Journal of Inequalities in Pure and Applied Mathematics

http://jipam.vu.edu.au/

Volume 7, Issue 5, Article 187, 2006

A NOTE ON THE MARTINGALE INEQUALITY

YU MIAO

COLLEGE OF MATHEMATICS AND INFORMATION SCIENCE
HENAN NORMAL UNIVERSITY
453007 HENAN, CHINA.
yumiao728@yahoo.com.cn

Received 26 April, 2006; accepted 29 September, 2006
Communicated by F. Qi

ABSTRACT. In this paper, we will establish a martingale inequality, which extends the classic Hoeffding inequality in some sense. In addition, our inequality improves the results of Lee and Su [7] (2002) in some cases.

Key words and phrases: Bounded Martingale; Deviation bound; Hoeffding inequality; Martingale inequality.

2000 *Mathematics Subject Classification.* 60G42, 60E15.

1. INTRODUCTION

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathcal{F}_0 = \{\phi, \Omega\} \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}$, an integrable random variable $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ can be written as

$$X - \mathbb{E}X = \sum_{k=1}^{n} \left[\mathbb{E}(X|\mathcal{F}_k) - \mathbb{E}(X|\mathcal{F}_{k-1})\right] := \sum_{k=1}^{n} d_k,$$

where d_k is a martingale difference. An early inequality result is the following. If for any k, there exist constants a_k and b_k, such that $\mathbb{P}(d_k \in [a_k, b_k]) = 1$, then for any $t > 0$, we have the following classic Hoeffding inequality (cf. [5])

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2\exp\left\{-\frac{2t^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right\}.$$

De la Peña [2, 3] discussed a general class of exponential inequalities for bounded martingale difference and ratios by the decoupling theory. Andreas [9] gave exponential deviation inequalities for one-sided bounded martingale difference sequences. In the case of the length of longest increasing subsequences and the independence number of sparse random graphs, Lee and Su [7] have utilised the symmetry argument in the martingale inequality.

ISSN (electronic): 1443-5756

I wish to thank Y.Q. Chen and K. Zheng of Henan Normal University for many suggestions and helpful discussions during our writing of this paper. I am also grateful to the conscientious anonymous referee for his very serious and valuable report. His suggestions and comments have largely contributed to Section 3.

123-06

For these phenomena of measure concentration, the usual procedure in analysis is via martingale methods, information-theoretic methods and Talagrand's induction method (see [6, 8, 10]). In most applications, X is a function of n independent (possibly vector valued) random variables $\xi_1, \xi_2, \dots, \xi_n$ and the filtration is $\mathcal{F}_k = \sigma(\xi_1, \xi_2, \dots, \xi_n)$. In this case we let $\{\xi_1^{'}, \xi_2^{'}, \dots, \xi_n^{'}\}$ be an independent copy of $\{\xi_1, \xi_2, \dots, \xi_n\}$ and define

$$\Delta_k = X(\xi_1, \xi_2, \dots, \xi_{k-1}, \xi_k, \xi_{k+1}^{'}, \dots, \xi_n^{'}) - X(\xi_1, \xi_2, \dots, \xi_{k-1}, \xi_k^{'}, \xi_{k+1}^{'}, \dots, \xi_n^{'}).$$

Let $d_k = \mathbb{E}(\Delta_k|\mathcal{F}_k)$. By definition, Δ_k is the change in the value of X resulting from a change only in one coordinate. So, if there exists a constant c_k, such that $|\Delta_k| \le c_k$ a.s., then $|d_k| \le c_k$ a.s. and we can apply the Hoeffding inequality to obtain a tail bound for X. However, in many cases, c_k grows too rapidly and so the Hoeffding inequality does not provide any reasonable tail bound. For improving the Hoeffding inequality, Lee and Su [7] obtained the following reasonable tail bound for X.

Theorem 1.1 (See Theorem 1 in Lee and Su [7])**.** *Assume that there exists a positive and finite constant c such that for all $k \le n$, $|\Delta_k| \le c$ a.s. and there exist $0 < p_k < 1$ such that for each $k \le n$, $\mathbb{P}(0 < |\Delta_k| \le c|\mathcal{F}_{k-1}) \le p_k$ a.s. Then, for every $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le 2\exp\left\{-\frac{t^2}{2c^2\sum_{k=1}^n p_k + 2ct/3}\right\}. \tag{1.1}$$

In this paper, we will demonstrate that if $\frac{t}{c\sum_{k=1}^n p_k}$ is larger, especially if $\frac{t}{c\sum_{k=1}^n p_k} \ge 2.83e^{2.83}$, we can obtain a more precise inequality than (1.1). In Section 2, we will give the main results and show our inequalities are more precise than (1.1) in some cases. In Section 3, we apply our results to the longest increasing subsequence.

