Technical Reports on Mathematical and Computing Sciences: TR-C195
title: An Error Bound for the Pseudo Expectation: 1D Case
author: Osamu Watanabe and Johannes Schneider
affiliation:
Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology
W8-25, Tokyo 152-8552
email: watanabe@is.titech.ac.jp
acknowledgments to financial supports:
Supported in part by a Grant-in-Aid for Scientific Research on Priority Areas "Statical-Mechanical Approach to Probabilistic Information Processing" 2002-2005.

Abstract. In [WSA03], the notion of *pseudo expectation* has been proposed for analyzing relatively simple Markov processes, which would be often seen as simple execution models of local search algorithms. In this note, we consider the simplest case — one dimensional Markov process — and investigate the approximation error bound of its pseudo expectation.

1 Preliminaries

Consider any Markov process M defined on a state space S. Then by definition, there must be some function f_M on S that satisfies

$$\mathrm{E}[\, s_{t+1} \,|\, s_t = s \,] \;=\; f_M(s)$$

for all $s \in S$. Then the *pseudo expectation* for M is simply defined by

$$\widehat{\mathrm{E}}_M(s) \;=\; f_M^t(s).$$

It has been proposed [WSA03] to use this pseudo expectation for approximating the real expectation $\mathrm{E}[\, s_t \,|\, s_0 = s\,]$, in particular, when analyzing Markov processes defined as simple execution models of local search algorithms.

Here we discuss its approximation error for one simple Markov process. The Markov process we will consider is defined as a probabilistic model for the simple "random ball exchange".

Random Ball Exchange

Alice and Bob has A and B balls respectively, total $N = A + B$. At each step, either from Alice to Bob, or from Bob to Alice, one ball is passed with the following probability.

$$\Pr[\,\text{Alice} \to \text{Bob}\,] \;=\; p(A) \;=\; \frac{WA}{WA + (N-A)} \;=\; \frac{WA}{N + (W-1)A}, \quad \text{and}$$
$$\Pr[\,\text{Alice} \to \text{Bob}\,] \;=\; q(A) \;=\; 1 - p(A) \;=\; \frac{N-A}{N + (W-1)A}.$$

Here W is some constant, which we called a *weight*. (Though it is a constant for the process, it may depend on N when we define the Markov process.) Without loss of generality, we may assume $W \geq 1$; otherwise, consider a weight on B, the number of Bob's balls, which should be then larger than 1.

Since the total number of balls $A + B$ does not change, we may express the execution of this process by, say, the number of balls of Alice. That is, a state can be expressed by a one number, namely, A; in this sense, it is a one dimensional Markov process. Let A_t denote the number of Alice's balls at (i.e., just after) the tth step.

Define $f(x)$ so that $f(A_t) = \mathrm{E}[A_{t+1}|A_t]$. More specifically, we define $f(x)$ as follows.

$$f(x) \;=\; p(x)(x+1) + q(x)(x-1) \;=\; x + 1 - \frac{2Wx}{N + (W-1)x}.$$

On the other hand, the pseudo expectation of this process is defined as follows.

$$\widehat{\mathrm{E}}(t,x) \;=\; f^t(x).$$

Our goal is to give some appropriate upper bound for it approximation error, which is defined as follows. (Throughout this note, we will use a to denote a state, often the initial state, and we assume that $a \in \{0, ..., N\}$. Also t will be used for a step, and it is assumed nonnegative integer.)

$$\mathit{err}(t,a) \;=\; |\mathrm{E}[\,A_t \,|\, A_0 = a\,] - \widehat{\mathrm{E}}(t,a)|.$$

Let us see here some basic facts on this process and the error function. Note first that $p(0) = 0$ and $p(N) = 1$ (hence, $f(0) = 1$ and $f(N) = -1$). Thus, A_t takes values from 0 to N (and so is $f(x)$ if $x \in [0, N]$). Thus, we have an obvious bound $\mathit{err}(t,a) \leq N$. What we would like to show is something significantly smaller than N; for example, a constant independent from N (nor W).

We also point here the following facts. That is, the pseudo expectation is accurate for $t \leq 1$ (by definition) and $W = 1$.

Fact 1. For any a, we have $\mathit{err}(0,a) = \mathit{err}(1,a) = 0$.

