ISSN 1342-2812

Research Reports on Mathematical and Computing Sciences

An extension of Azuma-Hoeffding inequality

Hayato Takahashi and Yasuaki Niikura

March 2006, C–222

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES C: Computer Science

An extension of Azuma-Hoeffding inequality

Hayato Takahashi*and Yasuaki Niikura†

Abstract

We propose an approximation of expectation (*pseudo expectation*) of Markov process, which is originally discussed by Watanabe and Sawai in the context of randomized decoding algorithm for LDPC. We show a bound of the difference between pseudo expectation and expectation, and give a deviation probability with respect to the the pseudo expectation. As an example, we study certain kind of urn model, which is a simplified model of Watanabe and Sawai algorithm, and apply our results to the model.

Keywords: Markov chain, urn model, concentration, large deviation, randomized algorithm

1 Introduction

In Watanabe et. al. [4], a randomized decoding algorithm for LDPCC (low density parity check coding) is proposed. Since the state space of the algorithm is large, it is hard to obtain the average behavior of the algorithm; therefore they considered an approximation of the average. By numerical simulation, they observed that processes of the algorithm are highly concentrated and concluded that expectation and their approximation of expectation are close. In this paper, we call their approximation as *pseudo expectation*. The definition is given later.

In this paper we give analytical results for such a concentration phenomena. In order to do so, we first give a simple model. The processes of certain type of randomized local search algorithms including random decoding algorithm for LDPC can be modeled as follows (In the following, let n be a natural number) :
Let S_0 be the initial state of the algorithm (we assume that S_0 is constant). If n-step state of the algorithm is S_n then $n+1$-step state is $S_{n+1} = S_n + X(S_n)$, where $X(S_n)$ is a bounded random variable, which depends on the current state S_n. And if $S_n = q$ for some step n, where q is the terminate state, then the algorithm stops.

In the above model, let f be a function such that

$$E(S_{n+1}|S_n) = f(S_n),\ a.s.$$

*Tokyo Institute of Technology Department of Mathematical and Computing Sciences,
Email: Hayato.Takahashi@ieee.org
†NTT DATA

The pseudo expectation z_n is defined by

$$z_n = f^n(z_o),\ z_0 = S_0\ (= \text{constant}).$$

If f is linear, $E(S_{n+1}) = EE(S_{n+1}|S_n) = E(f(S_n)) = f(E(S_n))$ and hence $E(S_n) = f^n(S_0) = z_n$ for all n. If f is non-linear, in general, $E(S_n) \neq z_n$. By this property, the pseudo expectation z_n is considered to be a linear approximation of $E(S_n)$.

For a random variable X and a measurable function g, let $\operatorname{ess\,sup}^P_X g(X) = \inf\{a\ |P(g(X) > a) = 0\}$ and $\operatorname{ess\,inf}^P_X g(X) = \sup\{a\ |P(g(X) < a) = 0\}$; if P is obvious from the context, we omit P and write $\operatorname{ess\,sup}_X$ and $\operatorname{ess\,inf}_X$. In this paper we prove the following;

Theorem 1.1 *Let $S_0, S_1, \cdots, S_n$ be Markov process, and the state space S is a subset of R. Let f be a function such that $E(S_{i+1}|S_i) = f(S_i)$ a.s.*
Let

$$d_i = \operatorname*{ess\,sup}_{S_{i-1}} \operatorname*{ess\,sup}^{P(\cdot|S_{i-1})} S_i - \operatorname*{ess\,inf}^{P(\cdot|S_{i-1})} S_i \tag{1}$$

