The symmetry in the martingale inequality

Sungchul Lee[a],[*],[1], Zhonggen Su[a],[b],[2]

[a]*Department of Mathematics, Yonsei University, Seoul 120-749, South Korea*
[b]*Department of Mathematics, Zhejiang University, Hangzhou 310028, People's Republic of China*

Abstract

In this paper, we establish a martingale inequality and develop the symmetry argument to use this martingale inequality. We apply this to the length of the longest increasing subsequences and the independence number of sparse random graphs. © 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 60D05; 60F05; secondary 60K35; 05C05; 90C27

Keywords: Bounded martingale inequality; Longest increasing subsequence; Independence number

1. Introduction and main results

A common feature in many probabilistic arguments is to show that with high probability a random variable is concentrated on its mean. The usual way to do this is via either the martingale inequality, the isoperimetric inequality, or the log-Sobolev inequality. See Godbole and Hitczenko (1998), Janson et al. (2000), McDiarmid (1997, 1989), Steele (1997), Talagrand (1995) and Vu (2001) for various extensions and beautiful applications. In this paper, we establish a martingale inequality and develop the symmetry argument to use this martingale inequality. We apply this to the length of the longest increasing subsequences and the independence number of sparse random graphs.

To motivate the discussion below, let us begin with the well-known Azuma's inequality. Given a probability space $(\Omega, \mathscr{F}, P)$ and a filtration $\mathscr{F}_0 = \{\emptyset, \Omega\} \subset \mathscr{F}_1 \subset \cdots \subset \mathscr{F}_n = \mathscr{F}$, an integrable random

---

[*] Corresponding author.
E-mail addresses: sungchul@yonsei.ac.kr (S. Lee), zgsu@mail.hz.zj.cn, zgsu2001@yahoo.com (Z. Su).

variable $X \in L^1(\Omega, \mathcal{F}, P)$ can be written as

$$X - EX = \sum_{k=1}^{n} E(X|\mathcal{F}_k) - E(X|\mathcal{F}_{k-1}) := \sum_{k=1}^{n} d_k.$$

Here d_k is a martingale difference. If there exist constants $c_k > 0$ such that $|d_k| \leqslant c_k$ a.s. for each $k \leqslant n$, then for every $t \geqslant 0$,

$$P(X \geqslant EX + t) \leqslant \exp\left(-\frac{t^2}{2\sum_{k=1}^{n} c_k^2}\right),$$

$$P(X \leqslant EX - t) \leqslant \exp\left(-\frac{t^2}{2\sum_{k=1}^{n} c_k^2}\right).$$

The above result appears in Azuma (1967) and is often called Azuma's inequality, although it was actually earlier given by Hoeffding (1963). In most applications, X is a function of n-independent (possibly vector valued) random variables $\xi_1, \xi_2, \ldots, \xi_n$ and the filtration is

$$\mathcal{F}_k = \sigma(\xi_1, \xi_2, \ldots, \xi_k). \tag{1.1}$$

In this case, we let $\{\xi_1', \xi_2', \ldots, \xi_n'\}$ be an independent copy of $\{\xi_1, \xi_2, \ldots, \xi_n\}$ and define

$$\Delta_k = X(\xi_1, \ldots, \xi_{k-1}, \xi_k, \xi_{k+1}', \ldots, \xi_n') - X(\xi_1, \ldots, \xi_{k-1}, \xi_k', \xi_{k+1}', \ldots, \xi_n'), \tag{1.2}$$

$$d_k = E(\Delta_k|\mathcal{F}_k). \tag{1.3}$$

By definition, Δ_k is the change in the value of X resulting from a change only in one coordinate. So, if $|\Delta_k| \leqslant c_k$ a.s., then $|d_k| \leqslant c_k$ a.s. and we can apply Azuma's inequality to obtain a tail bound for X. However, in many cases c_k grows too rapidly that Azuma's inequality does not provide any reasonable tail bound. A detailed analysis on various problems in our paper shows that our Δ_k's are much smaller than c_k most of the time and from this observation we can improve Azuma's inequality and obtain a reasonable tail bound for X. Our result is the following.

Theorem 1. *Let X be an integrable random variable defined on a probability space $(\Omega, \mathcal{F}, P)$ which is in fact a function of n-independent random variables $\xi_1, \xi_2, \ldots, \xi_n$. We define $\mathcal{F}_k$, Δ_k, d_k by (1.1)–(1.3). Assume that there exists a positive and finite constant c such that for all $k \leqslant n$*

$$|\Delta_k| \leqslant c \text{ a.s.} \tag{1.4}$$

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

$$P(0 < |\Delta_k| \leqslant c|\mathcal{F}_{k-1}) \leqslant p_k \text{ a.s.} \tag{1.5}$$

Then, for every $t > 0$

$$P(X \geqslant EX + t) \leqslant \exp\left(-\frac{t^2}{2c^2 \sum_{k=1}^{n} p_k + 2ct/3}\right), \tag{1.6}$$

$$P(X \leqslant EX - t) \leqslant \exp\left(-\frac{t^2}{2c^2 \sum_{k=1}^{n} p_k + 2ct/3}\right). \tag{1.7}$$