Fact 2. If $W = 1$, then $\mathit{err}(t,a) = 0$ for any t and a.

Since our process is simple, we can give a precise formula for its stationary distribution[1] as follows.

[1] This Markov process is periodic with period 2. Thus, no stationary distribution exists. In the following, we consider states at even and odd steps separately. By the "average stationary state", we mean the arithmetic mean of the expectations for even and odd steps.

Proposition 1. Consider the process starting from $A_0 = a$. Then we have

$$\lim_{h\to\infty} \Pr[\, A_{2h} = m\,] \quad = \quad \begin{cases} 2P_\infty, & \text{if both } a+m \text{ is even, and} \\ 0, & \text{otherwise.} \end{cases}$$
$$\lim_{h\to\infty} \Pr[\, A_{2h+1} = m\,] \quad = \quad \begin{cases} 2P_\infty, & \text{if both } a+m \text{ is odd, and} \\ 0, & \text{otherwise.} \end{cases}$$

Where

$$P_\infty \;=\; \frac{1}{2}\cdot\frac{(W^m\cdot N + W^{m-1}\cdot m\cdot(1-W))}{N\cdot(1+W)^{N-1}}\cdot\binom{N}{m}.$$

The average stationary distribution is then calculated as follows. (Below we will often simply write, e.g., $\mathrm{E}[A_t|a]$ for $\mathrm{E}[A_t|A_0 = a]$.)

Corollary 2.

$$a_{\lim} \;\stackrel{\text{def}}{=}\; \frac{\lim \mathrm{E}[A_{2h}|a] + \mathrm{E}[A_{2h+1}|a]}{2} \;=\; \frac{N + (W-1)/2}{W+1}.$$

On the other hand, the fixed point a_{fix} of f is calculated as follows; thus, $\lim_{t\to\infty} f^t(a) = a_{\text{fix}}$ for any a.

Proposition 3.

$$a_{\text{fix}} \;=\; \frac{N}{W+1}.$$

Hence, in the limit, the error of the pseudo expectation is

$$\lim_{t\to\infty} err(t,a) \;=\; a_{\lim} - a_{\text{fix}} \;=\; \frac{1}{2} - \frac{1}{W+1}.$$

But unfortunately, $err(t,a)$ is not monotone in general, and this is not a general upper bound.

2 An Error Bound

In the following, we will prove the following bound for $N \geq W$.

Theorem 4. Assume that $N \geq W$. Then there are some constants c_1 and c_2 independent from N and W such that the following holds for any t and a.

$$err(t,a) \;\stackrel{\text{def}}{=}\; \mathrm{E}[A_t|A_0 = a] - f^t(a) \;\leq\; c_1 W + c_2.$$

Recall that no error occurs in the case that $W = 1$, which is because, intuitively, the function f becomes "linear". Our approach is to estimate the nonlinearity of f^t. More specifically, we introduce the following function, which we call *nonlinearity* of f^t at x.

$$dL(t,x) \;=\; p(x)f^t(x_-) + q(x)f^t(x_+) - f^t\Big(p(x)(x_-) + q(x)(x_+)\Big),$$

Here and throughout this note, we use the following notations.

$$t_- = t-1, \quad t_+ = t+1, \quad x_- = x-1, \quad x_+ = x+1.$$

To make this definition valid at the boundaries, we let $f^t(-1) = f^t(N+1) = 0$. In fact, it is easy to see that $dL(t,0) = dL(t,N) = 0$.

The following relation, though easy to show, by induction, plays a key role in our analysis.

Lemma 5. For any $t \geq 1$ and any x, $0 \leq x \leq N$,

$$err(t,x) \;=\; p(x)err(t_-,x_-) + q(x)err(t_-,x_+) + dL(t_-,x).$$

Here and in the following, we let $err(t,-1) = err(t,N+1) = 0$.

From this lemma, we have the following bound:

$$err(t,x) \;\leq\; \sum_{1\leq h\leq t-1} \max_{x\in[N]} dL(h,x). \tag{1}$$

Recall that $dL(h,0) = dL(h,N) = 0$; hence, for the maximum, we only have to consider the range $\{1,...,N-1\}$. Thus, in the following, by $\max_x dL(h,x)$ we mean $\max_{x\in\{1,...,N-1\}} dL(h,x)$. In more general, $dL(h,x)$ is estimated not only integral x; in such a case, we assume that x is taken from $(0,N)$.