Assume that $\forall i,\ d_i < \infty$. Let

$$r(x,y) = \begin{cases} (f(x)-f(y))/(x-y), & \text{if } x \neq y \\ 0, & \text{if } x = y \end{cases},$$

$$r(S_i) = r(S_i, z_i),\ r_i = \operatorname*{ess\,sup}_{S_i} |r(S_i)| \text{ for } 0 \le i \le n,$$

$$D_n = \sum_{i=0}^{n-1} (d_{n-i} \prod_{j=1}^{i} r_{n-j})^2,\ \text{where } \prod_{j=1}^{0} r_{n-j} = 1,$$

and

$$\beta_i = E(\gamma_i \delta_i | S_{n-i-1}), \tag{2}$$

where $\gamma_i = \prod_{j=1}^{i} r(S_{n-j}) - E(\prod_{j=1}^{i} r(S_{n-j})|S_{n-i-1})$, $\gamma_0 = 1$, and $\delta_i = S_{n-i} - E(S_{n-i}|S_{n-i-1})$. Then for any $n,\ t > 0$,

$$E(S_n) = z_n + \sum_{i=0}^{n-1} E(\beta_i), \tag{3}$$

and

$$P(|S_n - z_n - \sum_{i=0}^{n-1} \beta_i| > t) \le 2e^{-2t^2/D_n}. \tag{4}$$

Remark 1.1 *In the above theorem, β_i is the covariance of $\prod_{j=1}^{i} r(S_{n-j})$ and S_{n-i} conditioned by S_{n-i-1}. Thus by Schwarz inequality, we have*

$$|\beta_i| \le \sqrt{E(\gamma_i^2|S_{n-i-1})E(\delta_i^2|S_{n-i-1})}. \qquad (5)$$

If f is linear, $\gamma_i \equiv 0$ and hence $\beta_i \equiv 0$. If S_i and z_i are included in a closed interval $[a,b], -\infty \le a < b \le \infty$ and f is continuously differentiable on $[a,b]$, then by the mean value theorem we have

$$\begin{aligned} \inf_{x\in[a,b]} \frac{d}{dx} f(x) &\le \inf_{S_i} r(S_i) \le r(S_i) \\ &\le \sup_{S_i} r(S_i) \le \sup_{x\in[a,b]} \frac{d}{dx} f(x). \end{aligned} \qquad (6)$$

Thus we have $r_i \le \sup_{x\in[a,b]} |\frac{d}{dx} f(x)|$.

Corollary 1.1 *Under the same condition of Theorem 1.1, we have*

$$E(|S_n - z_n|) \le \sqrt{\frac{\pi}{2} D_n} + E(|\sum_{i=0}^{n-1} \beta_i|), \qquad (7)$$

and

$$\limsup_n \frac{|S_n - z_n - \sum_{i=0}^{n-1} \beta_i|}{g(n)} \le 1 \quad a.s., \qquad (8)$$

where $g(n) = \sqrt{D_n(\frac{1}{2}\log n + \log\log n)}$. Thus we have

$$S_n = z_n + \sum_{i=0}^{n-1} \beta_i + O(g(n)) \quad a.s. \qquad (9)$$

Let $r_{\sup} = \sup_{x,y\in S} |r(x,y)|$. In this paper, we call the process *contracting* if $r_{\sup} < 1$. For that case, the processes are highly concentrated as follows.

Corollary 1.2 *Assume that $r_{\sup} < 1$ and $\sup_n d_n < \infty$. Then $\lim_n D_n$ exists and $D = \lim_n D_n \le (\sup_n d_n)^2 \frac{1}{1-r_{\sup}^2} < \infty$. Under the same condition of Theorem 1.1, we have*
(a) z_n converges, and $\lim_n z_n$ is the unique solution of $f(z) = z$.
(b) $\sup_n |\sum_{i=0}^{n-1} \beta_i| \le \sup_n d_n \frac{1}{1-r_{\sup}} < \infty$ a.s.
(c) $\sup_n |E(S_n) - z_n| \le \sup_n E(|S_n - z_n|) \le \sqrt{\frac{\pi}{2} D} + \sup_n E(|\sum_{i=0}^{n-1} \beta_i|)$.
(d) Let $\tilde{g}(n) = \sqrt{D(\frac{1}{2}\log n + \log\log n)}$. Then $\limsup_n \frac{|S_n - z_n|}{\tilde{g}(n)} \le 1$ a.s., and hence $S_n = z_n + O(\tilde{g}(n))$ a.s.

If $\{S_n\}$ is a martingale sequence, i.e., $E(S_n|S_{n-1}, \cdots, S_0) = S_{n-1}$, *a.s.* then $r \equiv 1$ and $\beta_i \equiv 0$, *a.s.* For that case, Theorem 1.1 reduces to the following theorem, and hence Theorem 1.1 is an extension of the Azuma-Hoeffding inequality.