2. Main Results

In this section, we will continue to use the notions of Section 1.

Theorem 2.1. *Let X be an integrable random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is in fact a function of n independent random variables $\xi_1, \xi_2, \dots, \xi_n$. We define $\mathcal{F}_k$, Δ_k, d_k as in Section 1. Assume that there exist positive and finite constants c_k such that for all $k \le n$,*

$$|\Delta_k| \le c_k \quad \text{a.s.} \tag{2.1}$$

and there exist $0 < p_k < 1$ such that for each $k \le n$,

$$\mathbb{P}(0 < |\Delta_k| \le c_k|\mathcal{F}_{k-1}) \le p_k \quad \text{a.s.} \tag{2.2}$$

Then, for every $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le 2\exp\left\{-\frac{t^2}{2\sum_{k=1}^n e^{sc_k}c_k^2 p_k}\right\}, \tag{2.3}$$

where s satisfies the equation $s = \frac{t}{\sum_{k=1}^n e^{sc_k}c_k^2 p_k}$. In addition, if there exists a constant b, such that $s \ge b$, we will obtain

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le 2e^{-bt/2}. \tag{2.4}$$

Proof. In fact, we only prove the form $\mathbb{P}(X - \mathbb{E}X \ge t)$, and the other form $\mathbb{P}(X - \mathbb{E}X \le -t)$ is similar. By Jensen's inequality, for any $s > 0$, we have

$$\mathbb{E}(e^{sd_k}|\mathcal{F}_{k-1}) = \mathbb{E}(e^{s\mathbb{E}(\Delta_k|\mathcal{F}_k)}|\mathcal{F}_{k-1}) \le \mathbb{E}(e^{s\Delta_k}|\mathcal{F}_{k-1}), \quad \text{a.e.}$$

From (2.1), (2.2) and the following elementary inequality,

$$\forall x \in \mathbb{R}, \quad e^x \le 1 + x + \frac{|x|^2}{2} e^{|x|},$$

we can obtain

$$\begin{aligned}
\mathbb{E}(e^{s\Delta_k}|\mathcal{F}_{k-1}) &\le \mathbb{E}\left(1 + s\Delta_k + \frac{|s\Delta_k|^2}{2} e^{|s\Delta_k|}|\mathcal{F}_{k-1}\right) \\
&\le 1 + \frac{s^2}{2} e^{sc_k} \mathbb{E}(\Delta_k^2|\mathcal{F}_{k-1}) \\
&\le 1 + \frac{s^2}{2} e^{sc_k} c_k^2 p_k \\
&\le \exp\left\{\frac{s^2}{2} e^{sc_k} c_k^2 p_k\right\} \quad \text{a.e.}
\end{aligned}$$

It is easy to check that

$$X - \mathbb{E}X = \sum_{k=1}^{n} d_k.$$

Thus, by Markov's inequality, for any $s > 0$,

$$\begin{aligned}
\mathbb{P}(X - \mathbb{E}X \ge t) &\le e^{-st} \mathbb{E} e^{s(X - \mathbb{E}X)} \\
&\le e^{-st} \mathbb{E} e^{s \sum_{k=1}^{n} d_k} \\
&\le e^{-st} \mathbb{E}\left[e^{s \sum_{k=1}^{n-1} d_k} \mathbb{E}\left(e^{sd_n}|\mathcal{F}_{n-1}\right)\right] \\
&\le \exp\left\{-st + \frac{s^2}{2} e^{sc_n} c_n^2 p_n\right\} \mathbb{E} e^{s \sum_{k=1}^{n-1} d_k} \\
&\le \cdots \\
&\le \exp\left\{-st + \frac{s^2}{2} \sum_{k=1}^{n} e^{sc_k} c_k^2 p_k\right\}.
\end{aligned}$$

If we could take

$$s = \frac{t}{\sum_{k=1}^{n} e^{sc_k} c_k^2 p_k}, \tag{2.5}$$

(2.3) can be shown. In fact, putting $f_n(s) = \sum_{k=1}^{n} e^{sc_k} s c_k^2 p_k$, it is easy to see that for any n, $f_n(s)$ is a continuous function in s, and is nondecreasing on $[0, \infty)$ with $f_n(0) = 0$. Thus, for any $t > 0$, there exists only one solution that satisfies equation $s = \frac{t}{\sum_{k=1}^{n} e^{sc_k} c_k^2 p_k}$. The remainder of the proof is straightforward. □

Remark 2.2. It is easy to see that the solution of the equation $s = \frac{t}{\sum_{k=1}^{n} e^{sc_k} c_k^2 p_k}$ could not be given concretely. However, we can use the formula (2.4), by obtaining a low bound of s in many cases.

