg(\sum_{i=0}^{\lfloor\frac{N_0}{2}\rfloor}(\mathbf{prob}(A_k=2\cdot i)\cdot\underbrace{|2\cdot i-f(2\cdot i)|}_{\leq d})\Bigg)\leq d \qquad (3.2.7)
\end{aligned}$$

where $|2\cdot i-f(2\cdot i)|\leq d$, since in expectation at most d balls are taken or given to B in one step.

For N_0 even, the derivation is almost the same, only the sum needs to be changed slightly:

$$\lim_{k\to\infty}\Bigg(\sum_{i=1}^{\frac{N_0}{2}}(\mathbf{prob}(A_k=2\cdot i-1)(2\cdot i-1)-\mathbf{prob}(A_{k+1}=2\cdot i)\cdot 2\cdot i)\Bigg)$$

Obviously, for the arithmetic mean $\overline{ab}:=\frac{a+b}{2}$ of two numbers a,b with $|a-b|=c$ holds that:

$$|a-\overline{ab}|=\frac{|a-b|}{2}\leq\frac{c}{2} \qquad (3.2.8)$$

and also

$$|b - \overline{ab}| = \frac{|b-a|}{2} \leq \frac{c}{2} \tag{3.2.9}$$

Using the previous inequalities ((3.2.8) and (3.2.9)) together with the bound on the mean (3.2.6) for the oscillating expectations and the maximum difference between the expectation between two steps (3.2.7) completes the proof of lemma 3.2.2. □

Next the steady state error, e.g. the difference between the real and pseudo expectation after "infinitely" many steps, is calculated. It turns out to be independent of the number of balls N and the weight w, which is rather surprising. More precisely, the error is less than d, e.g. 1, if only one ball is moved per step.

Corollary 3.2.3. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.4, let $N = N_0 \cdot d$ with $N_0 \in \mathbb{N}$, then the error err_s for the steady state can be bounded by:*

$$err_s < d$$

Proof. Using lemma 3.2.2 and the expression for the fix point (2.4.3) gives

$$\begin{aligned}
err_s &\leq \max_{a\in[0,N]} \lim_{n\to\infty} err(n,a) \\
&= \max_{a\in[0,N]} \lim_{n\to\infty} \left(\mathbf{E}[A_n|A_0 = a] - f^n(a) \right) \\
&\leq \frac{d}{2} + \left|\frac{d\cdot(2\cdot N_0 + w - 1)}{2\cdot(w+1)} - \frac{N}{w+1}\right| \\
&\leq \frac{d}{2} + \left|\frac{d\cdot N_0}{w+1} + \frac{d\cdot(w-1))}{2\cdot(w+1)} - \frac{N}{w+1}\right| \\
&\leq d
\end{aligned}$$

□

3.3 Relation between state with maximum probability and fix point

It seems almost trivial that for a highly concentrated process the expected and the most likely state are close for any time step n.

Experiments have shown that for the considered example (see 2.4) the pseudo expectation and the most likely state almost coincide. But only for the steady state this could be confirmed analytically. In fact, it could be shown that the most likely state and the fix point are equal, given that the fix point is an integer e.g. an element of the finite state space.

Lemma 3.3.1. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.4 and let the fix point $b_f = \frac{N}{1+w}$ be an integer and let $\frac{2}{N} < w < \frac{N}{2}$, then the probability of the state b_f is the maximum steady state probability:*

$$arg \max_{i \in [0,N]} \mathbf{p_s}(A = i) = \frac{N}{1+w}$$

Intuitively, this holds due to lemma 3.3.2, which says that it is more likely to move towards the fix point than away from it, and the fact that the loop probabilities are 0 for all states.

Let ζ_a^b be a random variable, which denotes the number of visits of state a before the chain $A_T, A_{T+1}, \ldots$ hits b conditioned on $A_T = a$.

$$\zeta_a^b := \sum_{j=T}^{\infty} \mathbf{prob}(A_j = a \text{ and } A_{T+k} \neq b \ (\forall j > k \geq 0) \mid A_T = a)$$

Proof. The goal is to show that $\mathbf{p_s}(A = b_f) \geq \mathbf{p_s}(A = i)\ \forall i$.

Equivalently, it will be proven that:

$$\mathbb{E}[\zeta_{b_f}^i] > \mathbb{E}[\zeta_i^{b_f}] \tag{3.3.1}$$

Or in words: The expected number of visits of b_f given the chain starts from the fix point b_f before the state i is visited is at least the expected number of visits of i given the chain starts from state i before the fix point is hit.

Next we show that indeed

$$(3.3.1) \Rightarrow \mathbf{p_s}(A = b_f) \geq \mathbf{p_s}(A = i)$$

Since the transition matrix P is irreducible and has an invariant distribution (theorem 3.2.1), all states are positive recurrent (theorem 1.7.7, Norris [1]), meaning that it only takes finite time to go from any state to any other. So the hitting time of every state is finite. Therefore (3.3.1) intuitively suggests that in the long run b_f will be visited more often than i, independent of the starting state A_0.

Let the random variable T_b^a be the first passage time of b given that the chain starts from a. Assume, $A_0 = i_0$, where i_0 is an arbitrary state $\in [0, N]$. Using the strong Markov property (theorem 1.4.2, Norris [1]) for the chain $A_0, A_1, \ldots$ conditional on $A_{T_{b_f}^{i_0}} = b_f$, the process will start afresh after time $T_{b_f}^{i_0}$ from b_f independent of its past $A_0, A_1, \ldots, A_{T_{b_f}^{i_0}-1}$. Until b_f is hit for the first time (starting from i_0), at most $\mathbb{E}[\zeta_i^{b_f}]$ are made to i, since

$$\mathbb{E}[\text{Visits of } i \text{ before } b_f \text{ is hit}|A_0 = i_0] \;=\; \mathbb{E}[\zeta_i^{b_f}] \cdot \underbrace{\mathbf{prob}(i \text{ hit before } i_0|A_0 = i_0)}_{\leq 1}$$

Using the strong Markov property again for the chain $A_{n_0}, A_{n_0+1}, \ldots$ with $n_0 = T_{b_f}^{i_0}$, the process will start afresh from b_f after $n_0 + T_i^{b_f}$ steps from i and finally, after another random time $T_{b_f}^i$ the process will start again from b_f and the loop is complete. Within the first loop, starting and ending in i or b_f (depending on the starting state i_0), in expectation b_f will be hit more often, if

$$\mathbb{E}[\zeta_{b_f}^i] - \mathbb{E}[\zeta_i^{b_f}] > 0$$

That is to say (3.3.1) holds. Looking at the long term behavior, meaning at infinitely many cycles, we obtain that

$$\begin{aligned}
&\lim_{n\to\infty} (\mathbb{E}[\text{Visits of } b_f \text{ within } n \text{ steps}] \;-\; \mathbb{E}[\text{Visits of } i \text{ within } n \text{ steps}]) \\
&= \lim_{k\to\infty} k \cdot (\mathbb{E}[\zeta_{b_f}^i] - \mathbb{E}[\zeta_i^{b_f}]) \;-\; \mathbb{E}[\text{Visits of } i \text{ before } b_f \text{ is hit}|A_0 = i] \\
&\geq \lim_{k\to\infty} k \cdot (\mathbb{E}[\zeta_{b_f}^i] - \mathbb{E}[\zeta_i^{b_f}]) \;-\; \mathbb{E}[\zeta_i^{b_f}] = \infty \Leftrightarrow \;(3.3.1)
\end{aligned}$$

Using the ergodic theorem 1.10.2 together with theorem 1.7.7 (see Norris [1]), the steady state probability of a state j of the Markov chain is given by

$$\mathbf{p_s}(A = j) = \lim_{n\to\infty} \frac{\text{Visits of } j \text{ within } n \text{ steps}}{n}$$

This yields that (3.3.1) $\Rightarrow \mathbf{p_s}(A = b_f) \geq \mathbf{p_s}(A = i)$.

In order to show (3.3.1) the next lemma is useful. It says that the transition probability towards the fix point is greater than going further away from it.

Lemma 3.3.2. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.4 and let the fix point b_f be an integer and let $\frac{2}{N} < w < \frac{N}{2}$, then for the probability of the state b_f holds:*

$$\begin{aligned}
\mathbf{prob}(A_{n+1} = b_f + i + 1|A_n = b_f + i) \;&\leq\; \mathbf{prob}(A_{n+1} = b_f + i - 1|A_n = b_f + i) \\
\mathbf{prob}(A_{n+1} = b_f - i + 1|A_n = b_f - i) \;&\geq\; \mathbf{prob}(A_{n+1} = b_f - i - 1|A_n = b_f - i)
\end{aligned}$$

where $0 \leq i \leq N$ and $i = k \cdot d$ with $k \in \mathbb{N}_0$.

Proof. The proof will be done by straight forward calculation using the definition of the transition probability function for each case in the lemma.

The transition probability from the fix point (e.g. $i = 0$) is independent of the direction:

$$\mathbf{prob}(A_{n+1} = b_f \pm 1|A_n = b_f) = \frac{1}{2}$$

The likelihood of getting closer to the fix point, if the number of balls of B is $b_f + i$ ($i > 0$), is bigger than $\frac{1}{2}$.

$$\mathbf{prob}(A_{n+1} = b_f + i - 1 | A_n = b_f + i) \;=\; \frac{w \cdot N_0 + w^2 \cdot (i+1) + w \cdot (i+1)}{2 \cdot w \cdot N_0 + w^2 \cdot (i+1) - \cdot (i+1)} \;>\; \frac{1}{2}$$

Thus moving away from the fix point is $< \frac{1}{2}$.

To come closer to the fix point, if the number of balls is $b_f - i$ ($i > 0$) is greater than $\frac{1}{2}$, too.

$$\mathbf{prob}(A_{n+1} = b_f - i + 1 | A_n = b_f - i) \;=\; \frac{w \cdot N_0 + (i-1) - w \cdot (i-1)}{2 \cdot w \cdot N_0 + (i-1) - w^2 \cdot (i-1)} \;>\; \frac{1}{2}$$

Thus going further away is $< \frac{1}{2}$.

This concludes the enumeration of all cases and verifies lemma 3.3.2. □

Now (3.3.1) will be proven.

Assume $b_f > i$ for the following statements. Then

$$\mathbb{E}[\zeta_i^{b_f}] \;=\; \mathbb{E}[\text{Transitions from } b_f \text{ to } b_f - 1 \text{ until } i \text{ is hit} | A_0 = b_f] \;\cdot\; \mathbb{E}[\zeta_{b_f}^{b_f - 1}] \tag{3.3.2}$$

In this formula $\mathbb{E}[\zeta_{b_f}^{b_f-1}]$ gives the number of times, b_f is visited before the chain moves one step in the direction of i or in other words: before it does a transition from b_f to $b_f - 1$.

The other term

$$\mathbb{E}[\text{Transitions from } b_f \text{ to } b_f - 1 \text{ until } i \text{ is hit} | A_0 = b_f]$$

gives the number of times the chain moves back and forth between b_f and $b_f - 1$ plus 1 (for the last transition from b_f to $b_f - 1$) without hitting i.

So due to the strong Markov property (see theorem 1.4.2, Norris [1]) for every transition from b_f to $b_f - 1$ (at least 1 takes place before hitting i) in expectation b_f is visited $\mathbb{E}[\zeta_{b_f}^{b_f-1}]$ times.

Substituting (3.3.2) and the analogue for $\mathbb{E}[\zeta_{b_f}^{i}]$ into (3.3.1) gives

$$\begin{aligned}
&\mathbb{E}[\zeta_{b_f}^{i}] \;>\; \mathbb{E}[\zeta_i^{b_f}] \\
&\Leftrightarrow \\
&\mathbb{E}[\text{Transitions from } b_f \text{ to } b_f - 1 \text{ until } i \text{ is hit} | A_0 = b_f] \;\cdot\; \mathbb{E}[\zeta_{b_f}^{b_f-1}] \\
&> \mathbb{E}[\text{Transitions from } i \text{ to } i+1 \text{ until } b_f \text{ is hit} | A_0 = i] \;\cdot\; \mathbb{E}[\zeta_i^{i+1}]
\end{aligned} \tag{3.3.3}$$

Comparing the corresponding terms is the last part of the proof.

Let us consider $\mathbb{E}[\zeta_{b_f}^{b_f-1}]$ first. With $\mathbf{prob}(A_{n+1} = b_f + 1 | A_n = b_f)$ the process goes to $b_f + 1$. (Observe that there always exist states $b_f \pm 1$ because $b_f = \frac{N}{1+w}$ must be an integer and $0 < w < \infty$). Since the Markov chain is positive recurrent, the process will certainly return to b_f in finite time before i is hit (Recall that $b_f > i$ is assumed).

The number of visits to other states than b_f is not of any interest and also the number of steps it takes to do them. Accordingly, it is appropriate to consider a Markov chain C_n instead, which transition matrix is defined by the graph of the chain A_n with all states greater than b_f removed:

$$\mathbf{prob}(C_{n+1} = b_f + 1 | C_n = b_f) = 0$$

and the loop probability of b_f equal to

$$\mathbf{prob}(C_{n+1} = b_f | C_n = b_f) \ = \ \mathbf{prob}(A_{n+1} = b_f + 1 | A_n = b_f)$$

For C_n the expectation of $\zeta_{b_f}^{b_f - 1}$ is

$$\mathbb{E}[\zeta_{b_f}^{b_f-1}] \ = \ \sum_{n=0}^{\infty} \mathbf{prob}(C_{n+1} = b_f | C_n = b_f)^n \ = \ \sum_{n=0}^{\infty} \mathbf{prob}(A_{n+1} = b_f + 1 | A_n = b_f)^n$$

which is the same as the expectation of $\zeta_{b_f}^{b_f-1}$ for A_n.

Analogously, $\mathbb{E}[\zeta_i^{i+1}] \ = \ \sum_{n=0}^{\infty} \mathbf{prob}(A_{n-1} = i - 1 | A_n = i)^n$.