We begin with some basic properties of our f^t.

Fact 3. For any $t \geq 1$, the first derivative $(f^t)'(x)$ is computed inductively as follows.

$$(f^t)'(x) \;=\; \prod_{0\leq i\leq t-1}\left(1 - \frac{2WN}{(N+(W-1)f^i(x))^2}\right).$$

In the following we let f^0 be the identity function, i.e., $f^0(x) = x$.

Fact 4. For any t, the following holds:
(1) both $f^t(x)$ and $(f^t)'(x)$ are monotone and increasing,
(2) $0 \leq f^t(x) \leq x$ and $0 \leq (f^t)'(x) \leq 1$.

From the monotonicity of $(f^t)'(x)$, the function $f^t(x)$ must be concave, which implies the following properties.

Fact 5. The nonlinearity $dL(h,x)$ is nonnegative for any $h \geq 0$ and x. Thus, $\mathrm{E}[A_t|a] \geq f^t(a)$ for any t and a.

Let us replace $dL(h,x)$ with a function that is easier to compute. As one can easily expect, the nonlinearity $dL(h,x)$ can be bounded by the second derivative. But here by using the monotonicity of f^h and $(f^t)'$, we can also prove the following bound.

Lemma 6. For any $h \geq 0$ and x, we have

$$dL(h,x) \;\leq\; \frac{1}{2}((f^h)'(x_+) - (f^h)'(x_-)).$$

Proof. For simplicity, let us use g to denote f^h here. By using the fact that g' is monotone and increasing, we have

$$\begin{array}{rclcl} g(p(x)x_- + q(x)x_+) &=& g(x_- + 2q(x_-)) &\geq& g(x_-) + 2q(x)g'(x_-), \quad \text{and} \\ g(p(x)x_- + q(x)x_+) &=& g(x_+ - 2p(x_+)) &\geq& g(x_+) - 2p(x)g'(x_+). \end{array}$$

Hence,

$$\begin{array}{rcl} dL(h,x) &=& p(x)g(x_-) + q(x)g(x_+) - g(p(x)(x_-) + q(x)(x_+)) \\ &\leq& p(x)g(x_-) + q(x)g(x_+) \\ && \quad - \Big(\, p(x)\Big(g(x_-) + 2q(x)g'(x_-)\Big) + q(x)\Big(g(x_+) - 2p(x)g'(x_+)\Big)\,\Big) \\ &=& 2p(x)q(x)(g'(x_+) - g'(x_-)) \;\leq\; \dfrac{1}{2}(g'(x_+) - g'(x_-)). \end{array}$$

□

Thus, in the following define $df(h,x) = (f^h)'(x_+) - (f^h)'(x_-)$, and we will bound $\max_x df(h,x)$. Again unless explicitly stated, we will assume that $x \in \{1, ..., N-1\}$, or $x \in (0,N)$ in general. Also h will be a nonnegative integer.

Consider first $df(0,x)$ and $df(1,x)$. We have

$$\begin{array}{rcl} df(0,x) &=& (f^0)'(x+1) - (f^0)'(x-1) \;=\; 0, \quad \text{and} \\ df(1,x) &=& f'(x+1) - f'(x-1) \\ &=& \left(1 - \dfrac{2WN}{(N+(W-1)(x+1))^2}\right) - \left(1 - \dfrac{2WN}{(N+(W-1)(x-1))^2}\right) \\ &\leq& \dfrac{4(W-1)WN\Delta_0}{(N+(W-1)(x-1))^3}, \end{array}$$

where $\Delta_0 = (x+1) - (x-1) = 2$. In general, we have the following bound for any $h \geq 1$.

$$df(h,x) \;\leq\; \alpha_+(h-1)\Big(df(h-1,x) + \beta(h-1)\cdot P(h-2)\Big), \tag{2}$$

where

$$
\begin{aligned}
\alpha_+(i) &= 1 - \frac{2WN}{(N+(W-1)f^i(x+1))^2}, \quad \alpha_-(i) \;=\; 1 - \frac{2WN}{(N+(W-1)f^i(x-1))^2},\\
\beta(i) &= \frac{4(W-1)WN\Delta_i}{(N+(W-1)f^i(x-1))^3} \text{ (where } \Delta_i = f^i(x+1) - f^i(x-1)\text{), and}\\
P(i) &= \prod_{j=0}^{i} \alpha_-(j).
\end{aligned}
$$

Note that the product $\prod_{j=0}^{-1} \cdots$ is 1 as usual.