Corollary 2.3. *Under the conditions of Theorem 1.1, we assume that for all* $1 \le k \le n$, $c_k = c$. *Then, for every* $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le 2 \exp\left\{-\frac{t^2}{2e^{sc} c^2 \sum_{k=1}^{n} p_k}\right\}, \tag{2.6}$$

where s satisfies the equation $s = \frac{t}{e^{sc}c^2\sum_{k=1}^n p_k}$. In addition, if there exists a constant b, such that $s \geq b$, we obtain

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2e^{-bt/2}. \tag{2.7}$$

Next, we will show that, in some cases, the condition $s \geq b$ in Corollary 2.3 could be obtained and our results are better than inequality (1.1).

Proposition 2.4. *Under the conditions of Corollary 2.3,*

(R_1)**:** *Assuming that for any given $t > 0$,*

$$\frac{t}{c\sum_{k=1}^n p_k} \geq 2.83e^{2.83}, \tag{2.8}$$

then we have the following inequality

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2e^{-2.83t/(2c)}, \tag{2.9}$$

and in this case, our bound $e^{-2.83t/(2c)}$ is better than (1.1).

(R_2)**:** *Conversely, if for any given $t > 0$,*

$$\frac{t}{c\sum_{k=1}^n p_k} \leq 2.82e^{2.82}, \tag{2.10}$$

then (1.1) is better than our result.

Proof. By $s = \frac{t}{e^{sc}c^2\sum_{k=1}^n p_k}$ and $\frac{t}{c\sum_{k=1}^n p_k} \geq 2.83e^{2.83}$, it is easy to see that

$$sce^{sc} \geq 2.83e^{2.83} \qquad \text{and} \qquad sc \geq 2.83.$$

From Corollary 2.3, (2.9) can be obtained.

Next we will show that our bound $e^{-2.83t/(2c)}$ is better than (1.1). For $\frac{t}{c\sum_{k=1}^n p_k} \geq 3e^3$, we know

$$\begin{aligned}
&\frac{t}{c\sum_{k=1}^n p_k}(1/c - s/3) < s, \qquad s = \frac{t}{e^{sc}c^2\sum_{k=1}^n p_k}; \\
&\Leftrightarrow \frac{t}{c^2\sum_{k=1}^n p_k} < \frac{ts}{3c\sum_{k=1}^n p_k} + s, \qquad se^{sc} = \frac{t}{c^2\sum_{k=1}^n p_k}; \\
&\Leftrightarrow se^{sc} < \frac{ts}{3c\sum_{k=1}^n p_k} + s, \qquad se^{sc} = \frac{t}{c^2\sum_{k=1}^n p_k}; \\
&\Leftrightarrow e^{sc} < \frac{t}{3c\sum_{k=1}^n p_k} + 1, \qquad se^{sc} = \frac{t}{c^2\sum_{k=1}^n p_k}; \\
&\Leftrightarrow c(e^{sc} - 1)\sum_{k=1}^n p_k < t/3, \qquad se^{sc} = \frac{t}{c^2\sum_{k=1}^n p_k}; \\
&\Leftrightarrow 2c^2e^{sc}\sum_{k=1}^n p_k < 2c^2\sum_{k=1}^n p_k + 2ct/3, \qquad se^{sc} = \frac{t}{c^2\sum_{k=1}^n p_k}.
\end{aligned} \tag{2.11}$$

Thus, by comparing (2.6) and (1.1), the proof of (R_1) is given under the condition $\frac{t}{c\sum_{k=1}^n p_k} \geq 3e^3$. To proving remainders, by (2.11), we only show the following relations

$$\begin{cases} \frac{t}{c\sum_{k=1}^n p_k}(1/c - s/3) \geq s, \text{ if } 2.83e^{2.83} \leq \frac{t}{c\sum_{k=1}^n p_k} < 3e^3; \\ \frac{t}{c\sum_{k=1}^n p_k}(1/c - s/3) \leq s, \text{ if } \frac{t}{c\sum_{k=1}^n p_k} < 2.82e^{2.82}. \end{cases} \tag{2.12}$$