Theorem 1.2 (Azuma-Hoeffding[1, 2]) *Let $X_1, X_2, \cdots X_n$ be a random process, and let $S_n = \sum_{i=1}^{n} X_i + S_0$. Assume (1). If $S_1, S_2, \cdots S_n$ is a martingale sequence, then for any n, $t > 0$,*

$$P(|S_n - S_0| \ge t) \le 2e^{-2t^2/\sum_{i=1}^{n} d_i^2}.$$

1.1 Comparing to other methods

Let $\forall i$, $Y_i = E(g|X_1, \cdots, X_i)$, then $\{Y_i\}$ is a martingale sequence. If we have a bound $d_i, 1 \le i \le n$ such that

$$\operatorname*{ess\,sup}_{X_1,\cdots,X_{i-1}} \; \operatorname*{ess\,sup}_{X_i}^{P(\cdot|X_1,\cdots,X_{i-1})} Y_i - \operatorname*{ess\,inf}_{X_i}^{P(\cdot|X_1,\cdots,X_{i-1})} Y_i \le d_i \qquad (10)$$

then the following bounded difference inequality holds for any n, $t > 0$ and for any random variables $X_1, \cdots, X_n$ [3],

$$\begin{aligned} &P(|g(X_1, \cdots, X_n) - Eg(X_1, \cdots, X_n| \ge t) \\ &\le 2e^{-2t^2/\sum_{i=1}^{n} d_i^2}. \end{aligned}$$

Equation (4) is considered to be a variant of the bounded difference inequality. However it is hard to obtain the difference d_i in (10) for our model below. Instead of computing d_i, we introduced r_i whose bound is easily obtained by (6).

In [5], under a certain assumption, an approximation of stochastic processes is studied. The approximation is the solution of a certain differential equation. Our method is similar to [5]; we proposed an approximation of the mean (pseudo expectation) and gave a concentration inequality around pseudo expectation of Markov process. To obtain the pseudo expectation, we simply iterate f. Thus regardless the size of the state space, we can easily compute the pseudo expectation. In particular our method is effective for contracting process as shown in the next section.

2 Example

In this section, we examine our result by simple models, which is considered to be an urn model. First consider the following game.
1. Player A and B have $A_n \ge 0$ and $B_n \ge 0$ 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. Probability of the event {A wins} depends on A_n.
Let S be the total number of balls. Then we see

$$\forall n, A_n + B_n = S. \qquad (11)$$

Let $P(A_n)$ and $P(B_n)$ be the probability of the event {A wins} and {B wins} respectively. Then we have

$$\forall n, P(A_n) + P(B_n) = 1.$$

Thus we see

$$E\left(\left(\begin{array}{c} A_{n+1} \\ B_{n+1} \end{array}\right) | \left(\begin{array}{c} A_n \\ B_n \end{array}\right)\right) = \left(\begin{array}{c} A_n \\ B_n \end{array}\right)$$
$$+\left(\begin{array}{c} -\min\{d, A_n\}P(A_n) + \min\{d, B_n\}P(B_n) \\ \min\{d, A_n\}P(A_n) - \min\{d, B_n\}P(B_n) \end{array}\right).$$

We approximate above equation by

$$E\left(\left(\begin{array}{c} A_{n+1} \\ B_{n+1} \end{array}\right) | \left(\begin{array}{c} A_n \\ B_n \end{array}\right)\right)$$
$$= \left(\begin{array}{c} A_n \\ B_n \end{array}\right) + d\left(\begin{array}{c} -P(A_n) + P(B_n) \\ P(A_n) - P(B_n) \end{array}\right) \quad (12)$$

2.1 linear case

In the above model, let $P(A_n)$ and $P(B_n)$ be A_n/S and B_n/S respectively. Then by (11) we have

$$E(A_{n+1}|A_n) = (1 - \frac{2d}{S})A_n + d, \ a.s.$$

In our notation,

$$\forall n, \ d_n = 2d, \ f(A_n) = (1-\frac{2d}{S})A_n+d, \ r \equiv (1-\frac{2d}{S}).$$

Hence we have

$$D_n = 4d^2\frac{1 - r^{2n}}{1 - r^2},$$

where $r = 1-\frac{2d}{S}$. Since f is linear, we see $\beta \equiv 0$ a.s. and $z_n = E(A_n)$ for all n (see Remark 1.1). Since the process is contracting ($r_{\sup} = 1 - \frac{2d}{S} < 1$), by solving $z = f(z)$, we see $\lim_n z_n = S/2$.