Here let us define $u(h,x)$ for $h \geq 0$ as follows. (Define $u(0,x) = 0$.)

$$
u(h,x) \;=\; \alpha_+(h-1)\left(u(h-1) + \beta'(h-1)\cdot P(h-2)\right),
$$

where $\beta'(i)$ is defined a bit more simply as

$$
\beta'(i) \;=\; \frac{8W^2N}{(N+(W-1)f^i(x-1))^3}.
$$

Then it is easy to see that $u(h,x)$ bounds $df(h,x)$ and that it has the following closed form.

$$
u(h,x) \;=\; \sum_{i=0}^{h-1}\left(\prod_{j=i+1}^{h-1}\alpha_+(j)\right)\beta'(i)P(i-1). \tag{3}
$$

We will discuss below a bound for this $u(h,x)$.

Consider any x and h, and let them be fixed for a while. Now we introduce an important parameter $R = N/W$, which is larger than equal to 1 when $W \leq N$ (as assumed in our theorem). Recall that $a_{\text{fix}} = N/(W+1)$ ($\approx R$) is the fixed point of f; hence, $f^i(x_-)$ gets decreased if x_- is much larger than R, say, $x_- > 2R$. A technical key of our analysis is to bound terms in $u(h,x)$ by stages depending on $f^i(x_-)$ w.r.t. R. More specifically, we consider the following K stages, where $h_0 = 0$ and $K = W-2$.

$$
\begin{array}{llcccl}
i = h_0 \sim h_1 - 1 & (W-1)R & < & f^i(x_-) & \leq & WR\ (=N)\\
i = h_1 \sim h_2 - 1 & (W-2)R & < & f^i(x_-) & \leq & (W-1)R\\
i = h_2 \sim h_3 - 1 & (W-3)R & < & f^i(x_-) & \leq & (W-2)R\\
\vdots & & & \vdots & & \\
i = h_{K-1} \sim h_K - 1 & 2R & < & f^i(x_-) & \leq & 3R\\
i = h_K \sim h - 1 & 0 & \leq & f^i(x_-) & \leq & 2R.
\end{array}
$$

Since $f^i(x_-)$ gets decreased if $x_- > 2R$, the sequence $h_0 \leq h_1 \leq \cdots \leq h_K$ can be defined. Note that if $x_- \leq cR$ for some integer $c < W$, then we would have $h_1 = h_2 = \cdots = h_{W-c} = 0$. Similarly, if $cR < f^{h-1}(x_-) \leq (c+1)R$ for some integer $c \geq 2$, then we would have $h_{W-c} = h_{W-c+1} = \cdots = h_K = h$. Thus, these definitions are valid for all x_- and h.

For each k, $0 \le k \le K$, let $c_k = W - k$ and $s_k = h_{k+1} - h_k$; $s_k \ge 0$ is the length of each stage.

Consider any $i \in \{h_k, ..., h_{k+1}-1\}$ such that $(c_k - 1)R < f^i(x_-) \le c_k R$ holds with some $c_k \ge 3$. Then we have

$$\begin{aligned}
\alpha_+(i) &= 1 - \frac{2WN}{(N+(W-1)f^i(x+1))^2} \le 1 - \frac{2WN}{(N+(W-1)(c_k R+2))^2} \\
&\le 1 - \frac{2RW^2}{R^2W^2(c_k+1+2/R)^2} = 1 - \frac{2}{R(c_k+1+2/R)^2}, \quad \text{and} \\
\beta'(i) &= \frac{8W^2N}{(N+(W-1)f^i(x-1))^3} \le \frac{8W^2N}{(N+(W-1)(c_k-1)R)^3} \\
&\le \frac{8RW^3}{(c_k-1)^3R^3W^3} = \frac{8}{(c_k-1)^3R^2}.
\end{aligned}$$

For any i such that $0 \ge f^i(x_-) \le 2R$, we have the same bounds as above with $c_K = 2$.