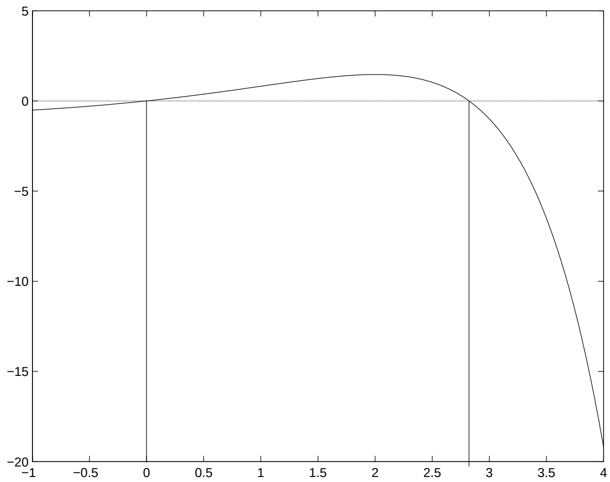

Figure 1

Since $s = \frac{t}{e^{sc}c^2\sum_{k=1}^n p_k}$, (2.12) is equivalent to the following relations

$$\begin{cases} ce^{sc}(1/c - s/3) \geq 1, & \text{if } 2.83e^{2.83} \leq \frac{t}{c\sum_{k=1}^n p_k} < 3e^3; \\ ce^{sc}(1/c - s/3) \leq 1, & \text{if } \frac{t}{c\sum_{k=1}^n p_k} < 2.82e^{2.82}. \end{cases} \tag{2.13}$$

Letting $f(s) = ce^{sc}(1/c - s/3) - 1$ and $sc = x$, we have $f(x) = e^x(1 - x/3) - 1$. It is not difficult to check that $f(x)$ is an increasing function in $[0, 2.82]$ and a decreasing function in $[2.83, \infty)$ (or see Figure 1). And $f(0) = 0$, $f(x_0) = 0$, where $x_0 \in [2.82, 2.83]$. The rest is obvious.

□

Remark 2.5. In the above proposition, though the bounds $2.82e^{2.82}$ and $2.83e^{2.83}$ are coarser, we can easily determine which inequalities are a little sharper by using these bounds.

Remark 2.6. The above results show that for given n (resp. t), our inequality is more precise in the case of sufficiently large t (resp. small n). However, in many cases, we need computer power to use our inequality. For example, assuming $\frac{t}{c\sum_{k=1}^n p_k} = B$, where B is given, then we often need to control the solution of the equation $xe^x = B$.

3. THE LONGEST INCREASING SUBSEQUENCE

In this section, we discuss the longest increasing subsequence as in Lee and Su [7] (2002) and show our results are little sharper. Consider the symmetric group S_n of permutations π on the number $1, 2, \ldots, n$, equipped with the uniform probability measure. Given a permutation $\pi = (\pi(1), \pi(2), \ldots, \pi(n))$, an increasing subsequence $i_1, i_2, \ldots, i_k$ is a subsequence of $1, 2, \ldots, n$ such that

$$i_1 < i_2 < \cdots < i_k, \quad \pi(i_1) < \pi(i_2) < \cdots < \pi(i_k).$$

We write $L_n(\pi)$ for the length of longest increasing subsequences of π.

Let $U_i = (X_i, Y_i)$, $i = 1, 2, \ldots, n$, be a sequence of i.i.d. uniform sample on the unit square $[0,1]^2$. $U_{i_1}, U_{i_2}, \ldots, U_{i_k}$ is called a monotone increasing chain of height k if

$$X_{i_j} < X_{i_j+1}, \ Y_{i_j} < Y_{i_j+1} \quad \text{for } j = 1, 2, \ldots, k-1.$$

Define $L_n(U)$ to be the maximum height of the chains in the sample $U_1, U_2, \ldots, U_n$.

By Hammersley [4] (1972) and Aldous and Diaconis [1] (1999), the following facts are known:

(F_1)**:** *$L_n(\pi)$ has the same distribution as $L_n(U)$.*
(F_2)**:** *$\frac{L_n(\pi)}{\sqrt{n}} \to 2$, in probability and in mean.*

Let $\{U_1', U_2', \ldots, U_n'\}$ be an independent copy of $\{U_1, U_2, \ldots, U_n\}$. It is easy to see that, letting

$$\Delta_k = L_n(U_1, \ldots, U_{k-1}, U_k, U_{k+1}', \ldots, U_n') - L_n(U_1, \ldots, U_{k-1}, U_k', U_{k+1}', \ldots, U_n')$$

Δ_k takes values only $+1$, 0, -1. Moreover, since $\mathbb{E}(\Delta_k|\mathcal{F}_{k-1})$, where $\mathcal{F}_{k-1} = \sigma(U_1, U_2, \ldots, U_{k-1})$, we have

$$\mathbb{P}(\Delta_k = +1|\mathcal{F}_{k-1}) = \mathbb{P}(\Delta_k = -1|\mathcal{F}_{k-1}).$$

Letting $p_k = 2\mathbb{E}L_{n-k+1}(U_k, U_{k+1}, \ldots, U_n)/(n-k+1)$, from Lee and Su [7] (2002), there is the following fact:

(F_3)**:** $\mathbb{P}(\Delta_k = +1|\mathcal{F}_{k-1}) \le p_k/2$.