Using lemma (3.3.2) gives that

$$\mathbf{prob}(C_{n+1} = i - 1 | C_n = i) \ < \ \mathbf{prob}(C_{n+1} = b_f + 1 | C_n = b_f)$$

and therefore

$$\mathbb{E}[\zeta_{b_f}^{b_f-1}] > \mathbb{E}[\zeta_i^{i+1}] \tag{3.3.4}$$

Before comparing the second term of (3.3.3), the following definition and lemma is required: The probability to return to a state a within two steps by visiting $a + 1$ is

$$\begin{aligned} \mathbf{prob}(\text{cycle hit } a + 1) \ &:= \ \mathbf{prob}(C_{n+2} = a \text{ and } C_{n+1} = a + 1 | C_n = a) \\ &= \ \mathbf{prob}(C_{n+1} = a + 1 | C_n = a) \cdot \mathbf{prob}(C_{n+2} = a | C_{n+1} = a + 1) \end{aligned}$$

Lemma 3.3.3. *For $a \in [b_f, N - 1]$*

$$\mathbf{prob}(\textit{cycle hit } a) > \mathbf{prob}(\textit{cycle hit } a + 1)$$

and for $a \in [0, b_f - 1]$

$$\mathbf{prob}(\textit{cycle hit } a) < \mathbf{prob}(\textit{cycle hit } a + 1)$$

Proof. The case $a \in [b_f, N - 1]$ is investigated first. By elementary calculation for $a = b_f$ it follows for $w \in]\frac{2}{N}, \frac{N}{2}[$ that

$$\begin{aligned} &\mathbf{prob}(\text{cycle hit } b_f) \ - \ \mathbf{prob}(\text{cycle hit } b_f + 1) = \frac{1}{2} \cdot \frac{w \cdot (w^3 + 3\,w^2 + 3\,w + 1)}{(w^2 + w \cdot N - 1) \cdot (w^2 + 2\,w \cdot N - 1)} > 0 \\ &\Rightarrow \mathbf{prob}(\text{cycle hit } b_f) > \mathbf{prob}(\text{cycle hit } b_f + 1) \end{aligned}$$

Due to this and the fact that **prob**(cycle hit a) is a continuous function for $a \in [b_f, N-1]$, it follows that if

$$\exists a \text{ with } \quad \textbf{prob}(\text{cycle hit } a) \; < \; \textbf{prob}(\text{cycle hit } a+1)$$

then there must also

$$\exists a \text{ with } \quad \textbf{prob}(\text{cycle hit } a) \; = \; \textbf{prob}(\text{cycle hit } a+1)$$

Solving the equation for a gives:

$$\begin{aligned} & \frac{w \cdot (a+1)(N-a)}{((w-1) \cdot a + N) \cdot (N + (a+1) \cdot w - (a+1))} \\ = & \frac{w \cdot (a+2)(N-(a+1))}{((w-1) \cdot (a+1) + N) \cdot (N + (a+2) \cdot w - (a+2))} \\ \Leftrightarrow & \, a = \frac{N \cdot w - 2 \cdot w}{1+w} < b_f \Rightarrow \; a \notin [b_f, N-1] \end{aligned}$$

For that reason there exists no such a and the claim follows. For $a \in [0, b_f - 1]$ the proof is analogous. □

The next statement is needed for the final step of the proof, where

$$\mathbb{E}[\text{Transitions from } b_f \text{ to } b_f - 1 \text{ before } i \text{ or } b_f + 1 \text{ is hit} | A_0 = b_f]$$

is expressed in terms of cycle probabilities.

Lemma 3.3.4. *For numbers c_j ($j \in [0, D]$) with $c_j > c_{j+1}$ holds that*
$c_0(1 + c_1(1 + c_2(1 + ..(1 + c_D)))) > c_D(1 + c_{D-1}(1 + c_{D-2}(1 + ..(1 + c_0))))$

Proof. The lemma follows from a simple expansion of both sides and comparing each term:

$$\begin{aligned} c_0(1 + c_1(1 + c_2(1 + ..(1 + c_D)))) \quad &> \quad c_D(1 + c_D(1 + c_{D-1}(1 + ..(1 + c_0)))) \\ \Leftrightarrow c_0 + c_0c_1 + c_0c_1c_2 + \ldots + \prod_{j=0}^{D} c_j \quad &> \quad c_D + c_Dc_{D-1} + c_Dc_{D-1}c_{D-2} + \ldots + \prod_{j=0}^{D} c_j \end{aligned}$$

□

Now the second terms of (3.3.3) are compared. More precisely:

$$\begin{aligned} & \mathbb{E}[\text{Transitions from } b_f \text{ to } b_f - 1 \text{ until } i \text{ is hit} | A_0 = b_f] \\ & \geq \; \mathbb{E}[\text{Transitions from } i \text{ to } i+1 \text{ until } b_f \text{ is hit} | A_0 = i] \\ & \Leftrightarrow \\ & \sum_{n_1+n_2+\ldots+n_{b_f-i-1} = b_f - i - 1, \; n_j \geq 1}^{\infty} (1 + \textbf{prob}(\text{cycle hit } b_f)^{n_1}) \; \cdot \; (1 + \textbf{prob}(\text{cycle hit } b_f - 1)^{n_2}) \; \cdot \\ & (1 + \ldots(1 + \textbf{prob}(\text{cycle hit } i+2)^{n_{b_f-i-1}}))) \\ & \geq \sum_{n_1+n_2+\ldots+n_{b_f-i-1} = b_f - i - 1, \; n_j \geq 1}^{\infty} (1 + \textbf{prob}(\text{cycle hit } i+1)^{n_1}) \; \cdot \; (1 + \textbf{prob}(\text{cycle hit } i+2)^{n_2}) \; \cdot \\ & (1 + \ldots(1 + \textbf{prob}(\text{cycle hit } b_f - 1)^{n_{b_f-i-1}}))) \end{aligned}$$

This follows by definition of the expectation. (One might also find the formula with the expanded product given in the proof of lemma 3.3.4 more instructive.)

Observe that both sums have the same number of summands and in both the same cycle probabilities occur, but in the opposite order (apart from the first term **prob**(cycle hit $i+1$) and **prob**(cycle hit b_f)). Once again, the individual terms are examined. Due to lemma 3.3.3

$$\mathbf{prob}(\text{cycle hit } i+1) \;\leq\; \mathbf{prob}(\text{cycle hit } b_f)$$

Applying lemma 3.3.4 for all common summands $D = b_f - i - 1$, $c_0 = c_D = 1$, $c_j = \mathbf{prob}(\text{cycle hit } j + i + 1)$ shows that

$$\begin{aligned}&(1+\mathbf{prob}(\text{cycle hit } b_f-1)^{n_2}) \;\cdot\; (1+...(1+\mathbf{prob}(\text{cycle hit } i+2)^{n_{b_f-i-1}})))\\ > &(1+\mathbf{prob}(\text{cycle hit } i+2)^{n_2}) \;\cdot\; (1+...(1+\mathbf{prob}(\text{cycle hit } b_f-1)^{n_{b_f-i-1}}))\end{aligned}$$

The case for $i < b_f$ is analogous.

Finally, this shows (3.3.3) and (3.3.1) which concludes the proof of lemma 3.3.1, that the probability is maximum at the fix point. □

Chapter 4

Recurrence Formula For Error Of Pseudo Expectation

In the current chapter a recurrence formula for the error is stated for the pseudo expectation depending on one and two variables (section 4.1 and 4.3). Unfortunately, it depends on the second moment e.g. on the variance for the pseudo expectation in one dimension. This is a serious drawback and might make it hard to use. An example application is only given in one dimension (see section 4.2).

4.1 1 Dimension

A recurrence formula for the error is given, which can be directly applied for almost any kind of f. As stated before, since the error bound depends on the variance at step n, which is usually not known, but might be bounded, the application is in general not straight forward.

Theorem 4.1.1. *Let $S_0, S_1, \ldots, S_n$ be a Markov process as defined in 2.2.1 and let its state space S be a subset of $\mathbb{R}$, where $\mathbf{prob}(S_0 = k) = 1$ for an arbitrary $k \in S$. Let the function f for the pseudo expectation (2.3.1) be such that it and its derivatives up to order 2 are continuous on I_S, then*

$$err(n+1, S_0) = f'(f^n(S_0)) \cdot err(n, S_0) + \frac{f''(\bar{c}_n)}{2} \cdot (var(S_n) + err(n, S_0)^2)$$

where $\bar{c}_n \in I_S$ and $f^n(S_0)$ denotes the iterated application of f.

Remark

1. If the function f is linear, then $f'' = 0$ and the approximation is exact.

2. In case of a deterministic process ($\forall n \; var(S_n) = 0$), there is also no error.

Proof. As a first step $\mathbb{E}[S_n]$ is expressed in terms of f and the probability distribution of the random variable S_n.

$$
\begin{aligned}
\mathbb{E}[S_n] &= \sum_{i\in S} \mathbf{prob}(S_n = i)\cdot i \\
&= \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot \mathbb{E}[S_n|S_{n-1}=i] \\
&= \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot f(i)
\end{aligned}
$$

Next we use the Taylor expansion of the function f at point $f^{n-1}(S_0)$ up to order 1. In order to do this, f must be continuous on I_S as well as its first derivative and the 2nd derivative has to exist on I_S. The Taylor expansion becomes:

$$
f(x) = \underbrace{f(f^{n-1}(S_0))}_{f^n(S_0)} + f'(f^{n-1}(S_0))\cdot(x - f^{n-1}(S_0)) + \frac{f''(c^x_{n-1})}{2}\cdot(x - f^{n-1}(S_0))^2 \qquad (4.1.1)
$$

where $x \in I_S$ and c^x_{n-1} denotes a constant in the interval $[S_0,x]$.

Using the previous Taylor expansion (4.1.1) and the obvious fact that all probabilities of all states add up to 1 ($\sum_{i\in S}\mathbf{prob}(S_{n-1}=i)=1$), the expectation of S_n becomes:

$$
\begin{aligned}
\mathbb{E}[S_n] &= \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot f(i) \\
&= \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot\Big(f^n(S_0) + f'(f^{n-1}(S_0))\cdot(i - f^{n-1}(S_0)) + \\
&\quad \frac{f''(c^i_{n-1})}{2}\cdot(i - f^{n-1}(S_0))^2\Big) \\
&= \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot f^n(S_0) + \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot f'(f^{n-1}(S_0))\cdot(i - f^{n-1}(S_0)) \\
&\quad + \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot\frac{f''(c^i_{n-1})}{2}\cdot(i - f^{n-1}(S_0))^2 \\
&= f^n(S_0)\cdot\sum_{i\in S} \mathbf{prob}(S_{n-1} = i) \\
&\quad + f'(f^{n-1}(S_0))\cdot\Big(\sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot i - \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot f^{n-1}(S_0)\Big) \\
&\quad + \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot\frac{f''(c^i_{n-1})}{2}\cdot(i - f^{n-1}(S_0))^2 \\
&= f^n(S_0) + f'(f^{n-1}(S_0))\cdot\Big(\mathbb{E}[S_{n-1}] - f^{n-1}(S_0)\Big) \\
&\quad + \sum_{i\in S} \mathbf{prob}(S_{n-1} = i)\cdot\frac{f''(c^i_{n-1})}{2}\cdot(i - f^{n-1}(S_0))^2 \qquad (4.1.2)
\end{aligned}
$$

The analysis of the last term $\sum_{i\in S}\mathbf{prob}(S_{n-1}=i)\cdot\frac{f''(c^i_{n-1})}{2}\cdot(i-f^{n-1}(S_0))^2$ requires some prerequisites.

Lemma 4.1.2. *Let $d(i)$ and $n(j)$ be two sequences of non-negative numbers and $\forall i \in [1, d_0]$ let $d(i) \in [a, b]$, then there exists $\overline{d} \in [a, b]$ such that*

$$\sum_{i=1}^{d_0} d(i) \cdot n(i) = \overline{d} \cdot \sum_{i=1}^{d_0} n(i)$$

.

Proof. In case $\sum_{i=1}^{d_0} n(i) = 0$, $\overline{d}$ can be any number in $[a, b]$. If not, then one can define the function p as $p : i \mapsto \frac{n(i)}{\sum_{j=1}^{d_0} n(j)}$. Obviously $p(i) \geq 0$ and $\sum_{i=1}^{d_0} p(i) = 1$. So p is a probability distribution and $\overline{d} = \sum_{i=1}^{d_0} d(i) \cdot p(i)$ is simply the expectation, which exists and is always within $[a, b]$. □

Since f'' is continuous on I_S, there exist for all $f''_0 \in [\min_j f''(c_{n-1}^j), \max_j f''(c_{n-1}^j)]$ an $x \in [\min_j c_{n-1}^j, \max_j c_{n-1}^j]$, such that $f''(x) = f''_0$. Therefore Lemma 4.1.2 can be applied with $d(i) = f''(c_{n-1}^i)$ and $n(i) = \frac{1}{2} \cdot \mathbf{prob}(S_{n-1} = i) \cdot (i - f^{n-1}(S_0))^2$ which yields that there exists a $\overline{c}_{n-1}$ such that

$$\begin{aligned}
&\textstyle\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot \frac{f''(c_{n-1}^i)}{2} \cdot (i - f^{n-1}(S_0))^2 \\
&= \tfrac{f''(\overline{c}_{n-1})}{2} \cdot \textstyle\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot (i - f^{n-1}(S_0))^2 \\
&= \tfrac{f''(\overline{c}_{n-1})}{2} \cdot \textstyle\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot \Big(i - \mathbb{E}[S_{n-1}] + err(n-1, S_0)\Big)^2 \\
&= \tfrac{f''(\overline{c}_{n-1})}{2} \cdot \Bigg[\underbrace{\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot \Big(i - \mathbb{E}[S_{n-1}]\Big)}_{\mathbb{E}[(S_{n-1} - \mathbb{E}[S_{n-1}])^2] = var(S_{n-1})} \\
&\quad - \underbrace{\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot 2 \cdot err(n-1, S_0) \cdot \mathbb{E}[S_{n-1}])}_{2 \cdot err(n-1, S_0) \cdot \mathbb{E}[S_{n-1}])} + \underbrace{\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot 2 \cdot i \cdot err(n-1, S_0)}_{2 \cdot err(n-1, S_0) \cdot \mathbb{E}[S_{n-1}])} \\
&\quad + \underbrace{\sum_{i \in S} \mathbf{prob}(S_{n-1} = i) \cdot err(n-1, S_0)^2}_{err(n-1, S_0)^2} \Bigg] \\
&= \tfrac{f''(\overline{c}_{n-1})}{2} \cdot \Big(var(S_{n-1}) + err(n-1, S_0)^2\Big) \qquad (4.1.3)
\end{aligned}$$

Combining (4.1.3) and (4.1.2) gives:

$$\mathbb{E}[S_n] \quad = \quad f^n(S_0) + f'(f^{n-1}(S_0)) \cdot err(n-1, S_0) + \frac{f''(\overline{c}_{n-1})}{2} \cdot \Big(var(S_{n-1}) + err(n-1, S_0)^2\Big)$$

which completes the proof. □

4.2 Example

In this section the previous recurrence formula for the error 4.1.1 is applied to the example process (defined in 2.4).