Here we define

$$\gamma(c) = 1 - \frac{2}{(R(c+1+2/R)^2)} \le 1 - \frac{2}{(R(c+3)^2)}.$$

The bound above holds if $R \ge 1$. In the following discussion, it is essential that $\gamma(c) < 1$ for any $c \ge 2$, and for this, it is enough that $R \ge 1$. (Here we need the assumption that $W \le N$.)

By using the above bounds for α_+ and β', we can bound a partial sum in (3) as follows.

$$\begin{aligned}
&\sum_{i=h_k}^{h_{k+1}-1} \left(\prod_{j=i+1}^{h-1} \alpha_+(j) \right) \beta'(i) P(i-1) = \sum_{i=h_k}^{h_{k+1}-1} \beta'(i) \prod_{j=i+1}^{h-1} \alpha_+(j) \prod_{j=0}^{i-1} \alpha_-(j) \\
&\le \frac{8}{(c_k-1)^3R^2} \sum_{i=h_k}^{h_{k+1}-1} \prod_{j=i+1}^{h-1} \alpha_+(j) \prod_{j=h_k}^{i-1} \alpha_+(j) = \frac{8s_k}{(c_k-1)^3R^2} \sum_{i=h_k}^{h_{k+1}-1} \prod_{j=0\,(j\neq i)}^{h-1} \alpha_+(j) \\
&\le \frac{8s_k}{(c_k-1)^3R^2} \gamma(c_0)^{s_0} \cdots \gamma(c_{k-1})^{s_{k-1}} \gamma(c_k)^{s_k-1} \gamma(c_{k+1})^{s_{k+1}} \cdots \gamma(c_K)^{s_K}.
\end{aligned}$$

Define $\widehat{u}(h,x)$ by

$$\widehat{u}(h,x) = \sum_{k=0}^{K} \frac{8s_k}{(c_k-1)^3R^2} \Gamma, \quad \text{where} \quad \Gamma = \prod_{k=0}^{K} \gamma(c_k)^{s_k}.$$

Then noting that $\gamma(2) \le \gamma(c_K) < \gamma(c_{K-1}) \cdots$, it follows from our discussion that

$$u(h,x) \le (\min_k \gamma(c_k))^{-1} \cdot \widehat{u}(h,x) = \gamma(2)^{-1} \cdot \widehat{u}(h,x) \le \frac{25}{23} \cdot \widehat{u}(h,x). \tag{4}$$

Here we need some knowledge on the convergence speed of f^t. Though a bit more detail analysis will be given later, the following simple bound is enough at this moment.

Fact 6. For any $x \in (0, N)$ and for any t, we have

$$f^t(x) \geq 2R \;\Rightarrow\; f^{t+1}(x) < f^t(x) - 1/3.$$

Note that for any h, the expressions defining $\widehat{u}(h,x)$ and $\widehat{u}(h,x')$ differ only on the choice of $s_0, ..., s_K$. On the other hand, we know from the above fact that each s_k is at most $3R$ (except for s_K). Thus, we have

$$\begin{aligned}
\widehat{u}(h,x) &= \sum_{k=0}^{K} \frac{8s_k}{(c_k-1)^3R^2} \prod_{k=0}^{K} \gamma(c_k)^{s_k} \;\leq\; \sum_{k=0}^{K-1} \frac{8\cdot 3R}{(c_k-1)^3R^2} \prod_{k=0}^{K} \gamma(c_k)^{s_k} + \frac{8s_K}{(c_K-1)^3R^2} \prod_{k=0}^{K} \gamma(c_k)^{s_k} \\
&\leq \gamma(2)^{s_K} \sum_{k=0}^{K-1} \frac{24}{(c_k-1)^3R} + \gamma(2)^{s_K}\frac{s_K}{R^2} \;\leq\; \gamma(2)^{s} \sum_{k=0}^{K-1} \frac{24}{(c_k-1)^3R} + \frac{s_K\gamma(2)^{s_K}}{R^2} \\
&\leq \frac{6\gamma(2)^s}{R} + \frac{s_K\gamma(2)^{s_K}}{R^2},
\end{aligned}$$