In Figure 1, we show $E(|S_n - z_n|)$, $(S_n = A_n)$ and its bound $\sqrt{\frac{\pi}{2}D_n}$ for this model.

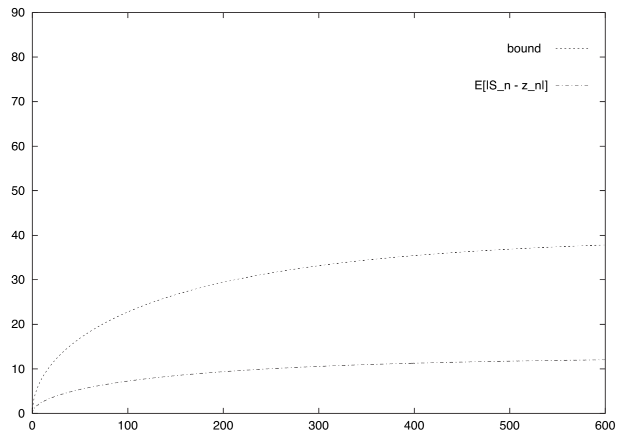

Figure 1: We show the graph $E(|S_n - z_n|)$ and its bound $\sqrt{\frac{\pi}{2}D_n}$ for $d = 1$ and $0 \le n \le 600$. The initial condition is $S_0 = A_0 = 500$ and $S = 1000$. The ratio $E(|S_n - z_n|) : \sqrt{\frac{\pi}{2}D_n}$ is about $1 : 4$.

2.2 non-linear case

Let $P(A_n)$ and $P(B_n)$ be $\frac{wA_n}{wA_n+B_n}$ and $\frac{B_n}{wA_n+B_n}$ respectively, where w is a positive constant such that $2d/S < w < S/2d$. Then by (11) we have

$$E(A_{n+1}|A_n) = A_n - d\frac{(w + 1)A_n - S}{(w - 1)A_n + S}, \ a.s.$$

In our notation,

$$\forall n, \ d_n = 2d, \ f(A_n) = A_n - d\frac{(w + 1)A_n - S}{(w - 1)A_n + S}.$$

In this case f is non-linear, clearly in general, $E(A_n) \neq z_n$. Since

$$\frac{d}{dx}f(x) = 1 - \frac{2wdS}{((w - 1)x + S)^2}, \quad (13)$$

we have (see Remark 1.1),

$$r_{\sup} \le f'(S) = 1 - \frac{2d}{wS} < 1, \quad (14)$$

and hence the process is contracting. By solving $z = f(z)$, we see $\lim_n z_n = S/(w+1)$. In this model, r_i is not constant. If we bound r_i by $1 - \frac{2d}{wS}$, we can not give a tight bound for $E(|A_n - z_n|)$. In order to give a sharp bound, we give a heuristic (not rigorous) approximation (15) below: Since the process is contracting, by Corollary 1.2, we see that A_n and z_n are close. Thus by (6), it seems that r_i is nearly equal to $f'(z_i)$ for $1 \le i \le n - 1$. Also by Figure 2, it seems that β_i is nearly equal to 0. Hence we arrive at the following heuristic approximation of the right-hand side of (7):

$$\sqrt{\frac{\pi}{2}D_n}, \text{ where } D_n = \sum_{i=0}^{n-1}(d_{n-i}\prod_{j=1}^{i} f'(z_{n-j}))^2. \quad (15)$$

In Figure 3, we show the graph $E(|S_n - z_n|)$ ($S_n = A_n$), and its approximated bound (15).

3 Proof of theorem

Let

$$\alpha(i, n) = (S_{n-i} - E(S_{n-i}|S_{n-i-1}))\prod_{j=1}^{i} r(S_{n-j}),$$

for $0 \le i \le n - 1$, where $\prod_{j=1}^{0} r(S_{n-j}) = 1$.