Since the terms in the recurrence formula depend on n, their maximum over all steps is used. Thus the coefficients become constant and the fix point of the recurrence equation can be calculated. Unfortunately, for the variance, a maximum could only be conjectured. Moreover, since the maximum is taken, the modified recurrence formula only converges for small w, say $N < \frac{16}{w^6}$ (See remark (4.2.1)).

The missing element to turn the conjecture into a theorem is to show that $var(S_n) \leq \frac{N}{4}$, where $\frac{N}{4}$ is the variance of a random variable X, which is distributed $binomial(n, p)$ with $n = N$ and $p = \frac{1}{2}$. Intuitively, this is reasonable since for $w = 1$ the bound is exact (e.g. $var(n)$ grows with n up to $\frac{N}{4}$. It was previously shown (see remark after 3.2.1), that the steady state distribution becomes the binomial distribution for $w = 1$).

The example process becomes more concentrated for bigger w and it is obvious that the variance becomes smaller. But a proof is still outstanding. For any value of w bigger than 1, experiments have shown, that the variance has a maximum, which is less than the conjectured bound of $\frac{N}{4}$ (see figures 2.3 and 2.8).

Since the variance gets smaller the more concentrated the process is, e.g. with decreasing w, $\frac{N}{4}$ is a bad bound for small w as can be seen be looking at the rough bounds stated in the remark section (e.g. expression (4.2.10) and (4.2.10)).

Conjecture 4.2.1. *Given the stochastic process defined in 2.4. Let $N > 2$ and $1 < w < \frac{N}{4}$ and let*

$$err_m := \frac{1 - f'_m - \sqrt{(f'_m - 1)^2 - 4\, {f''_m}^2 var_m}}{2 f''_m}$$

then if err_m is a real number and

$$0 \leq f'_m + f''_m \cdot x < 1 \qquad x \in [0, err_m] \tag{4.2.1}$$

it follows that

$$\max_n err(n, A_0) \leq err_m$$

where

$$f'_m := \max\{f'(A_0), f'(b_f)\}$$

$$f''_m := \max_{x \in \{A_0, b_f\}} \frac{4 \cdot (w-1)}{((w-1) \cdot x + N)^2}|$$

$$var_m := \frac{N}{4}$$

To begin with, it will be shown that f'_m and f''_m are indeed the maximum coefficients.

Let's consider f'_m. The next statement gives an interval for the possible values for the first derivative:

$$\max_{x\in[c,d]} f'(x) \in \{f'(c), f'(d)\} \tag{4.2.2}$$

For the non-linear case ($w \neq 1$) the second derivative is always non-zero in the interval [0,N]:

$$f''(x) = \frac{4 \cdot w \cdot (w-1) \cdot N}{(N + (w-1) \cdot x)^3}$$

If $1 < w < \frac{N}{4}$, $f'' > 0$. This implies that f' is strictly monotone increasing. Because of this property of f'', f' has neither its maximum nor minimum inside the interval, but instead on the boundary, which shows (4.2.2).

Additionally, the first derivative is always within $]0, \frac{1}{2}[$, since $\forall N > 0,\ 1 < w < \frac{N}{4}$

$$\frac{1}{2} < f'(x) = 1 - \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot x)^2} < 1 \tag{4.2.3}$$

Since $0 < f' < 1$ the limit $\lim_{n\to\infty} f^n(A_0)$ converges to a fix point b_f and the convergence is monotone, meaning that for a start point A_0 less than the fix point b_f holds $\forall n \geq 0$:

$$f^{n+1}(A_0) > f^n(A_0)$$

and for $A_0 > b_f$

$$f^{n+1}(A_0) < f^n(A_0)$$

It also follows that the iteration $f^n(A_0)$ always stays within the starting point and the fix point b_f:

$$\max_n f^n(A_0) \in \{A_0, b_f\} \tag{4.2.4}$$

Due to the claim (4.2.2), that f' takes its maximum on the boundary and the monotone convergence of f^n, the coefficient for $err(n, A_0)$ is bounded by

$$\max_n f'(f^n(A_0)) = \max\{f'(A_0), f'(b_f)\} =: f'_m$$

Now a bound for f''_m will be calculated. Recall that $f''(\overline{c}_n)$ is an expectation

$$f''(\overline{c}_n) = \sum_{i=0}^{N} f''(c_n^i) \cdot p_i$$

where p_i ($0 \leq i \leq N$) is a distribution. For that reason $f''(\overline{c}_n) < \max_{i\in[0,N],n\in[0,\infty]} f''(c_n^i)$. Solving the Taylor expansion (4.1.1) of $f(i)$ given by

$$f(i) = f(a) + f'(a) \cdot (i-a) + \frac{f''(c_n^i)}{2} \cdot (i-a)^2$$

where $a = f^{n-1}(A_0)$ for the second derivative gives

$$\begin{aligned} f''(c_n^i) &= 2 \cdot \frac{f(i) - f(a) + f'(a)(a-i)}{(i-a)^2} \\ &= \frac{4 \cdot (w-1)\, w \cdot N}{((w-1) \cdot i + N)\, ((w-1) \cdot a + N)^2} \end{aligned}$$

For $i = 0$ the previous expression becomes maximum:

$$\max_i f''(c_n^i) = \frac{4 \cdot w \cdot (w-1)}{((w-1) \cdot a + N)^2}$$

The second derivative $f''(c_n^i)$ also decreases with a. The maximum for a is $\max_n f^n(A_0) = \max_{a \in \{A_0, b_f\}} a$ (see (4.2.4)). For that reason

$$\max_{i,a} f''(c_n^i) = \max_{a \in \{A_0, b_f\}} \frac{4 \cdot (w-1)}{((w-1) \cdot a + N)^2}$$

This proofs the claim for f''_m.

Taking a closer look at each of the terms of the bound given by theorem 4.1.1 reveals that: For $w > 1$,

$$err(n+1, A_0) = \underbrace{f'(f^n(A_0))}_{\in [f'(A_0), f'(b_f)]} \cdot err(n, A_0) + \underbrace{\underbrace{\frac{f''(\bar{c}_n)}{2}}_{>0} \cdot (\underbrace{var(A_n)}_{>0} + \underbrace{err(n, A_0)^2}_{\geq 0})}_{>0} \tag{4.2.5}$$

The error bound is defined as $errb(1) = errb(0) = 0$ and

$$errb(n+1) := f'_m \cdot errb(n) + \frac{f''_m}{2} \cdot (var_m + errb(n)^2) \tag{4.2.6}$$

Since the coefficients in (4.2.5) have no sign change and the signs for the terms containing $err(n, A_0)$ are equal, the maximum coefficients can be taken to bound the error: $errb(n) \geq err(n, A_0)$.

The iteration $errb(n)$ might not converge. In that case the fix point equation $errb(n+1) = errb(n)$ has two solutions with an imaginary part, but if it does converge the fix point equation has the real solution(s):

$$errb_{f\pm} = \frac{1 - f'_m \pm \sqrt{(f'_m - 1)^2 - 4\, {f''_m}^2 var_m}}{2 \cdot f''_m}$$

Both solutions are feasible in the sense that $errb_{f\pm} > 0$, since $f''_m > 0$ and

$$0 > \underbrace{1 - f'_m}_{\in]0,1[, \text{ see } (4.2.3)} \pm \sqrt{\underbrace{(f'_m - 1)^2 - 4\, {f''_m}^2 var_m}_{<(1-f'_m)^2}}$$

Finally, $errb(n)$ converges towards $errb_{f-}$, since the first derivative of $errb(n+1)$ (4.2.6) has to be within $]0,1[$ in the interval $[0, err_{b_f-}]$, e.g. condition (4.2.1) has to be satisfied.

Remark 4.2.1.

1. To get a better understanding, when the iteration converges in terms of N and w, the two extreme cases $A_0 = 0$ and $A_0 = N$ are considered. By looking at the definition of f'_m and f''_m it is not hard to see that the error of one of the two cases gives the maximum possible error for any starting state $A_0 \in [0, N]$. Furthermore, if the bound exists (e.g. err_m in conjecture 4.2.1 is a real number and 4.2.1 is satisfied) for $A_0 = 0$ and $A_0 = N$, it exists for any $A_0 \in [0, N]$, too.

   The bound for $A_0 = N$ becomes a function of w and N. In this case $1 - f'_m = 1 - f'(N) = \frac{2}{w \cdot N}$ and for b_f the second derivative $f''_m = \frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^2 \cdot N^2}$ is maximized. So the fix point equals:

$$\begin{aligned} errb_{f-} &= \frac{\frac{2}{w \cdot N} - \sqrt{(\frac{2}{w \cdot N})^2 - 4 \cdot (\frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^2 \cdot N^2})^2 \cdot \frac{N}{4}}}{2 \cdot \frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^2 \cdot N^2}} \\ &= \frac{4 \cdot w \cdot N - \sqrt{16 \cdot w^2 \cdot N^2 - (w+1)^6 \cdot (w-1)^2 \cdot N}}{2 \cdot (w+1)^3 \cdot (w-1)} \end{aligned}$$

   The iteration converges if the root is a real number:

$$\begin{aligned} 0 &< 16 \cdot w^2 \cdot N^2 - (w+1)^6 \cdot (w-1)^2 \cdot N \\ \Leftarrow w &< \sqrt[6]{N} - 1 \end{aligned}$$

   Similarly for $A_0 = 0$ the bound is maximized for $1 - f'_m = 1 - f'(b_f) = \frac{(w+1)^2}{2 \cdot w \cdot N}$ and $f''_m = \frac{4 \cdot w \cdot (w-1)}{N^2}$ is maximized for 0. So the bound for the error becomes:

$$\begin{aligned} errb_{f-} &= \frac{\frac{(w+1)^2}{2 \cdot w \cdot N} - \sqrt{(\frac{(w+1)^2}{2 \cdot w \cdot N})^2 - 4 \cdot (\frac{4 \cdot w \cdot (w-1)}{N^2})^2 \cdot \frac{N}{4}}}{2 \cdot \frac{4 \cdot w \cdot (w-1)}{N^2}} \\ &= \frac{2 \cdot (w+1)^2 \cdot N - \sqrt{(w+1)^4 - \frac{64 \cdot w^4 \cdot (w-1)^2}{N}}}{16 \cdot w^2 \cdot (w-1)} \end{aligned}$$

   But in this case the number of balls N has only to be approximately w^{-2}:

$$0 < (w+1)^4 - \frac{64 \cdot w^4 \cdot (w-1)}{N} \tag{4.2.7}$$

$$\Leftarrow 64 \cdot (w-1)^2 < N \tag{4.2.8}$$

$$\Leftarrow w < \frac{\sqrt{N}}{8} \tag{4.2.9}$$

   For illustration some plots showing the error for fixed N and variable weight w for the extreme cases with $A_0 = N$ and $A_0 = 0$ are given (see figure 4.3 and 4.1). Besides it is also shown in figure that the condition 4.2.1 is satisfied for the considered range of w (see figure 4.4 and 4.2).

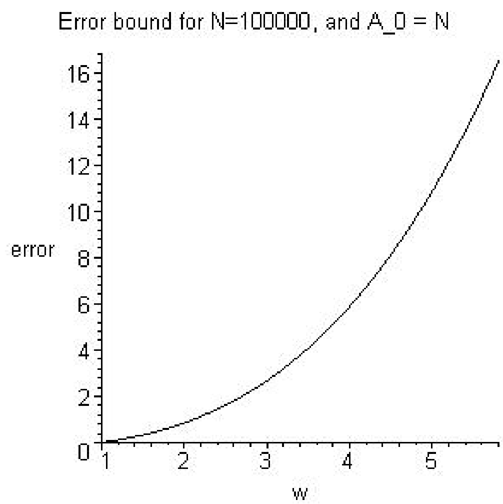

Figure 4.1: Error bound for $N = 100000$ and $A_0 = N$

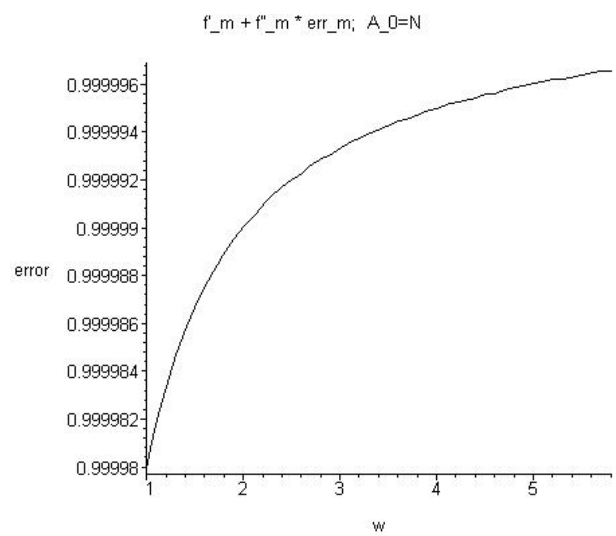

Figure 4.2: Condition 4.2.1 < 1 for $A_0 = N$, $N = 100000$

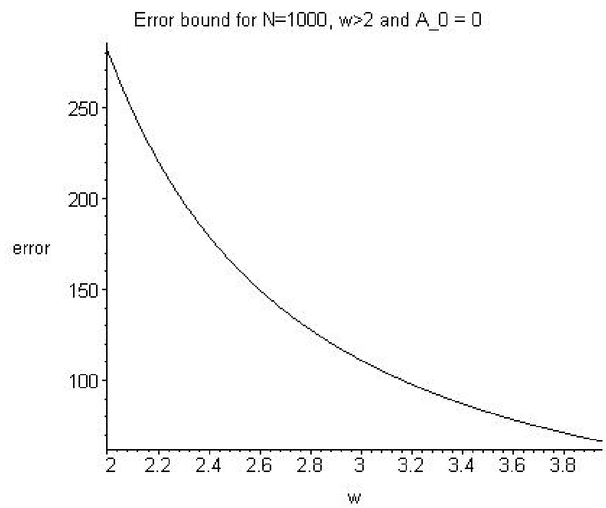

Figure 4.3: Error bound for $N = 1000$ and $A_0 = 0$

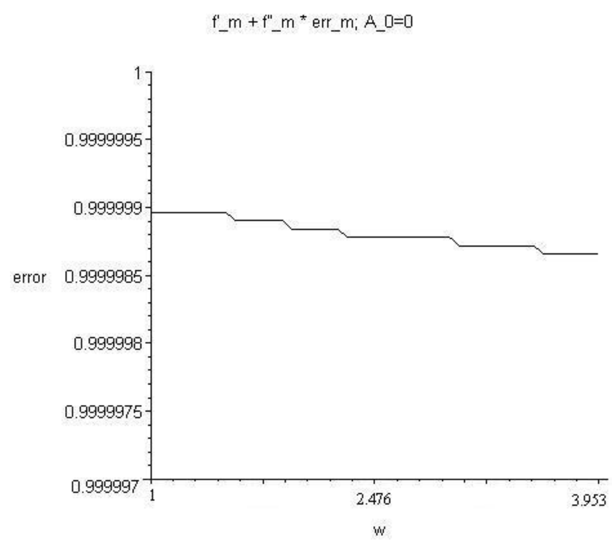

Figure 4.4: Condition (4.2.1) < 1 for $A_0 = 0$, $N = 1000$

2. A rough bound looks as follows: For $A_0 \leq b_f$ and $N = w^k$ with $k \geq 6$, we have:

$$1 - f'_m = 1 - f'(N) = 2 \cdot w^{-1-k}$$

and

$$f''_m = \frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^{2\cdot(1+k)}}$$

$$\begin{aligned} errb_{f+} &= \frac{2 \cdot w^{-1-k} + \sqrt{4 \cdot w^{2\cdot(-1-k)} - 4 \cdot (\frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^{2\cdot(1+k)}})^2 \cdot \frac{w^k}{4}}}{2 \cdot \frac{(w+1)^3 \cdot (w-1)}{2 \cdot w^{2\cdot(1+k)}}} \\ &\approx \frac{4 \cdot w^{1+k} + \sqrt{16 \cdot w^{2\cdot(1+k)} - w^8 \cdot w^k}}{2 \cdot w^4} \\ &= 4 \cdot w^{k-3} \end{aligned} \quad (4.2.10)$$