where $s = h - 3RW$, since $s_K \geq h - 3R(W-2) \geq h - 3RW$. The last bound is from the fact that $2^{-3} + 3^{-3} + \cdots < 1/4$. Noting that $v(s) = s\gamma(2)^s$ is increasing for $s < 11.5R$ and decreasing for $s \geq 11.5$, we further have

$$\widehat{u}(h,x) \;\leq\; \frac{6\gamma(2)^s}{R} + \begin{cases} \dfrac{h\gamma(2)^h}{R^2}, & \text{if } h < 11.5R, \\ \dfrac{11.5R\gamma(2)^{11.5R}}{R^2}, & \text{if } 11.5R \leq h < (3R(W-2) + 11.5R, \text{ and} \\ \dfrac{t\gamma(2)^t}{R^2}, & \text{if } h \leq 3WR + 11.5R \text{ (here let } t = h - 3WR). \end{cases}$$

Therefore, we finally have the following desired bound.

$$\begin{aligned}
\textstyle\sum_{h\geq 0} \max_x \widehat{u}(h,x) &\leq \sum_{h=0}^{3RW-1} \frac{6}{R} + \sum_{s\geq 0} \frac{6\gamma(2)^s}{R} + \frac{(3R(W-2)+1)(11.5R)\gamma(2)^{11.5R}}{R^2} + \sum_{t\geq 0} \frac{t\gamma(2)^t}{R^2} \\
&\leq 18W + 75 + (3\cdot 11.5\cdot e^{-11.5/12.5})W + \frac{625}{4} \;\leq\; 18W + 14W + 75 + 157 \\
&= 32W + 232.
\end{aligned}$$

2.1 A Bound for a Large Weight W

Theorem 4 gives a good bound when W is regarded as a constant. On the other hand, it would not be satisfiable for large W; in fact, our above analysis is not applicable for the case $W > N$. On the other hand, for the case that W is proportional to N, *some* bound is provable by a much simpler argument.

Consider, for example, the case that $W = N$. Note first that $a_{\text{lim}} \approx a_{\text{fix}} \approx N/W = 1$; that is, from any state, the process converges to 1. In fact, we have

$$p(x) \;=\; \frac{Wx}{(N+(W-1)x)} \;\approx\; \frac{Nx}{N(1+x)} \;=\; \frac{x}{1+x}.$$

Hence $p(x)$ is close to 1 for large x; this means that for such large x, x gets decreased by 1 almost deterministically. Similarly, since we have

$$f^t(x) \;=\; f^{t-1}(x) - 1 + \frac{2W f^{t-1}(x)}{N + (W-1)f^{t-1}(x)} \;\approx\; f^{t-1}(x) - 1 + \frac{2}{1 + f^{t-1}(x)},$$

$f^t(x)$ gets decreased almost by 1 from $f^{t-1}(x)$ so long as $f^{t-1}(x)$ is large. Thus, in such a situation that we may assume that A_t is large, both the real expectation $\mathrm{E}[A_t|a]$ and the pseudo expectation $f^t(a)$ can be approximated as $a - t$, and hence, only small difference would be expected between them. This observation leads us to the following bound.

Theorem 7. Let $\alpha = (W-1)/N > 0$. Then there exists a constant c_3 independent from N and W such that the following holds for any t and a.

$$\mathit{err}(t,a) \;\le\; \frac{c_3}{\sqrt{\alpha}} \cdot \sqrt{N}.$$

Proof. We may assume that $\alpha > 1/N$; otherwise, we can assume that W is small, and Theorem 4 is applicable. Note that both a_{fix} and a_{lim}, which can be approximated as $1/\alpha$, are much less than $\sqrt{N/\alpha}$ $(= \sqrt{\alpha N} \cdot (1/\alpha))$. Hence, for any initial state $a \le \sqrt{N/\alpha}$, and for any t, we certainly have both $\mathrm{E}[A_t|a] \le \sqrt{N/\alpha}$ and $f^t(a) \le \sqrt{N/\alpha}$; that is, the bound trivially holds with $c_3 = 1$. Thus, we only need to consider the case that the process starts from some $a > \sqrt{N/\alpha}$.