Lemma 3.1 $S_n - z_n = \sum_{i=0}^{n-1} \alpha(i, n)$.

Proof. By the definition of f, r, and z_n, we see

$$S_n - z_n = S_n - E(S_n|S_{n-1}) + E(S_n|S_{n-1}) - z_n$$
$$= S_n - E(S_n|S_{n-1}) + f(S_{n-1}) - f(z_{n-1})$$
$$= S_n - E(S_n|S_{n-1}) + r(S_{n-1})(S_{n-1} - z_{n-1}).$$

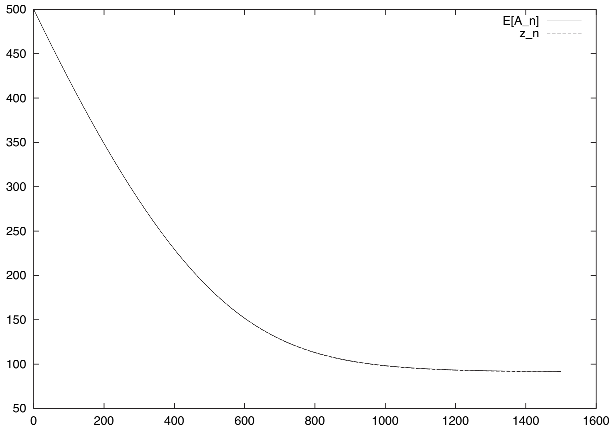

Figure 2: We show the graph $E(S_n)$ and z_n for $w = 10$, $d = 1$ and $0 \leq n \leq 1500$. The initial condition is $S_0 = A_0 = 500$ and $S = 1000$. We see that $E(S_n)$ and z_n are almost same in this scale and they converge to $S/(w+1) = 1000/11$.

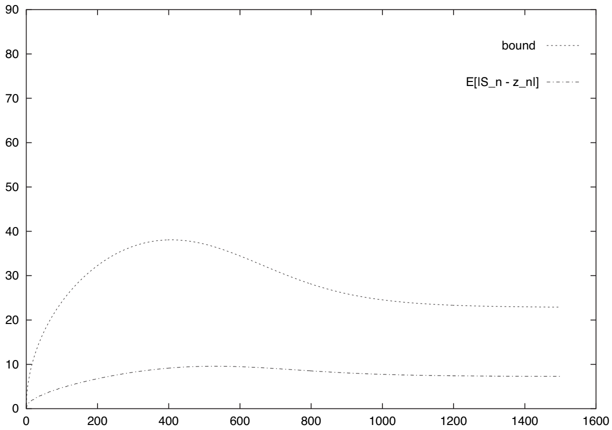

Figure 3: We show the graph $E(|S_n - z_n|)$ and approximated bound (15) for $w = 10$, $d = 1$ and $0 \leq n \leq 1500$. The initial condition is $S_0 = A_0 = 500$ and $S = 1000$.

By iteratively applying this identity to $S_{n-i} - z_{n-i}, 1 \leq i \leq n$, we have the lemma. Note that $S_0 = z_0$. ∎

Let

$$\beta_i = \beta(i,n) = E(\alpha(i,n)|S_{n-i-1}).$$

Then

$$\beta(i,n) = E((\prod_{j=1}^{i} r(S_{n-j}) - E(\prod_{j=1}^{i} r(S_{n-j})|S_{n-i-1})) (S_{n-i} - E(S_{n-i}|S_{n-i-1}))|S_{n-i-1}),$$

for $0 \leq i \leq n-1$.

Proof of Theorem 1.1.

Proof of (3).)

By Lemma 3.1, we see

$$E(S_n) - z_n = \sum_{i=0}^{n-1} E(\alpha(i,n)) = \sum_{i=0}^{n-1} EE(\alpha(i,n)|S_{n-i-1}) = \sum_{i=0}^{n-1} E(\beta(i,n)),$$

and (3) holds.

Proof of (4).)