For $A_0 > b_f$ and $N = w^k$ with $k \geq 2$

$$1 - f'_m = 1 - f'(b_f) = \frac{(w+1)^2}{2 \cdot w^{k+1}}$$

and

$$f''_m = 4 \cdot w^{1-2\cdot k} \cdot (w-1)$$

So the bound for the error becomes:

$$\begin{aligned} errb_{f+} &= \frac{\frac{(w+1)^2}{2 \cdot w^{k+1}} + \sqrt{(\frac{(w+1)^2}{2 \cdot w^{k+1}})^2 - 4 \cdot (4 \cdot w^{1-2\cdot k} \cdot (w-1))^2 \cdot \frac{w^k}{4}}}{2 \cdot 4 \cdot w^{1-2\cdot k} \cdot (w-1)} \\ &\approx \frac{\frac{w^{1-k}}{2} + \sqrt{(\frac{w^{2\cdot(1-k)}}{4} - 16 \cdot w^{4\cdot(1-k)} \cdot w^k}}{8 \cdot w^{2\cdot(1-k)}} \\ &= \frac{w^{k-1} + \sqrt{w^{2\cdot(k-1)} - 64 \cdot w^k}}{16} \\ &= \frac{w^{k-1}}{8} \end{aligned} \quad (4.2.11)$$

4.3 2 Dimensions

Theorem 4.3.1 is a direct extension of theorem 4.1.1.

Theorem 4.3.1. *Let $(A,B)_0 = (a,b), (A,B)_1, \ldots, (A,B)_n$ be a Markov process given by definition 2.2.2 and $\overrightarrow{f}$ be a function as stated in 2.3.2. Let f_A and f_B be such that they and all of their partial derivatives up to order 2 are continuous on $I_{S_A} \times I_{S_B}$, then*

$$\begin{aligned} err_A(n,a,b) \;=\; & \frac{\partial f_A}{\partial A}(\overrightarrow{f}^{n-1}(a,b)) \cdot err_A(n-1,a,b) \\ & + \frac{\partial f_A}{\partial B}(\overrightarrow{f}^{n-1}(a,b)) \cdot err_B(n-1,a,b) \\ & + \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial B^2}(c(n-1)^{(A,dB)}, c(n-1)^{(B,dB)}) \cdot \left(var(B_{n-1}) + err_B(n-1,a,b)^2 \right) \\ & + \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)}) \cdot \left(var(A_{n-1}) + err_A(n-1,a,b)^2 \right) \\ & + \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^{(A,dAB)}, c(n-1)^{(B,dAB)}) \cdot \\ & \left(cov[A_{n-1}, B_{n-1}] + err_A(n-1,a,b) \cdot err_B(n-1,a,b) \right) \end{aligned}$$

where $c(n-1)^{(A,dA)}$, $c(n-1)^{(A,dB)}$ and $c(n-1)^{(A,dAB)}$ denote constants in the interval I_{S_A} and in the same way $c(n-1)^{(B,dA)}$, $c(n-1)^{(B,dB)}$ and $c(n-1)^{(B,dAB)}$ are constants in the interval I_{S_B}.

Remark 4.3.1.

1. The formula for $err_B(A_n, B_n)$ is symmetric, e.g. switching the indexes of err and f from A to B or the other way, results in a formula of the same type.
2. If the function f_A is linear in its arguments, then the partial derivatives $\frac{\partial^2 f_A}{\partial A \partial B} = 0, \frac{\partial^2 f_A}{\partial B^2} = 0$ and $\frac{\partial^2 f_A}{\partial A^2} = 0$ and therefore the approximation is exact for A. The same holds for f_B and B.
3. In case of a deterministic process ($\forall n \; var(A_n) = 0$, $var(B_n) = 0$ and $cov[A_n, B_n]) = 0$, there is also no error.

Since theorem 4.3.1 is a straight forward extension of the 1 dimensional case (e.g. theorem 4.1.1), the proof also follows the ideas for one dimension.

The derivation of the formula for the error of A_n $err_A(A_n, B_n)$ and the error of B_n $err_B(A_n, B_n)$ is exactly the same. Thus only the proof for $err_A(A_n, B_n)$ will be given.

Proof. As a first step $\mathbb{E}_A[A_n, B_n]$ is expressed in terms of f_A and the joint probability distribution of the random variables A_n and B_n.

$$
\begin{aligned}
\mathbb{E}_A[A_n, B_n] &= \sum_{(i,j)\in S} \mathbf{prob}((A_n, B_n) = (i,j)) \cdot i \\
&= \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \mathbb{E}_A[A_n, B_n | A_{n-1}, B_{n-1}] \\
&= \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot f_A(A_{n-1}, B_{n-1})]
\end{aligned}
$$

Next we use the Taylor expansion of the function f_A at point $\overrightarrow{f}^{n-1}(a,b)$ up to order 1. In order to do this, f_A must be continuous on $I_{S_A} \times I_{S_B}$ as well as its first partial derivatives and the 2nd order partial derivatives have to exist on $I_{S_A} \times I_{S_B}$. The Taylor expansion becomes:

$$
\begin{aligned}
f_A(x,y) = & \underbrace{f_A(\overrightarrow{f}^{n-1}(a,b))}_{f_A^n(a,b),\text{ due to definition, see 2.3.2}} \\
& + \frac{\partial f_A}{\partial A}(\overrightarrow{f}^{n-1}(a,b)) \cdot (x - f_A^{n-1}(a,b)) \\
& + \frac{\partial f_A}{\partial B}(\overrightarrow{f}^{n-1}(a,b)) \cdot (y - f_B^{n-1}(a,b)) \\
& + \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^A, c(n-1)^B) \cdot (x - f_A^{n-1}(a,b))^2 \\
& + \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial B^2}(c(n-1)^A, c(n-1)^B) \cdot (y - f_B^{n-1}(a,b))^2 \\
& + \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^A, c(n-1)^B) \cdot (x - f_A^{n-1}(a,b)) \cdot (y - f_B^{n-1}(a,b))
\end{aligned}
\tag{4.3.1}
$$

where $x \in I_{S_A}$, $y \in I_{S_B}$ and $c(n-1)^A$, $c(n-1)^B$ denote constants in the interval $[A_0,x]$ and $[B_0,y]$.

Using the previous Taylor expansion (4.3.1) and the obvious fact that the probabilities of all states add up to 1 ($\sum_{i,j\in S}\mathbf{prob}((A_n,B_n)=(i,j))=1$), the expectation of A_n becomes:

$$
\begin{aligned}
\mathbb{E}_A[A_n,B_n] &= \sum_{(i,j)\in S}\mathbf{prob}((A_{n-1},B_{n-1})=(i,j))\cdot f_A(A_{n-1},B_{n-1})] \\
&= \sum_{(i,j)\in S}\mathbf{prob}((A_{n-1},B_{n-1})=(i,j))\cdot\Big(f_A^n(a,b) \\
&\quad +\frac{\partial f_A}{\partial A}(\overrightarrow{f}^{n-1}(a,b))\cdot(i-f_A^{n-1}(a,b)) \\
&\quad +\frac{\partial f_A}{\partial B}(\overrightarrow{f}^{n-1}(a,b))\cdot(j-f_B^{n-1}(a,b)) \\
&\quad +\frac{1}{2}\cdot\frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot(i-f_A^{n-1}(a,b))^2 \\
&\quad +\frac{1}{2}\cdot\frac{\partial^2 f_A}{\partial B^2}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot(j-f_B^{n-1}(a,b))^2 \\
&\quad +\frac{\partial^2 f_A}{\partial A\partial B}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot \\
&\quad (i-f_A^{n-1}(a,b))\cdot(j-f_B^{n-1}(a,b))\Big) \\
&= f_A^n(a,b) \\
&\quad +\frac{\partial f_A}{\partial A}(\overrightarrow{f}^{n-1}(a,b))\cdot\underbrace{(\mathbb{E}_A[(A_{n-1},B_{n-1})]-f_A^{n-1}(a,b))}_{err_A(n-1,a,b)} \\
&\quad +\frac{\partial f_A}{\partial B}(\overrightarrow{f}^{n-1}(a,b))\cdot\underbrace{(\mathbb{E}_B[(A_{n-1},B_{n-1})]-f_B^{n-1}(a,b))}_{err_B(n-1,a,b)} \\
&\quad +\sum_{(i,j)\in S}\mathbf{prob}((A_{n-1},B_{n-1})=(i,j))\cdot \\
&\quad \Big(\frac{1}{2}\cdot\frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot(i-f_A^{n-1}(a,b))^2 \\
&\quad +\frac{1}{2}\cdot\frac{\partial^2 f_A}{\partial B^2}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot(j-f_B^{n-1}(a,b))^2 \\
&\quad +\frac{\partial^2 f_A}{\partial A\partial B}(c(n-1,(i,j))^A,c(n-1,(i,j))^B)\cdot \\
&\quad (i-f_A^{n-1}(a,b))\cdot(j-f_B^{n-1}(a,b))\Big) \qquad (4.3.2)
\end{aligned}
$$

In the above formula $c(n-1,(i,j))$ denotes a constant, which depends on the step $n-1$ and the state $(i,j)\in S$. Next this constant will be replaced for each term (e.g. the ones containing $\frac{\partial^2 f_A}{\partial A\partial B},\frac{\partial^2 f_A}{\partial A^2}$ and $\frac{\partial^2 f_A}{\partial B^2}$) by another one, which is independent of the state (i,j).

To do so, some prerequisites are required.

Lemma 4.3.2. *Let $d(i,j)$ and $n(k,l)$ be two sequences of non-negative numbers and $\forall(i,j)\in S$ let $d(i,j)\in[a,b]$, then there exists $\overline{d}\in[a,b]$ such that*

$$\sum_{(i,j)\in S}d(i,j)\cdot n(i,j)=\overline{d}\cdot\sum_{(i,j)\in S}n(i,j)$$

.

Proof. In case $\sum_{(i,j)\in S} n(i,j) = 0$, $\overline{d}$ can be any number in $[a,b]$. If not, then one can define the function p as $p : (i,j) \mapsto \frac{n(i,j))}{\sum_{(k,l)\in S} n(k,l)}$. Obviously $p(i,j) \geq 0$ and $\sum_{(i,j)\in S} p(i,j) = 1$. So p is a probability distribution and $\overline{d} = \sum_{(i,j)\in S} d(i,j) \cdot p(i,j)$ is simply the expectation, which exists and is always within $[a,b]$.

□

Lemma 4.3.2 will now be applied. This can be done since all partial derivatives are continuous on the considered interval. The details will only be given for $\frac{\partial^2 f_A}{\partial A^2}$. Because of its continuity, there exist for any n and all

$$\begin{aligned} f_0'' \;\in\; & [\min_{(i,j)} \frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B), \\ & \max_{(i,j)} \frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B)] \end{aligned}$$

an

$$x \in [\min_{(i,j)} c(n-1,(i,j))^A, \max_{(i,j)} c(n-1,(i,j))^A]$$

and a

$$y \in [\min_{(i,j)} c(n-1,(i,j))^B, \max_{(i,j)} c(n-1,(i,j))^B]$$

such that $\frac{\partial^2 f_A}{\partial A^2}(x,y) = f_0''$ where $(i,j) \in S$.

For that reason Lemma 4.3.2 can be applied with

$$d(i,j) = \frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B)$$

and

$$n(i,j) = \frac{1}{2} \cdot \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - f_A^{n-1}(a,b))^2$$

which yields that there exists a pair $(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)})$ such that

$$\begin{aligned} & \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \frac{1}{2} \cdot \\ & \frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B) \cdot (i - f_A^{n-1}(a,b))^2 \\ & = \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)}) \cdot \\ & \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - f_A^{n-1}(a,b))^2 \end{aligned}$$

In the above formula $c(n-1)^{(B,dA)}$ is a constant, which depends on step $n-1$. The symbol dA should indicate, that this constant is related to the term containing $\frac{\partial^2 f_A}{\partial A^2}$.