For the longest increasing subsequence, we have the following result.

Theorem 3.1. *There exists a constant $\delta < 1/2$, such that for all sufficiently large n and any $r > 0$,*

$$\mathbb{P}(|L_n(U) - \mathbb{E}L_n(U)| > rn) \le 2\exp\left\{-\frac{\delta rn\log n}{2}\right\}. \tag{3.1}$$

Proof. For any $r > 0$ and sufficiently large n, s in Corollary 2.3 needs to satisfy the equation $s = \frac{rn}{e^s\sum_{k=1}^n p_k}$. Since

$$\frac{1}{\sqrt{n}}\mathbb{E}L_n(U) \to 2 \quad \text{as} \quad n \to \infty,$$

we have

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{\mathbb{E}L_k(U)}{k} \to 4 \quad \text{as} \quad n \to \infty, \quad \text{i.e.,} \quad n^{-1/2}\sum_{k=1}^n p_k \to 4.$$

By the equation $s = \frac{rn}{e^s\sum_{k=1}^n p_k}$, we know that for sufficiently large n, $se^s = O(\sqrt{n})$. Thus there exists a constant $\delta < 1/2$, such that $se^s > e^{\delta\log n}\delta\log n$, i.e., $s \ge \delta\log n$. By Corollary 2.3, we have the result. □

Remark 3.2. By Proposition 2.4, we know our results are sharper than the ones in Lee and Su [7] to a certainty. Lee and Su [7] gave the following result by an application of inequality (1.1).

Theorem LS. *Given any $\varepsilon > 0$, for all sufficiently large n and any $t > 0$,*

$$\mathbb{P}\left(|L_n(\pi) - \mathbb{E}L_n(\pi)| \ge t\right) \le 2\left(-\frac{t^2}{(16+\varepsilon)\sqrt{n} + 2t/3}\right). \tag{3.2}$$

Here if taking $t = rn$, then $\mathbb{P}\left(|L_n(\pi) - \mathbb{E}L_n(\pi)| \ge rn\right) \le O(e^{-n})$, which is coarser than (3.1)

REFERENCES

[1] D.J. ALDOUS AND P. DIACONIS, Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem, *Bull. Amer. Math. Soc.*, **36** (1999), 413–432.

[2] V.H. DE LA PEÑA, A bound on the moment generating function of a sum of dependent variables with an application to simple random sampling without replacement, *Ann. Inst. H. Poincaré Probab. Staticst.*, **30** (1994), 197–211.

[3] V.H. DE LA PEÑA, A general class of exponential inequalities for martingales and ratios, *Ann. Probab.*, **27** (1999), 537–564.

[4] J.M. HAMMERSLEY, A few seedlings of research, *Proceedings of the Sixth Berkeley Symposium on Mathematical and Statistical Probability*, University of California Press, Berkeley, CA, (1972), 345–394.

[5] W. HOEFFDING, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, **58** (1963), 12–20.

[6] M. LEDOUX, Isoperimetry and gaussian analysis, *Lectures on Probability Theory and Statistics*, Ecole d'Et'e de Probabilités de St-Flour XXIV-1994, P. Bernard (Editor), LNM 1648, Springer-Verlag, Berlin, 1996, 165–294.

[7] S. LEE AND Z.G. SU, The symmetry in the martingale inequality, *Stat. Probab. Letters,* **56** (2002), 83–91.

[8] P. MASSART, About the constants in Talagrand's concentration inequalities for empirical processes, *Ann. Probab.,* **28**(2) (2000), 863–884.

[9] A. MAURER, A bound on the deviation probability for sums of non-negative random variables, *J. Inequal. Pure. Appl. Math.*, **4**(1) (2003), Art. 15. [ONLINE: `http://jipam.vu.edu.au/article.php?sid=251`].

[10] M. TALAGRAND, Concentration of measure and isoperimetric inequalities in product spaces, *Publ. Math. del'IHES,* **81** (1995), 73–205.