By Lemma 3.1, we have

$$S_n - z_n - \sum_{i=0}^{n-1} \beta(i,n) = \sum_{i=0}^{n-1} (\alpha(i,n) - \beta(i,n)). \quad (16)$$

By the definition of α and β, we see

$$E(\alpha(i,n) - \beta(i,n)|S_{n-i-1}) = 0, \quad (17)$$

for all $0 \leq i \leq n-1$.

Fix S_{n-i-1}. Then by assumption, the length of the support set of S_{n-i} is bounded by d_{n-i}. Thus there are constants a and b such that $\alpha(i,n) \in [a,b]$, *a.s.*, and $b - a \leq d_{n-i} \prod_{j=1}^{i} r_{n-j}$. Also $\beta(i,n) \in [a,b]$. Hence there are constants a' and b' such that

$$\alpha(i,n) - \beta(i,n) \in [a', b'],\ a.s. \text{ and } b' - a' \leq d_{n-i} \prod_{j=1}^{i} r_{n-j}. \quad (18)$$

Recall that $\beta(i,n)$ is a function of S_{n-i-1}.

Let $Y_k^n = \sum_{i=n-k}^{n-1} (\alpha(i,n) - \beta(i,n)),\ 1 \leq k \leq n$. Then by (17) we have $E(Y_k^n|S_{k-1}) = Y_{k-1}^n$, and Y_n^n equals to left hand side of (16). Thus by (18) and applying $\{Y_k^n\}_{1 \leq k \leq n}$ to Azuma-Hoeffding inequality (Theorem 1.2), we have

$$P(|Y_k^n| > t) \leq 2e^{-2t^2/\sum_{i=n-k}^{n-1}(d_{n-i}\prod_{j=1}^{i} r_{n-j})^2}.$$

In particular by letting $k = n$ in the above inequality, we have (4). ∎

Proof of Corollary 1.1.

Proof of (7).

By (4), we have

$$E(|S_n - z_n - \sum_{i=0}^{n-1} \beta_i|) = \int_0^\infty P(|S_n - z_n - \sum_{i=0}^{n-1} \beta_i| > t)dt \leq \int_0^\infty 2e^{-2t^2/D_n} dt = \sqrt{\frac{\pi}{2} D_n}.$$

Thus we have (7).

Proof of (8) and (9)).

By (4), we have

$$\sum_n P(|S_n - z_n - \sum_{i=0}^{n-1} \beta_i| > g(n))$$
$$\leq \sum_n 2e^{-2g(n)^2/D_n} = \sum_n \frac{2}{n(\log n)^2} < \infty.$$

Thus by Borell-Cantelli lemma we have (8) and (9). ■

Proof of Corollary 1.2.
Since $D_n = \sum_{i=0}^{n-1}(d_{n-i}\prod_{j=1}^{i} r_{n-j})^2 \leq (\sup_n d_n)^2 \frac{1}{1-r_{\text{sup}}^2} < \infty$, we see $\lim_n D_n$ exists and $D = \lim_n D_n < \infty$. Also since f is a contracting map i.e., $\forall x, y,\ |f(x) - f(y)| < r_{\text{sup}}|x - y|$, we have (a). Also since

$$|\sum_{i=0}^{n-1} \beta_i| \leq \sum_{i=0}^{n-1} |\beta_i| \leq \sup_n d_n \sum_i r_{\text{sup}}^i \leq \sup_n d_n \frac{1}{1 - r_{\text{sup}}},$$

where the second inequality follows from (2), and hence we have (b). The first inequality of part (c) is trivial, and the second inequality of part (c) follows from Corollary 1.1 (a). Part (d) follows from Corollary 1.1 (b) and part (b) of this corollary. ■

References

[1] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Math. J.*, 19(3):357–367, 1967.

[2] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 53:13–30, 1963.

[3] C. McDiarmid. Concentration. In M. Habib et.al, editor, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248. Springer, Berlin, 1998.

[4] Osamu Watanabe, Takeshi Sawai, and Hayato Takahashi. Analysis of a randomized local search algorithm for LDPCC decoding problem. In *Lecture Notes in Comp. Sci.*, volume 2827, 2003.

[5] N. C. Wormald. Differential equations for random processes and random graphs. *Ann. Applied Prob.*, pages 1217–1235, 1995.