To further simplify the remaining expression, remember that the error was defined as $err_A(A,B) := \mathbb{E}_A[A,B] - f_A(A,B)$.

$$
\begin{aligned}
&\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \frac{1}{2} \cdot \\
&\frac{\partial^2 f_A}{\partial A^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B) \cdot (i - f_A^{n-1}(a,b))^2 \\
&= \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)}) \cdot \\
&\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - \mathbb{E}_A[A_{n-1}, B_{n-1}] + err_A(n-1,a,b))^2 \\
&= \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)}) \cdot \\
&\Bigg(\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - \mathbb{E}_A[A_{n-1}, B_{n-1}])^2 \\
&+ \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - \mathbb{E}_A[A_{n-1}, B_{n-1}]) \cdot err_A(n-1,a,b) \\
&+ \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot err_A(n-1,a,b)^2 \Bigg) \\
&= \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial A^2}(c(n-1)^{(A,dA)}, c(n-1)^{(B,dA)}) \cdot \Big(var(A_{n-1}) + err_A(n-1,a,b)^2 \Big) \qquad (4.3.3)
\end{aligned}
$$

The derivation for the term in (4.3.2) containing $\frac{\partial^2 f_A}{\partial B^2}$ is completely analogous. Therefore only the result is given:

$$
\begin{aligned}
&\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \frac{1}{2} \cdot \\
&\frac{\partial^2 f_A}{\partial B^2}(c(n-1,(i,j))^A, c(n-1,(i,j))^B) \cdot (i - f_A^{n-1}(a,b))^2 \\
&= \frac{1}{2} \cdot \frac{\partial^2 f_A}{\partial B^2}(c(n-1)^{(A,dB)}, c(n-1)^{(B,dB)}) \cdot \\
&\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - f_A^{n-1}(a,b))^2 \qquad (4.3.4)
\end{aligned}
$$

Next the term in (4.3.2) containing $\frac{\partial^2 f_A}{\partial A \partial B}$ is examined.

$$
\begin{aligned}
&\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \\
&\frac{\partial^2 f_A}{\partial A \partial B}(c(n-1,(i,j))^A, c(n-1,(i,j))^B) \cdot (i - f_A^{n-1}(a,b)) \cdot (j - f_B^{n-1}(a,b)) \\
&= \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^{(A,dAB}, c(n-1)^{(B,dAB)}) \\
&\cdot \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - f_A^{n-1}(a,b)) \cdot (j - f_B^{n-1}(a,b)) \\
&= \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^{(A,dAB}, c(n-1)^{(B,dAB)}) \\
&\cdot \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \\
&(i - \mathbb{E}_A[A_{n-1}, B_{n-1}] + err_A(n-1,a,b)) \cdot (j - \mathbb{E}_B[A_{n-1}, B_{n-1}] + err_B(n-1,a,b)) \\
&= \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^{(A,dAB}, c(n-1)^{(B,dAB)}) \cdot \\
&\Bigg(\sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot (i - \mathbb{E}_A[A_{n-1}, B_{n-1}]) \cdot (j - \mathbb{E}_B[A_{n-1}, B_{n-1}]) \\
&+ \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot \Big((i - \mathbb{E}_A[A_{n-1}, B_{n-1}]) \cdot err_B(n-1,a,b) \\
&+ (j - \mathbb{E}_B[A_{n-1}, B_{n-1}]) \cdot err_A(n-1,a,b) \Big) \\
&+ \sum_{(i,j)\in S} \mathbf{prob}((A_{n-1}, B_{n-1}) = (i,j)) \cdot err_B(n-1,a,b) \cdot err_A(n-1,a,b) \\
&= \frac{\partial^2 f_A}{\partial A \partial B}(c(n-1)^{(A,dAB}, c(n-1)^{(B,dAB)}) \cdot \\
&\Big(cov[A_{n-1}, B_{n-1}] + err_A(n-1,a,b) \cdot err_B(n-1,a,b) \Big) \qquad (4.3.5)
\end{aligned}
$$

Plugging the results for the partial derivatives (4.3.3),(4.3.4) and (4.3.5) into (4.3.2) and using the definition of the error completes the proof. □

Chapter 5

Linearisation error

This chapter includes a technique to bound the error of the pseudo expectation in one and two dimensions (see section 5.1 and 5.3). It is applied to an example process in section 5.2. The technique is independent of the probability distribution of the states and uses only the properties of the function f, used to calculated the pseudo expectation.

5.1 Error bound in 1 dimension

In this section a new concept for expressing the error is introduced. The linearisation error $dL(n, a)$ for step n and state a is defined as:

$$dL(n, a) := (\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot f^n(x)) - f^{n+1}(a)$$

Since the pseudo expectation is correct for linear functions f, the error depends only on the non-linear part of the function f^n. The Taylor expansion of f^n at $f(a)$ can be written as:

$$f^n(x) = \underbrace{f^n(f(a))}_{f^{n+1}(a)} + (f^n)'(f(a)) \cdot \Big(x - f(a)\Big) + \frac{(f^n)''(c)}{2} \cdot \Big(x - f(a)\Big)^2 \qquad (5.1.1)$$

where c is a constant in $[x, f(a)]$. In order to do this $(f^n)'$ must be continuous and $(f^n)''$ must exist in the interval $[x, f(a)]$. As will be deduced later, the 1^{st} derivative of f^n can be expressed as a product of $f'(f^i(x))$ where $f^i(x) \in I_S$ (see 5.1.3), therefore it is enough to show that f' is continuous on I_S. The 2^{nd} derivative of f^n can as well be expressed in terms of products and sums of $f'(f^n(x))$ and $f''(f^n(x))$ (see 5.1.5) which implies that the Taylor expansion is possible, if $f''(x)$ exists for $x \in I_S$.

Remark:
Of course, a Taylor expansion of a higher (or lower) order can be considered as well (given that f fulfills the requirements to do so). A lower order expansion was considered in section 5.2.2.

Theorem 5.1.1. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.2.1, and let f' be continuous on S and let f'' exist on S then for the linearisation error $dL(n,a)$, it holds that*

$$dL(n,a) \leq \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \frac{(f^n)''(c_x)}{2} \cdot \Big(x - f(a) \Big)^2$$

where c_x denotes a constant in $[x, f(a)]$.

Proof. Using the Taylor expansion (5.1.1) in the definition of $dL(n,a)$ yields:

$$\begin{aligned} dL(n,a) &= \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \Big(f^{n+1}(a) + (f^n)'(f(a)) \cdot \Big(x - f(a) \Big) \\ &\quad + \frac{(f^n)''(c_x)}{2} \cdot \Big(x - f(a) \Big)^2 \Big) - f^{n+1}(a) \\ &= \underbrace{\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot f^{n+1}(a)}_{f^{n+1}(a)} \\ &\quad + (f^n)'(f(a)) \cdot \underbrace{\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot (x - f(a)) +}_{=0} \\ &\quad \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \frac{(f^n)''(c_x)}{2} \cdot \Big(x - f(a) \Big)^2 - f^{n+1}(a) \\ &= \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \frac{(f^n)''(c_x)}{2} \cdot \Big(x - f(a) \Big)^2 \end{aligned}$$

□

Besides, the error of the pseudo expectation can be expressed as a sum of linearisation errors:

Lemma 5.1.2. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.2.1, then for the error of the pseudo expectation holds:*

$$err(n) \leq \sum_{i=2}^{n-1} \max_{a \in S} dL(i,a)$$

Proof.

$$\begin{aligned} err(n+1,a) &= \mathbb{E}[A_{n+1} | A_0 = a] - f^{n+1}(a) \\ &= (\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \mathbf{E}[A_n | A_1 = x]) - f^{n+1}(a) \\ &= (\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot (\mathbf{E}[A_n | A_1 = x] - f^n(x)) \\ &\quad + (\sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot f^n(x) - f^{n+1}(a)) \\ &= \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot err(n,x) + dL(n,a) \qquad (5.1.2) \end{aligned}$$

The recursive evaluation becomes quickly computationally infeasible for growing n and Markov chains, where a large number of states has more than one transition.

For the example process, described in section 2.4 holds that, each state (apart from 0 and N) has two transitions, thus to calculate a bound for the n^{th} step requires $O(2^n)$ given that $dL(n,a)$ is in $O(1)$.

Apart from that, it is hard to obtain a closed form for (5.1.2) in general. Because of this, expression (5.1.2) is further bounded:

$$\begin{aligned} err(n+1) &\leq \max_{b\in S} err(n,b) + dL(n,a) \\ &= \sum_{i=2}^{n} \max_{b\in S} dL(i,b) \end{aligned}$$

Observe that the error for the first step is 0 since the function f is defined as $f(A_n) := \mathbb{E}[A_{n+1}|A_n]$ (see 2.3.1). In mathematical notation: $err(1,a) = 0$. Thus the summation starts from 2.

Besides, note that the error bound is now independent of the starting state and moreover of the probability distribution for A_n. □

Corollary 5.1.3. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in section 2.2.1, then for the error of the pseudo expectation holds:*

$$err(n+1) \leq \sum_{i=2}^{n} \max_{b\in S} \sum_{x\in S} \mathbf{prob}(A_1 = x|A_0 = b) \cdot \frac{(f^i)''(c_x^b)}{2} \cdot \Big(x - f(b)\Big)^2$$

where c_x^b denotes a constant in $[x, f(b)]$.

Remark 5.1.1.
In total there are $n \cdot |S|$ constants).

As a next step, the second derivative $(f^n)''$ is expressed in terms of sums and products of f' and f'' depending on $f^i(x)$. This might be useful, since non-recursive expressions of the derivatives of f^n become complicated even for small n. This also holds for the function f^n itself, but in some cases it might be easier to deal with f^n than with its derivatives. For an example, see section 5.2.

Primarily, a recurrence and explicit formula in terms of f^i for computing the first derivative of f^n is given. For $n \geq 1$ it is true that:

$$\begin{aligned} (f^n(x))' &= (f(f^{n-1}(x)))' \\ &= f'(f^{n-1}(x)) \cdot (f(f^{n-2}(x)))' \\ &= f'(f^{n-1}(x)) \cdot (f(f^{n-2}(x)))' \\ &= \prod_{i=0}^{n-1} f'(f^i(x)) \end{aligned} \tag{5.1.3}$$

The product for $\prod_{i=0}^{-1} f'(f^i(x))$ is defined as 1. Because the derivative of $(f^0)(x) := x$ is 1 as well, the formula in product form is valid for $n \geq 0$.

Next the second derivative of f^n for $n \geq 0$ will be investigated. For n equals 0, the second derivative is 0:

$$(f^0(x))'' = (x)'' = 0$$

$(f^n(x))''$ can also be computed recursively using expression (5.1.3) for $(f^n)'$:

$$\begin{aligned}
(f^n(x))'' &= \Big(f'(f^{n-1}(x)) \cdot f^{n-1}(x)'\Big)' \\
&= f''(f^{n-1}(x)) \cdot (f^{n-1}(x)')^2 + f'(f^{n-1}(x)) \cdot f^{n-1}(x)'' \\
&= f'(f^{n-1}(x)) \cdot f^{n-1}(x)'' + f''(f^{n-1}(x)) \cdot (\prod_{i=0}^{n-1} f'(f^i(x)))^2 \\
&= f'(f^{n-1}(x)) \cdot f^{n-1}(x)'' + f''(f^{n-1}(x)) \cdot \prod_{i=0}^{n-1} f'(f^i(x))^2 \qquad (5.1.4)
\end{aligned}$$

This can also be expressed non-recursively in a straight forward fashion, which can easily be seen by substituting $(f^n)''$ by (5.1.4) and replacing again $(f^{n-1})''$ in the result by (5.1.4) and so on.

$$
\begin{aligned}
(f^n(x))'' &= f'(f^{n-1}(x)) \cdot f^{n-1}(x)'' + f''(f^{n-1}(x)) \cdot \prod_{i=0}^{n-1} f'(f^i(x))^2 \\
&= f'(f^{n-1}(x)) \cdot \left(f'(f^{n-2}(x)) \cdot f^{n-2}(x)'' + f''(f^{n-2}(x)) \cdot \prod_{i=0}^{n-2} f'(f^i(x))^2 \right) \\
&\quad + f''(f^{n-1}(x)) \cdot \prod_{i=0}^{n-1} f'(f^i(x))^2 \\
&= f'(f^{n-1}(x)) \cdot f'(f^{n-2}(x)) \cdot f^{n-2}(x)'' \\
&\quad + f'(f^{n-1}(x)) \cdot f''(f^{n-2}(x)) \cdot \prod_{i=0}^{n-2} f'(f^i(x))^2 \\
&\quad + f''(f^{n-1}(x)) \cdot \prod_{i=0}^{n-1} f'(f^i(x))^2 \\
&= f'(f^{n-1}(x)) \cdot f'(f^{n-2}(x)) \cdot f'(f^{n-3}(x)) \cdot f^{n-3}(x)'' \\
&\quad + f'(f^{n-1}(x)) \cdot f'(f^{n-2}(x)) \cdot f''(f^{n-3}(x)) \cdot \prod_{i=0}^{n-3} f'(f^i(x))^2 \\
&\quad + f'(f^{n-1}(x)) \cdot f''(f^{n-2}(x)) \cdot \prod_{i=0}^{n-2} f'(f^i(x))^2 \\
&\quad + f''(f^{n-1}(x)) \cdot \prod_{i=0}^{n-1} f'(f^i(x))^2 \\
&= \sum_{i=0}^{n-1} \left(\prod_{k=i+1}^{n-1} f'(f^k(x)) \cdot f''(f^i(x)) \cdot \prod_{j=0}^{i} f'(f^j(x))^2 \right) \\
&= \prod_{k=0}^{n-1} f'(f^k(x)) \cdot \left(\sum_{i=0}^{n-1} f''(f^i(x)) \cdot \prod_{j=0}^{i} f'(f^j(x)) \right) \qquad (5.1.5)
\end{aligned}
$$

5.2 Bound for example

Now the technique presented in the previous section 5.1 is applied to the example process defined in 2.4. First (in section 5.2.1) the Taylor expansion of f up to order 1, e.g. involving expressions up to the second derivative is used. In 5.2.2 this bound is improved by avoiding the second derivative.

5.2.1 Using 2nd derivative

Theorem 5.2.1. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.4, then for the error of the pseudo expectation holds:*

$$err(n) \leq 8 + \frac{1}{16}(w-1)\cdot(w+1)^3$$

In essence, theorem 5.2.1 is an application of the technique presented in the previous section 5.1, where the error can be bounded by a sum of second derivatives $(f^n)''$. These can be written in terms of sums and products of f' and f'' (see 5.1.5). This expression for $(f^n)''$ will be upper bounded to prove theorem 5.2.1.

Due to definition the linearisation error $dL(n,a)$ for step n at state a is

$$\begin{aligned} dL(n,a) &= \mathbf{prob}(A_{n+1} = a-1 | A_n = a) \cdot f^n(a-1) \\ &\quad + \mathbf{prob}(A_{n+1} = a+1 | A_n = a) \cdot f^n(a+1) - f^{n+1}(a) \end{aligned}$$

Thus for the example, graphically speaking, the linearisation error $dL(n,a)$ equals the distance between the function f^n and a straight line between the two points $f^n(a-1)$ and $f^n(a+1)$ for the argument $f(a)$ (see figure 5.1).