Consider the following simple Markov process $\{X_t\}_{t\ge 0}$:

$$\begin{array}{l} \text{for any } x > \sqrt{N/\alpha}, \\ \quad \Pr[\, X_t = X_{t-1} - 1 \,|\, X_{t-1} = x \,\} \;=\; p_0 \;\stackrel{\mathrm{def}}{=}\; \dfrac{\sqrt{\alpha N}}{1+\sqrt{\alpha N}}, \\ \quad \Pr[\, X_t = X_{t-1} + 1 \,|\, X_{t-1} = x \,\} \;=\; 1 - p_0 \;=\; \dfrac{1}{1+\sqrt{\alpha N}}; \;\text{ and} \\ \text{for any } x \le \sqrt{N/\alpha}, \\ \quad \Pr[\, X_t = X_{t-1} \,|\, X_{t-1} = x \,\} \;=\; 1. \end{array}$$

Note that for any $x > \sqrt{N/\alpha}$, we have

$$\Bigl(\; \Pr[\, A_t = A_{t-1} - 1 \,|\, A_{t-1} = x \,] \;=\; \Bigr) \;=\; p(x) \;=\; \frac{Wx}{N + (W-1)x} \;\ge\; \frac{\alpha x}{1 + \alpha x} \;>\; \frac{\sqrt{\alpha N}}{1+\sqrt{\alpha N}}.$$

Hence it holds that

$$\mathrm{E}[\, X_t \,|\, a \,] \;\ge\; \mathrm{E}[\, A_t \,|\, a \,] \;\ge\; f^t(a).$$

On the other hand, since

$$\mathrm{E}[\, X_t \,|\, X_{t-1} = x \,] \;=\; x - \frac{\sqrt{\alpha N}}{1+\sqrt{\alpha N}} + \frac{1}{1+\sqrt{\alpha N}} \;=\; x - 1 + \frac{2}{1+\sqrt{\alpha N}} \;\le\; x - 1 + \frac{2}{\sqrt{\alpha N}}$$

and $f^t(x) \geq f^{t-1}(x) - 1$, we have

$$a - t + \frac{2t}{\sqrt{\alpha N}} \;\geq\; \mathrm{E}[\, X_t \,|\, a \,] \;\geq\; \mathrm{E}[\, A_t \,|\, a \,] \;\geq\; f^t(a) \;\geq\; a - t.$$

This implies the following bound for any $t \leq N$.

$$err(t, a) \;=\; \mathrm{E}[\, A_t \,|\, a \,] - f^t(a) \;\leq\; \frac{2N}{\sqrt{\alpha N}} \;\leq\; 2\sqrt{\frac{N}{\alpha}}.$$

On the other hand, if $t > N$, the above error bound is clearly satisfied because both $\mathrm{E}[X_t|a]$ and $f^t(x)$ are smaller than $\sqrt{N/\alpha}$. $\square$

3 Concluding Remarks

We have shown that some approximation error bounds for the pseudo expectation. When W is small, i.e., it is regarded as a constant, the bound of Theorem 4 is good. Some $o(N)$ bound is also provable (i.e., Theorem 7) for the case that W is proportional to N; altogether the error can be bounded by $o(N)$. We, however, conjecture that the error could be bounded by a much smaller function, maybe, a constant independent from N and W.

Notice that the argument of Theorem 4 could be called a *mechanical approach*, because no statistical property of the process is used. On the other hand, we would expect some strong concentration [TN03], which may be helpful to get a tighter bound. Since our example is a simple 1D case, one may expect to solve the formula for the expected state completely by solving a corresponding differential equation. Unfortunately, though, a straightforward application of such method only provides us $o(N)$ bound, and more careful consideration may be necessary to obtain a better bound.

Acknowledgments

The author would like to thank Remi Monasson for pointing out the idea leading me to the argument of Section 2.1.

References

[TN03] H. Takahashi and Y. Niikura, An extension of Azuma-Hoeffding inequalities and its application to an analysis of randomized local search algorithms, in *Proc. Workshop IBIS2003*, 177–181, 2003.

[WSA03] 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.