Before stating the proof of 5.2.1 some prerequisites are derived.

Corollary 5.2.2. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.2.1, then for the error of the pseudo expectation holds:*

$$err(n) \leq \sum_{i=1}^{n-1} \max_{a\in[1,N-1]} dL(i,a)$$

Proof. Essentially, this follows from theorem 5.1.2. Only a small detail namely the range of states is not the continuous interval between the minimum and maximum state of the state space $S = 0, 1, \ldots, N$, e.g. $[0, N]$, but instead $[1, N-1]$. For the linearisation error for $a \in \{0, N\}$ holds:

$$\begin{aligned} dL(n,0) &= \underbrace{\mathbf{prob}(A_{n+1} = -1 | A_n = 0)}_{=0} \cdot f^n(-1) + \underbrace{\mathbf{prob}(A_{n+1} = 1 | A_n = 0)}_{=1} \cdot f^n(1) - \underbrace{f^{n+1}(0)}_{f^n(1)} \\ &= dL(n-1, 1) \end{aligned}$$

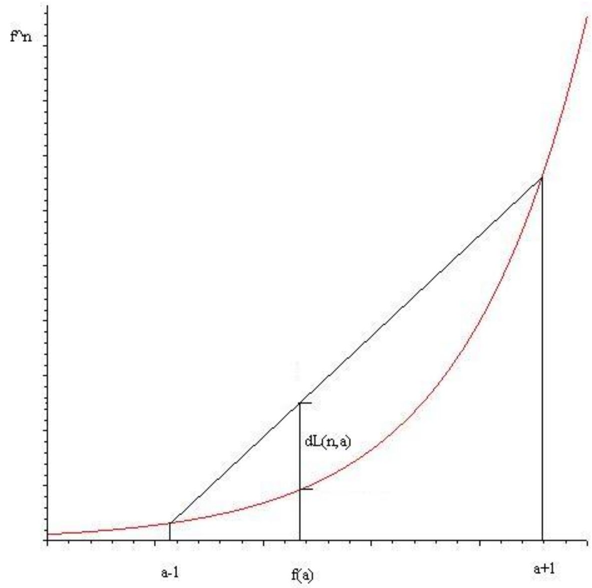

Figure 5.1: Linearisation error for example

and analogously

$$dL(n, N) = dL(n-1, N-1)$$

In other words, if the state a lies on the boundary the linearisation error for the first step is 0. □

Lemma 5.2.3. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as in 2.4, then for the linearisation error $dL(n,a)$ holds that*

$$dL(n,a) \leq \max_{x \in [0,N]} \frac{(f^n)''(x)}{2}$$

Proof. The results basically follows from theorem 5.1.1. Only some details need to be derived.

Recall that

$$f(a) := a - \mathbf{prob}(A_{n+1} = a-1 | A_n = a) + \mathbf{prob}(A_{n+1} = a+1 | A_n = a)$$

and

$$\mathbf{prob}(A_{n+1} = a-1 | A_n = a) + \mathbf{prob}(A_{n+1} = a+1 | A_n = a) = 1$$

Because of this,

$$\begin{aligned}(a-1) - f(a) &= (a - \mathbf{prob}(A_{n+1} = a-1 | A_n = a) - \mathbf{prob}(A_{n+1} = a+1 | A_n = a)) \\ &\quad -(a - \mathbf{prob}(A_{n+1} = a-1 | A_n = a) + \mathbf{prob}(A_{n+1} = a+1 | A_n = a)) \\ &= -2 \cdot \mathbf{prob}(A_{n+1} = a+1 | A_n = a) \qquad (5.2.1)\end{aligned}$$

and in the same way

$$\begin{aligned}
(a+1)-f(a) &= (a+\mathbf{prob}(A_{n+1}=a-1|A_n=a)+\mathbf{prob}(A_{n+1}=a+1|A_n=a)) \\
&\quad -(a-\mathbf{prob}(A_{n+1}=a-1|A_n=a)+\mathbf{prob}(A_{n+1}=a+1|A_n=a)) \\
&= 2\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a) \qquad (5.2.2)
\end{aligned}$$

The first derivative f' and the second f'' are obviously continuous for $x \in [0, N]$ as shown in 2.4.

Thus theorem 5.1.1 can be applied. (5.2.1) and (5.2.2) will also be used:

$$\begin{aligned}
dL(n,a) &= \sum_{x\in S}\mathbf{prob}(A_1=x|A_0=a)\cdot\frac{(f^n)''(c_x)}{2}\cdot\Big(x-f(a)\Big)^2 \\
&= \mathbf{prob}(A_{n+1}=a-1|A_n=a)\cdot\left(\frac{(f^n)''(c_{-1})}{2}\cdot\Big(a-1-f(a)\Big)^2\right) \\
&\quad +\mathbf{prob}(A_{n+1}=a+1|A_n=a)\cdot\left(\frac{(f^n)''(c_{+1})}{2}\cdot\Big(a+1-f(a)\Big)^2\right) \\
&= \mathbf{prob}(A_{n+1}=a-1|A_n=a)\cdot\left(\frac{(f^n)''(c_{-1})}{2}\cdot 2\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a)^2\right) \\
&\quad +\mathbf{prob}(A_{n+1}=a+1|A_n=a)\cdot\left(\frac{(f^n)''(c_{+1})}{2}\cdot 2\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a)^2\right) \\
&= 2\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a)\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a)\cdot \\
&\quad \Big((f^n)''(c_{-1})\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a) \\
&\quad +(f^n)''(c_{+1})\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a)\Big)
\end{aligned}$$

The constants c_{+1} and c_{-1} are undetermined values in the interval $[f(a), a+1]$ and $[a-1, f(a)]$ respectively.

As maximum 1 ball is moved per step $a-1 \le f(a) \le a+1$ and also $0 \le a-1 < a+1 \le N$, a constant $c_m \in [0, N]$ can be chosen such that

$$(f^n)''(c_{-1}) \le (f^n)''(c_m) \text{ and also } (f^n)''(c_{+1}) \le (f^n)''(c_m)$$

This allows to bound $dL(n, a)$ and to write it in a simpler way:

$$\begin{aligned}
dL(n,a) &= 2\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a)\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a)\cdot \\
&\quad \Big((f^n)''(c_{-1})\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a) \\
&\quad +(f^n)''(c_{+1})\cdot\mathbf{prob}(A_{n+1}=a-1|A_n=a)\Big) \\
&\le 2\cdot\underbrace{\mathbf{prob}(A_{n+1}=a-1|A_n=a)\cdot\mathbf{prob}(A_{n+1}=a+1|A_n=a)}_{\le\frac{1}{4}}\cdot\max_{c_m\in[0,N]}(f^n)''(c_m) \\
&\le \max_{c_m\in[0,N]}\frac{(f^n)''(c_m)}{2}
\end{aligned}$$

This completes the proof of lemma 5.2.3. □

Finally, the proof of theorem 5.2.1 will be pointed out.

Proof. Note that $\max_{x\in[0,N]}(f^n(x))''$ can be bounded as follows:

$$\begin{aligned}\max_{x\in[0,N]}(f^n)''(x) &= \max\{\max_{x\in[0,b_f]}(f^n)''(x), \max_{x\in[b_f,N]}(f^n)''(x)\} \\ &\leq \max_{x\in[0,b_f]}(f^n)''(x) + \max_{x\in[b_f,N]}(f^n)''(x) \qquad (5.2.3)\end{aligned}$$

Also observe that for any $j \in N_0$

$$f^j(x) \in [0, b_f] \text{ for } x \in [0, b_f]$$

and

$$f^j(x) \in [b_f, N] \text{ for } x \in [b_f, N]$$

since $0 \leq f'(x) < 1$ as shown in 2.4. Thus for an arbitrary function g, e.g. f' and f'', and for $x \in [a, b]$, where either $[a, b] = [0, b_f]$ or $[a, b] = [b_f, N]$ holds that

$$\max_{k\in N_0} g(f^k(x)) = \max_{c\in[a,b]} g(c) \qquad (5.2.4)$$

This allows to select the maximum independent of step k.

Using the previous observation and the non recursive formula (5.1.5) for the interval $[a, b] \in \{[0, b_f], [b_f, N]\}$ allows to obtain a bound for $(f^n)''(x)$ with $x \in [a, b]$ without any sums and products, which limits depends on n.

$$\begin{aligned}\max_{x\in[a,b]}(f^n)''(x) &= \max_{x\in[a,b]} \prod_{k=0}^{n-1} f'(f^k(x)) \cdot \left(\sum_{i=0}^{n-1} f''(f^i(x)) \cdot \prod_{j=0}^{i} f'(f^j(x)) \right) \\ &\leq \max_{c_1,c_2\in[a,b]} \prod_{k=0}^{n-1} f'(c_1) \cdot \left(\sum_{i=0}^{n-1} f''(c_2) \cdot \prod_{j=0}^{i} f'(c_1) \right) \\ &= \prod_{k=0}^{n-1} f'(b) \cdot \left(\sum_{i=0}^{n-1} f''(a) \cdot \prod_{j=0}^{i} f'(b) \right) \qquad (5.2.5) \\ &= f''(a) \cdot f'(b)^n \cdot \left(\sum_{i=0}^{n-1} f'(b)^i \right) \\ &= \frac{f''(a) \cdot f'(b)^n \cdot (f'(b)^n - 1)}{f'(b) - 1} \qquad (5.2.6)\end{aligned}$$

In order to show step (5.2.5) of this deduction, the first and second derivative are analyzed:

$$\max_{x\in[a,b]} f''(x) = \max_{x\in[a,b]} \frac{4 \cdot w \cdot (w-1) \cdot N}{(N + (w-1) \cdot x)^3}$$

The maximum depends on whether $w < 1$ or $w > 1$. As stated in the conditions for theorem (5.2.1) to be true, only the case $w > 1$ is considered. This gives the biggest f'' for the smallest value in the considered interval.

For $f'(f^i(x))$ the same advisements are valid. But by looking at

$$f'(x) = 1 - \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot x)^2}$$

it can be seen that the maximum is reached for the maximum possible value for x.

Next $\sum_{n=1}^{\infty} \max_{a_n \in [0,N]} (f^n)''(a_n)$ will be examined. When doing this, the bound (5.2.6) for $\max_{x \in [a,b]} (f^n)''(x)$, where $x \in [a,b]$ is used as well as (5.2.3).

$$\begin{aligned}
\sum_{n=1}^{\infty} \max_{a_n \in [0,N]} (f^n)''(a_n) &\leq \sum_{n=1}^{\infty} \Bigg(\max_{x \in [0,b_f]} (f^n)''(x) + \max_{x \in [b_f,N]} (f^n)''(x) \Bigg) \\
&= \sum_{n=1}^{\infty} \Bigg(\frac{f''(0) \cdot f'(b_f)^n \cdot (f'(b_f)^n - 1)}{f'(b_f) - 1} \\
&\quad + \frac{f''(b_f) \cdot f'(N)^n \cdot (f'(N)^n - 1)}{f'(N) - 1} \Bigg) \\
&= \sum_{n=1}^{\infty} \Bigg(\frac{f''(0) \cdot f'(b_f)^n \cdot (f'(b_f)^n - 1)}{f'(b_f) - 1} \Bigg) \\
&\quad + \sum_{n=1}^{\infty} \Bigg(\frac{f''(b_f) \cdot f'(N)^n \cdot (f'(N)^n - 1)}{f'(N) - 1} \Bigg)
\end{aligned} \tag{5.2.7}$$

Next a closed formula for an expression of the form (5.2.7) is stated.

Let $0 \leq c_1 < 1$ and let c_2 be an arbitrary positive constant then

$$\begin{aligned}
\sum_{n=1}^{\infty} \Bigg(\frac{c_2 \cdot c_1^n \cdot (c_1^n - 1)}{c_1 - 1} \Bigg) &= \frac{c_2}{c_1 - 1} \cdot \sum_{n=0}^{\infty} \Bigg(c_1^{2 \cdot n + 1} - c_1^n \Bigg) \\
&= \frac{c_2}{c_1 - 1} \cdot \Bigg(\frac{c_1}{1 - c_1^2} - \frac{1}{1 - c_1} \Bigg) \\
&= \frac{c_2}{(c_1 - 1) \cdot (1 - c_1)} \cdot \Bigg(\frac{c_1}{1 + c_1} - 1 \Bigg) \\
&= \frac{c_2}{(1 - c_1)^2 \cdot \underbrace{(1 + c_1)}_{>1}} \\
&\leq \frac{c_2}{(1 - c_1)^2}
\end{aligned} \tag{5.2.8}$$

The previous formula can be applied to our problem, since $\frac{N}{2} > w > 1$ and therefore $0 < f' < 1$.

Using it for the first case, meaning $A_0 \leq b_f$ and therefore $a = f'(b_f)$ and $b = f''(0)$, where

$$\begin{aligned}
f'(\frac{N}{w+1}) &= 1 - \frac{2 \cdot w \cdot N}{(N + (w-1) \cdot \frac{N}{w+1})^2} \\
&= 1 - \frac{2 \cdot w}{N \cdot (1 + \frac{w-1}{w+1})^2} \\
&= 1 - \frac{(w+1)^2}{2 \cdot w \cdot N}
\end{aligned}$$

and

$$f''(0) = \frac{4 \cdot w \cdot (w-1)}{N^2}$$

yields:

$$\begin{aligned}
\frac{f''(0)}{(f'(b_f)-1)^2} &= \frac{\frac{4 \cdot w \cdot (w-1)}{N^2}}{(\frac{(w+1)^2}{2 \cdot w \cdot N})^2} \\
&= \frac{16 \cdot w^3 \cdot (w-1)}{(w+1)^4} < 16 \qquad (5.2.9)
\end{aligned}$$

This result is quite surprising, since the error is less than a constant for any N and $1 < w < \frac{N}{2}$.

Using (5.2.8) for the second case ($A_0 \geq b_f$) implying $a = f'(N)$ and $b = f''(b_f)$, where

$$f'(N) = 1 - \frac{2}{w \cdot N}$$

and

$$f''(\frac{N}{w+1}) = \frac{(w-1) \cdot (w+1)^3}{2 \cdot (w \cdot N)^2}$$

gives

$$\begin{aligned}
\frac{f''(b_f)}{(f'(N)-1)^2} &= \frac{\frac{(w-1) \cdot (w+1)^3}{2 \cdot (w \cdot N)^2}}{(\frac{2}{w \cdot N})^2} \\
&= \frac{1}{8}(w-1) \cdot (w+1)^3 \qquad (5.2.10)
\end{aligned}$$

This result is much worse, since it depends on w and for $w > \sqrt[4]{N}$ it becomes bigger than the trivial bound,e.g. $err(n) \leq N$.

Adding the bounds for ($A_0 \geq b_f$) and ($A_0 < b_f$) gives due to (5.2.7):

$$\begin{aligned}
\sum_{n=1}^{\infty} \max_{a_n \in [0,N]} (f^n)''(a_n) &\leq \sum_{n=1}^{\infty} \left(\frac{f''(0) \cdot f'(b_f)^n \cdot (f'(b_f)^{n+1} - 1)}{f'(b_f) - 1} \right) \\
&\quad + \sum_{n=1}^{\infty} \left(\frac{f''(b_f) \cdot f'(N)^n \cdot (f'(N)^{n+1} - 1)}{f'(N) - 1} \right) \\
&= 16 + \frac{1}{8}(w-1) \cdot (w+1)^3
\end{aligned}$$

Combining this with both corollary 5.2.2 and lemma 5.2.3:

$$\begin{aligned} err(n) &\leq \sum_{i=1}^{\infty} \max_{a_i \in [0,N]} dL(i, a_i) \\ &\leq \frac{1}{2} \cdot \sum_{i=1}^{\infty} \max_{a_i \in [0,N]} (f^i)''(a_i) \\ &\leq \frac{1}{2} \cdot \sum_{i=1}^{\infty} \max_{a_i \in [0,b_f]} (f^i)''(x) + \frac{1}{2} \cdot \sum_{i=1}^{\infty} \max_{a_i \in [b_f,N]} (f^i)''(x) \text{ due to ref (5.2.3)} \\ &\leq 8 + \frac{1}{16} \cdot (w-1) \cdot (w+1)^3 \end{aligned}$$

This concludes the proof of theorem 5.2.1.

□

5.2.2 Using 1st derivative

Theorem 5.2.4. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.4, then for the error of the pseudo expectation holds:*

$$err(n) \leq 21 \cdot w + 132$$

Remark 5.2.1.
Experiments have indicated that the bound must be in $O(w)$. Thus it seems that $O(w)$ is the best one can hope for, when the probability distribution of the states is ignored.

Proof. In essence, theorem 5.2.1 is an application of the technique presented in the previous section 5.1, where the Taylor expansion is only done up to the first derivative:

$$f^n(x) = \underbrace{f^n(f(a))}_{f^{n+1}(a)} + (f^n)'(c) \cdot \Big(x - f(a) \Big) \tag{5.2.11}$$

The constant c is in $[x, f(a)]$. In order to do the expansion $(f^n)'$ must exist for the interval $[x, f(a)]$. As was deduced earlier (see 5.1.3), the 1^{st} derivative of f^n can be expressed as a product of $f'(f^i(x))$ where $f^i(x) \in I_S$, therefore it is enough to show that f' exists on I_S, which is the case for the example process (see (2.4.2)).

Thus analogously as for the 2nd derivative $dL(n, a)$ we have:

Lemma 5.2.5. *Let $A_0, A_1, \ldots, A_n$ be a Markov process as defined in 2.4, and let f' exist on S then for the linearisation error $dL(n, a)$, it holds that*

$$dL(n, a) \leq \sum_{x \in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot (f^n)'(c_x) \cdot (x - f(a))$$

where c_x denotes a constant in $[x, f(a)]$.

Proof. Using the Taylor expansion (5.2.11) in the definition of $dL(n,a)$ yields:

$$\begin{aligned} dL(n,a) &= \sum_{x\in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot \Big(f^{n+1}(a) + (f^n)'(c_x) \cdot \Big(x - f(a) \Big) - f^{n+1}(a) \\ &= \underbrace{\sum_{x\in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot f^{n+1}(a)}_{f^{n+1}(a)} \\ &\quad + \sum_{x\in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot (f^n)'(c_x) \cdot (x - f(a)) - f^{n+1}(a) \end{aligned}$$

□

Applying lemma 5.2.5 and for the difference $a \pm 1 - f(a)$ (see (5.2.1) and (5.2.2)) to the example process given in section 2.4 yields:

$$\begin{aligned} dL(n,a) &= \sum_{x\in S} \mathbf{prob}(A_1 = x | A_0 = a) \cdot (f^n)'(c_x) \cdot (x - f(a)) \\ &= \mathbf{prob}(A_{n+1} = a-1 | A_n = a) \cdot \Big((f^n)'(c_{-1}) \cdot \Big(a - 1 - f(a) \Big) \Big) \\ &\quad + \mathbf{prob}(A_{n+1} = a+1 | A_n = a) \cdot \Big((f^n)'(c_{+1}) \cdot \Big(a + 1 - f(a) \Big) \Big) \qquad (5.2.12) \\ &= -\mathbf{prob}(A_{n+1} = a-1 | A_n = a) \cdot \Big((f^n)'(c_{-1}) \cdot 2 \cdot \mathbf{prob}(A_{n+1} = a+1 | A_n = a) \Big) \\ &\quad + \mathbf{prob}(A_{n+1} = a+1 | A_n = a) \cdot \Big((f^n)'(c_{+1}) \cdot 2 \cdot \mathbf{prob}(A_{n+1} = a-1 | A_n = a) \Big) \\ &\leq 2 \cdot \underbrace{\mathbf{prob}(A_{n+1} = a-1 | A_n = a) \cdot \mathbf{prob}(A_{n+1} = a+1 | A_n = a)}_{\geq \frac{1}{4}} \cdot \\ &\quad ((f^n)'(a+1) - (f^n)'(a-1)) \\ &\leq \frac{1}{2} \cdot ((f^n)'(a+1) - (f^n)'(a-1)) \qquad (5.2.13) \end{aligned}$$

Using $c_{+1} = a+1$ and $c_{-1} = a-1$ to get an upper bound is doable due to the monotonicity of f' (see (2.4.2)) and the recurrence formula for $(f^n)'$ (see (5.1.3).

Substituting the recurrence formula for $(f^n)'$ gives:

$$\begin{aligned} & (f^n)'(x+1) - (f^n)'(x-1) \qquad (5.2.14) \\ &= f'(f^{n-1}(x+1)) \cdot \prod_{i=0}^{n-2} f'(f^i(x+1)) - f'(f^{n-1}(x-1)) \cdot \prod_{i=0}^{n-2} f'(f^i(x-1)) \\ &= f'(f^{n-1}(x+1)) \cdot \Big(\prod_{i=0}^{n-2} f'(f^i(x+1)) - \prod_{i=0}^{n-2} f'(f^i(x-1)) \Big) \\ &+ \Big(f'(f^{n-1}(x+1)) - f'(f^{n-1}(x-1)) \Big) \cdot \prod_{i=0}^{n-2} f'(f^i(x-1)) \\ &= \sum_{j=0}^{n-1} \Big(\prod_{i=j+1}^{n-2} f'(f^i(x+1)) \Big) \cdot \Big(f'(f^j(x+1)) - f'(f^j(x-1)) \Big) \cdot \prod_{k=0}^{j-1} f'(f^k(x-1)) \end{aligned}$$

As a next step, several stages for $f^i(x-1)$ are considered. Let the parameter R be greater than the fix point, e.g. $R > b_f$. Let $0 \leq h_1 \leq \ldots \leq h_K$, where $K = w - 2$. Then the stages for $f^i(x-1)$ become:

$$\begin{array}{lrcccl}
i = 0 \sim h_1 - 1 & (w-1)\cdot R & < & f^i(x-1) & \leq & w \cdot R \ (\leq N-1) \\
i = h_1 \sim h_2 - 1 & (w-2)\cdot R & < & f^i(x-1) & \leq & (w-1)\cdot R \\
i = h_2 \sim h_3 - 1 & (w-3)\cdot R & < & f^i(x-1) & \leq & (w-2)\cdot R \\
\vdots & & & \vdots & & \\
i = h_{K-1} \sim h_K - 1 & c_K \cdot R & < & f^i(x-1) & \leq & (c_K+1)\cdot R \\
i = h_K \sim h - 1 & 0 & \leq & f^i(x-1) & \leq & c_K \cdot R.
\end{array}$$

Note that if $x - 1 \leq c \cdot R$ for some integer $c < w$, then it follows $h_1 = h_2 = \cdots = h_{w-c} = 0$. Similarly if $c \cdot R < f^{n-1}(x-1) \leq (c+1) \cdot R$ for some integer $c \geq 2$, then we would have $h_{W-c+1} = \cdots = h_K = n$. Thus, these definitions are valid for all $x - 1$ and n. For each k, $0 \leq k \leq K$, let $c_k = w - k$ and let $s_k = h_{k+1} - h_k$ be the length of each stage.

Next observe that

$$\begin{aligned}
& f'(f^i(x+1)) - f'(f^i(x-1)) \\
=& \left(1 - \frac{2 \cdot w \cdot N}{(N + (w-1)\cdot f^i(x+1))^2})\right) - \left(1 - \frac{2 \cdot w \cdot N}{(N + (w-1)\cdot f^i(x-1))^2})\right) \\
=& \frac{2 \cdot w \cdot N \cdot (w-1) \cdot (f^i(x+1) - f^i(x-1)) \cdot \Big((w-1)\cdot(f^i(x+1) + f^i(x-1)) + 2 \cdot N\Big)}{(N + (w-1)\cdot f^i(x-1))^2 \cdot (N + (w-1)\cdot f^i(x+1))^2} \\
\leq& \frac{4 \cdot w \cdot N \cdot (w-1) \cdot (f^i(x+1) - f^i(x-1))}{(N + (w-1)\cdot f^i(x-1))^3} \\
\leq& \frac{8 \cdot w^2 \cdot N}{(N + (w-1)\cdot f^i(x-1))^3}
\end{aligned}$$

Consider any $i \in \{h_k, ..., h_{k+1} - 1\}$ such that $(c_k - 1) \cdot R < f^i(x-1) \leq c_k \cdot R$ holds with some $c_k \geq 3$. Then it follows, using the definition of R

$$\begin{aligned}
f'(f^i(x+1)) &= 1 - \frac{2 \cdot w \cdot N}{(N + (w-1)\cdot f^i(x+1))^2} \\
&\leq 1 - \frac{2 \cdot w \cdot N}{(N + (w-1)\cdot (c_k \cdot R + 2))^2} \\
&\leq 1 - \frac{2}{R \cdot (c_k + 1 + \frac{2}{R})^2} \\
&\leq 1 - \frac{2}{R \cdot (c_k + 3)^2}
\end{aligned}$$

The last bound holds because $R \geq b_f > 1$, since the weight by definition(see section 2.4) must be s.t. $1 \leq w \leq \frac{N}{2}$.

Later, it will be needed that (Recall $c_k \geq 2$).

$$\max_k \frac{1}{1 - \frac{2}{R \cdot (c_k+3)^2}} \leq \frac{25}{23}$$

and

$$
\begin{aligned}
f'(f^i(x+1)) - f'(f^i(x-1)) &\leq \frac{8 \cdot w^2 \cdot N}{(N + (w-1) \cdot f^i(x-1))^3} \\
&\leq \frac{8 \cdot w^2 \cdot N}{(N + (w-1) \cdot (c_k - 1) \cdot R)^3} \\
&\leq \frac{8}{R^2 \cdot (c_k - 1)^3}
\end{aligned}
$$

For any i such that $0 < f^i(x-1) \leq 2 \cdot R$, we have the same bounds as above with $c_k = 2$.

Using the last few inequalities a partial sum (for one stage) in (5.2.14) can be bounded as follows:

$$
\begin{aligned}
&\sum_{j=h_k}^{h_{k+1}-1} \left(\prod_{i=j+1}^{n-2} f'(f^i(x+1)) \right) \cdot \left(f'(f^j(x+1)) - f'(f^j(x-1)) \right) \cdot \prod_{k=0}^{j-1} f'(f^k(x-1)) \\
&\leq \frac{8}{R^2 \cdot (c_k - 1)^3} \cdot \sum_{j=h_k}^{h_{k+1}-1} \prod_{i=j+1}^{n-2} f'(f^i(x+1)) \cdot \prod_{k=0}^{j-1} f'(f^k(x-1)) \\
&\leq \frac{8}{R^2 \cdot (c_k - 1)^3} \cdot \sum_{j=h_k}^{h_{k+1}-1} \prod_{i=0, i \neq j}^{n-2} f'(f^i(x+1)) \\
&\leq \left(\max_k \frac{1}{1 - \frac{2}{R \cdot (c_k+3)^2}} \right) \cdot \frac{8}{R^2 \cdot (c_k - 1)^3} \cdot \sum_{j=h_k}^{h_{k+1}-1} \prod_{k=0}^{K} \left(1 - \frac{2}{R \cdot (c_k + 3)^2} \right)^{s_k} \\
&\leq \frac{9 \cdot s_k}{R^2 \cdot (c_k - 1)^3} \cdot \prod_{k=0}^{K} \left(1 - \frac{2}{R \cdot (c_k + 3)^2} \right)^{s_k}
\end{aligned}
$$

Note that for any h, the previous expression depends only on the choice of $s_0, ..., s_K$ and R and not on x.

Corollary 5.2.6. *For any x, $0 \leq x \leq N$, and for any $t \geq 0$, we have*

$$
f^t(x) \geq 2 \cdot R \implies f^{t+1}(x) < f^t(x) - \frac{1}{3}.
$$

Proof. For every $x \in [N, 2 \cdot b_f]$ at least $\frac{1}{3}$ of a ball is taken per step.

$$
f(x) - x = 1 - \frac{2 \cdot w \cdot x}{N + (w-1) \cdot x}
$$

This expression is minimal for the smallest possible value of $x \in [2 \cdot b_f, N]$

$$
\begin{aligned}
f(\frac{2 \cdot N}{w+1}) - \frac{2 \cdot N}{w+1} &= 1 - \frac{2 \cdot w \cdot \frac{2 \cdot N}{w+1}}{N + (w-1) \cdot \frac{2 \cdot N}{w+1}} \\
&\leq 1 - \frac{4 \cdot N \cdot \frac{w}{w+1}}{3 \cdot N} \leq -\frac{1}{3}
\end{aligned}
$$

□

We know from corollary 5.2.6 that each s_k is at most $3 \cdot R$ (except for s_k). Thus, it follows:

$$
\begin{aligned}
f'(f^i(x+1)) - f'(f^i(x-1)) &\le \sum_{k=0}^{K} \frac{9 \cdot s_k}{R^2 \cdot (c_k - 1)^3} \cdot \prod_{k=0}^{K}\left(1 - \frac{2}{R \cdot (c_k+3)^2}\right)^{s_k} \\
&\le \sum_{k=0}^{K-1} \frac{9 \cdot 3 \cdot R}{R^2 \cdot (c_k - 1)^3} \cdot \prod_{k=0}^{K}\left(1 - \frac{2}{R \cdot (c_k+3)^2}\right)^{s_k} \\
&\quad + \frac{9 \cdot s_K}{R^2 \cdot (c_K - 1)^3} \cdot \prod_{k=0}^{K}\left(1 - \frac{2}{R \cdot (c_k+3)^2}\right)^{s_k} \\
&\le \left(\frac{23}{25}\right)^{s_K} \cdot \sum_{k=0}^{K-1} \frac{27}{R \cdot (c_k - 1)^3} + \left(\frac{23}{25}\right)^{s_K} \cdot \frac{9 \cdot s_k}{R^2 \cdot 8} \\
&\le \left(\frac{23}{25}\right)^{s_K} \cdot \sum_{k=0}^{K-1} \frac{27}{R \cdot (c_k - 1)^3} + \left(\frac{23}{25}\right)^{s_K} \cdot \frac{9 \cdot s_k}{R^2 \cdot 8} \\
&\le \left(\frac{23}{25}\right)^{s_K} \cdot \frac{7}{R} + \left(\frac{23}{25}\right)^{s_K} \cdot \frac{9 \cdot s_k}{R^2 \cdot 8}
\end{aligned}
$$

where $s = h - 3 \cdot R \cdot w$, since $s_K \ge h - 3 \cdot R \cdot (w-2) \ge h - 3 \cdot R \cdot w$. The last bound follows from $2^{-3} + 3^{-3} + \cdots < \frac{1}{4}$.

Noting that $s \cdot (\frac{23}{25})^s$ is increasing for $s < 11.5 \cdot R$ and decreasing for $s \ge 11.5$. Therefore, we further have

$$
f'(f^n(x+1)) - f'(f^n(x-1)) \le \left(\frac{23}{25}\right)^s \cdot \frac{7}{R} + \frac{9}{8} \cdot \begin{cases} \frac{n \cdot (\frac{23}{25})^n}{R^2}, & \text{if } n < 11.5 \cdot R, \\ \frac{11.5R(\frac{23}{25})^{11.5R}}{R^2}, & \text{if } 11.5R \le h < 3R(w-2) + 11.5R, \text{ and} \\ \frac{t(\frac{23}{25})^t}{R^2}, & \text{if } 3R(w-2) + 11.5R \le h \ (t = h - 3wR) \end{cases}
$$

Finally the sum of the linearisation error over all time steps (and thus the error (see theorem 5.2.2)) is bounded by:

$$
\begin{aligned}
&err(n) \le \frac{1}{2} \cdot \sum_{i=0}^{\infty} \max_{x \in [1, N-1]} f'(f^i(x+1)) - f'(f^i(x-1)) \\
&\le \frac{1}{2} \cdot \Bigg(\sum_{h=0}^{3 \cdot R \cdot w - 1} \frac{7}{R} + \sum_{s=0}^{\infty} \left(\frac{23}{25}\right)^{s_K} \cdot \frac{7}{R} + \frac{9 \cdot (3 \cdot R \cdot (w-2) - 1) \cdot 11.5 \cdot R}{8 \cdot R^2} \cdot \left(\frac{23}{25}\right)^{11.5 \cdot R} \\
&+ \sum_{t=0}^{\infty} \left(\frac{23}{25}\right)^t \cdot \frac{9 \cdot t}{8 \cdot R^2} \Bigg) \\
&\le \frac{1}{2} \cdot \left(21 \cdot w + 88 + \frac{9}{8} \cdot \frac{R \cdot (3 \cdot R \cdot w)}{R^2} + \underbrace{\frac{9}{8} \cdot (3 \cdot 11.5 \cdot e^{-11.5/12.5})}_{<16} \cdot w + 176 \right) \\
&\le 21 \cdot w + 132
\end{aligned}
$$

This concludes the proof of theorem 5.2.4 □

5.3 Error bound in 2 dimensions

The two dimensional case turns out to be (not surprisingly) a straight forward extension of the one dimensional bound of section 5.1.

The linearisation error $\overrightarrow{dL}(n,a,b)$ for step n and state a is defined as:

$$\overrightarrow{dL}(n,a,b) = (dL_A(n,a,b), dL_B(n,a,b))^T$$

with

$$dL_A(n,a,b) := (\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot f_A^n(x,y)) - f_A^{n+1}(a,b)$$

and $dL_B(n,a,b)$ defined in the same manner. For the rest of the chapter only $dL_A(n,a,b)$ will be considered, since $dL_B(n,a,b)$ can be deduced in exactly the same way.

Since the pseudo expectation is correct for linear functions $\overrightarrow{f}$ (for a proof see (4.3.1), the error depends only on the non-linear part of the function f^n. The Taylor expansion of $\overrightarrow{f}^n$ at $\overrightarrow{f}(a,b)$ can be written as:

$$\begin{aligned} f_A^n(x,y) &= \underbrace{f_A^n(\overrightarrow{f}(a,b))}_{f_A^{n+1}(a,b),\text{ due to definition, see 2.3.2}} \\ &+\frac{\partial f_A^n}{\partial A}(\overrightarrow{f}(a,b))\cdot(x-f_A(a,b)) \\ &+\frac{\partial f_A^n}{\partial B}(\overrightarrow{f}(a,b))\cdot(y-f_B(a,b)) + R_{(x,y),(a,b)} \end{aligned} \quad (5.3.1)$$

The remainder $R_{(x,y),(a,b)}$ of the Taylor expansion is defined as:

$$\begin{aligned} R_{(x,y),(a,b)} &:= \frac{1}{2}\cdot\frac{\partial^2 f_A^n}{\partial A^2}(c_{(f_A(a,b),x)}, c_{(f_B(a,b),y)})\cdot(x-f_A(a,b)^2 \\ &+\frac{1}{2}\cdot\frac{\partial^2 f_A^n}{\partial B^2}(c_{(f_A(a,b),x)}, c_{(f_B(a,b),y)})\cdot(y-f_B(a,b)^2 \\ &+\frac{\partial^2 f_A^n}{\partial A\partial B}(c_{(f_A(a,b),x)}, c_{(f_B(a,b),y)})\cdot(x-f_A(a,b))\cdot(y-f_B(a,b)) \end{aligned}$$

The constants $c_{(f_A(a,b),x)}$ and $c_{(f_B(a,b),y)}$ are undetermined values in the interval $[f_A(a,b),x]$ and $[f_B(a,b),y]$ respectively.

In order to perform the Taylor expansion, f_A^n and f_B^n must be continuous on I_{S_A} and I_{S_B} respectively and as well as their first partial derivatives and the 2nd order partial derivatives have to exist in this interval. As will be deduced later, all partial derivatives (of 1st and 2nd order) of f_A^n and f_B^n can be expressed as a product of partial derivatives of f_A and f_B. Therefore it is enough to show that the first order partial derivatives are continuous and the 2^{nd} order partial derivatives exist on the considered interval.

Remark 5.3.1.
Of course, a Taylor expansion of a higher (or lower) order can be considered as well (given that f_A and f_B fulfill the requirements to do so).

Theorem 5.3.1. *Let $(A_0, B_0), (A_1, B_1), \ldots, (A_n, B_n)$ be a Markov process as defined in 2.2.2, and let f' be continuous on S and let f'' exist on S then for the linearisation error $dL_A(n, a, b)$ holds that*

$$dL_A(n,a,b) \leq \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot R_{(x,y),(a,b)}$$

Proof. Using the Taylor expansion (5.3.1) in the definition of $dL(n, a, b)$ yields:

$$\begin{aligned}
dL_A(n,a,b) &= (\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot f_A^n(x,y)) - f_A^{n+1}(a,b) \\
&= \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot \Bigg(f_A^{n+1}(a,b) \\
&\quad + \frac{\partial f_A^n}{\partial A}(\overrightarrow{f}(a,b)) \cdot (x - f_A(a,b)) + \frac{\partial f_A^n}{\partial B}(\overrightarrow{f}(a,b)) \cdot (y - f_B(a,b)) \\
&\quad + R_{(x,y),(a,b)} \Bigg) - {f_A}^{n+1}(a,b) \\
&= \underbrace{\frac{\partial f_A^n}{\partial A}(\overrightarrow{f}(a,b)) \cdot \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot (x - f_A(a,b))}_{=0} \\
&\quad \underbrace{+ \frac{\partial f_A^n}{\partial B}(\overrightarrow{f}(a,b)) \cdot \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot (y - f_B(a,b))}_{=0} \\
&\quad + \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot R_{(x,y),(a,b)} \\
&= \sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot R_{(x,y),(a,b)}
\end{aligned}$$

□

Lemma 5.3.2. *Let $(A_0, B_0), (A_1, B_1), \ldots, (A_n, B_n)$ be a Markov process as defined in 2.2.2 then for the linearisation error $dL_A(n, a, b)$ holds*

$$err_A(n,a,b) \leq \sum_{t=1}^{n-1} \max_{(x,y)\in S} dL_A(t,x,y)$$

Proof.

$$\begin{aligned}
&err_A(n+1,a,b) = \mathbf{E}_A[A_{n+1},B_{n+1}|(A_0,B_0)=(a,b)] - f_A^{n+1}(a,b)\\
&= (\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot \mathbf{E}_A[A_n,B_n|(A_1,B_1)=(x,y)]) - f_A^{n+1}(a,b)\\
&= (\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot \mathbf{E}_A[A_n,B_n|(A_1,B_1)=(x,y)] - f_A^{n}(x,y))\\
&+(\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot f_A^{n}(x,y)) - f_A^{n+1}(a,b)\\
&= (\sum_{(x,y)\in S} \mathbf{prob}((A_1,B_1)=(x,y)|(A_0,B_0)=(a,b)) \cdot err_A(n,x,y) + dL_A(n,a,b)\\
&\leq \max_{(x,y)\in S} err_A(n,x,y) + dL_A(n,a,b)\\
&\leq \sum_{t=2}^{n} \max_{(x,y)\in S} dL_A(t-1,x,y) \qquad (5.3.2)
\end{aligned}$$

Observe that the error for the first step is 0 due to definition of f_A (see (2.3.5)). For that reason the summation starts from 2. □

Next a recurrence formula for computing the first order partial derivative $\frac{\partial f_A^n(a,b)}{\partial A}$ is given. For $n \geq 1$ it is true that:

$$\begin{aligned}
\frac{\partial f_A^n(a,b)}{\partial A} &= \frac{\partial f_A(f^{n-1}(a,b))}{\partial A}\\
&= \frac{\partial f_A}{\partial A}(f^{n-1}(a,b)) \cdot \frac{\partial f_A^{n-1}(a,b)}{\partial A} + \frac{\partial f_A}{\partial B}(f^{n-1}(a,b)) \cdot \frac{\partial f_B^{n-1}(a,b)}{\partial A}
\end{aligned}$$

and in the same way for $\frac{\partial f_B^n(a,b)}{\partial A}$:

$$\begin{aligned}
\frac{\partial f_B^n(a,b)}{\partial A} &= \frac{\partial f_A(f^{n-1}(a,b))}{\partial B}\\
&= \frac{\partial f_B}{\partial A}(f^{n-1}(a,b)) \cdot \frac{\partial f_A^{n-1}(a,b)}{\partial B} + \frac{\partial f_B}{\partial B}(f^{n-1}(a,b)) \cdot \frac{\partial f_B^{n-1}(a,b)}{\partial B}
\end{aligned}$$

The other first order partial derivatives follow the same pattern.

Next a recurrence formula for computing the second order partial derivative $\frac{\partial^2 f_B^n(a,b)}{\partial A^2}$ is given. For $n \geq 1$ it is true that:

$$\begin{aligned}
\frac{\partial^2 f_B^n(a,b)}{\partial A^2} &= \frac{\partial\left(\frac{\partial f_B}{\partial A}(f^{n-1}(a,b)) \cdot \frac{\partial f_A^{n-1}(a,b)}{\partial B} + \frac{\partial f_B}{\partial B}(f^{n-1}(a,b)) \cdot \frac{\partial f_B^{n-1}(a,b)}{\partial B}\right)}{\partial A}\\
&= \frac{\partial f_B}{\partial A}(f^{n-1}(a,b)) \cdot \frac{\partial^2 f_A^{n-1}(a,b)}{\partial A^2} + \frac{\partial f_B}{\partial B}(f^{n-1}(a,b)) \cdot \frac{\partial^2 f_B^{n-1}(a,b)}{\partial A^2}\\
&+\frac{\partial^2 f_B}{\partial A^2}(f^{n-1}(a,b)) \cdot \left(\frac{\partial f_A^{n-1}(a,b)}{\partial A}\right)^2 + \frac{\partial^2 f_B}{\partial B^2}(f^{n-1}(a,b)) \cdot \left(\frac{\partial f_B^{n-1}(a,b)}{\partial A}\right)^2\\
&+2 \cdot \frac{\partial f_A^{n-1}(a,b)}{\partial A} \cdot \frac{\partial f_B^{n-1}(a,b)}{\partial B} \frac{\partial f_B}{\partial A \partial B}(f^{n-1}(a,b))
\end{aligned}$$

Recurrence formulas for the remaining partial derivatives can be obtained in the same way.

Bibliography

[1] J.R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.

[2] M.A. Berger: *An Introduction to Probability and Stochastic Processes*, Springer Press, 1993.

[3] J. Medhi: *Stochastic Processes*, Wiley Eastern Ltd., 1982.

[4] D. Zeilberger: *A=B*, http://www.math.upenn.edu/%7Ewilf/AeqB.pdf, 1997.

[5] R.P. Agarwal: *Fixed Point Theory and Applications*, Cambridge University Press, 2001

[6] Y. Niikura and J. Schneider, and O. Watanabe, An approximation analysis of simple Markov processes, in *Proc. SIGAL Workshop*, IPSJ AL94-11, 2004.

[7] J. Schneider and O. Watanabe, An approximate analysis of Markov processes and its application to analysis of randomized algorithms, in *Proc. COMP Workshop*, IEICE COMP May,2004.

[8] H. Takahashi and Y. Niikura, An extension of Azuma-Hoeffding inequalities and its application to an analysis of randomized local search algorithms, in *Proc. 6th IBIS Workshop*, 177–181, 2003.

[9] O. Watanabe, T. Sawai, and H. Takahashi, Analysis of a randomized local search algorithm for LDPCC decoding problem, in *Proc. SAGA'03*, Lecture Notes in Comp. Sci. 2827,50–60, 2003.