Proof. We prove here (1.6) with $c=1$. From this one can easily get (1.6). The argument for (1.7) is the same. Let $\bar{p} = n^{-1} \sum_{k=1}^{n} p_k$ and let $g(x) = (e^x - 1 - x)/x^2$ with $g(0) = \frac{1}{2}$. Since $d_k = E(\Delta_k | \mathcal{F}_{k-1})$, by Jensen's inequality we have for any $s > 0$

$$E(e^{sd_k} | \mathcal{F}_{k-1}) = E(e^{sE(\Delta_k | \mathcal{F}_{k-1})} | \mathcal{F}_{k-1})$$

$$\leqslant E(e^{s\Delta_k} | \mathcal{F}_{k-1})$$

$$= E(1 + s\Delta_k + s^2 \Delta_k^2 g(s\Delta_k) | \mathcal{F}_{k-1}).$$

Since g is increasing and since $|\Delta_k| \leqslant 1$ a.s., by (1.5) we have

$$E(e^{sd_k} | \mathcal{F}_{k-1}) \leqslant E(1 + s\Delta_k + s^2 \Delta_k^2 g(s\Delta_k) | \mathcal{F}_{k-1})$$

$$\leqslant 1 + s^2 g(s) E(\Delta_k^2 | \mathcal{F}_{k-1})$$

$$\leqslant 1 + s^2 g(s) p_k$$

$$\leqslant e^{s^2 g(s) p_k} \text{ a.s.}$$

By Markov's inequality, then we have for any $s > 0$

$$P(X \geqslant EX + t) \leqslant e^{-st} E e^{s(X - EX)}$$

$$\leqslant e^{-st} E e^{s \sum_{k=1}^{n} d_k}$$

$$\leqslant e^{-st} E e^{s \sum_{k=1}^{n-1} d_k} E(e^{sd_n} | \mathcal{F}_{n-1})$$

$$\leqslant e^{-st} E e^{s \sum_{k=1}^{n-1} d_k} e^{s^2 g(s) p_n}$$

$$\leqslant \cdots$$

$$\leqslant e^{-st} e^{s^2 g(s) \sum_{k=1}^{n} p_k}$$

$$= e^{-st + (e^s - 1 - s) n \bar{p}}.$$

Letting $\phi(x) = (1 + x) \log(1 + x) - x$ we note that $\phi(x) \geqslant x^2/(2(1 + x/3))$. With the choice of $s = \log(1 + t/n\bar{p})$ we have then

$$P(X \geqslant EX + t) \leqslant \exp(-n\bar{p}\phi(t/n\bar{p}))$$

$$\leqslant \exp\left(-\frac{t^2}{2\sum_{k=1}^{n} p_k + 2t/3}\right). \qquad \square$$

In this paper we show a way, we call a symmetry argument, to use this inequality to obtain a reasonably good tail bound for X. More precisely, we show how to estimate the term $\sum_{k=1}^{n} p_k$ in Theorem 1. In some cases this is easy. See Luczak and McDiarmid (2001) for an example. However, in some other cases this is not an easy job at all. Actually, in his very beautiful survey McDiarmid (1997) introduced Talagrand's isoperimetric inequality by showing that the isoperimetric inequality gives a reasonable tail bound for the longest increasing subsequence whereas the martingale inequality

does not. We were suspicious about this and this was one of our motivations of this research. What we find in this paper is that by the symmetry argument we can control the term $\sum_{k=1}^{n} p_k$ effectively and surprisingly by the above martingale inequality we can provide a comparable tail bound for the longest increasing subsequence. In Section 2, we tell this story. In Section 3, we again use the martingale inequality and the symmetry, and provide a tail bound for the independence number. Compared to the isoperimetric inequality and the log-Sobolev inequality, the martingale inequality is rather elementary. Therefore, if one knows the symmetry argument in this paper, in many cases one can use the elementary martingale inequality to obtain a reasonably good tail bound.

2. The longest increasing subsequence

Consider the symmetric group S_n of permutations π on the numbers $1, 2, \ldots, n$, 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

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

We write $L_n(\pi)$ for the length of the longest increasing subsequences of π. It turns out that L_n provides an entry to a rich and diverse circle of mathematical ideas. The recent surveys Aldous and Diaconis (1999), Deift (2000) are useful references for the properties of L_n, various associated results and some of the history. We will present below only what we need for the proof of Theorem 2.

Let $U_i = (X_i, Y_i)$, $i = 1, 2, \ldots, n$, be a sequence of iid 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

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

Note that by definition we do not require $i_j < i_{j+1}$ and hence $U_{i_1}, U_{i_2}, \ldots, U_{i_k}$ is in general not a subsequence of $U_1, U_2, \ldots, U_n$. Define $L_n(U)$ to be the maximum height of the chains in the sample $U_1, U_2, \ldots, U_n$.

A key observation, due to Hammersley (1972), is that $L_n(\pi)$ has the same distribution as $L_n(U)$. In fact, based on this equivalent formulation Hammersley (1972) first proved that there is a constant c_2 such that

$$\frac{L_n(\pi)}{\sqrt{n}} \to c_2 \text{ in probability and in mean.}$$

The constant c_2 is now known to be 2. See Aldous and Diaconis (1999) for details.

Our main result regarding the longest increasing subsequence is as follows.

Theorem 2. *Given any $\varepsilon > 0$, for all sufficiently large n and any $t > 0$*

$$P(|L_n(\pi) - EL_n(\pi)| \geqslant t) \leqslant 2 \exp\left(-\frac{t^2}{(16 + \varepsilon)\sqrt{n} + 2t/3}\right). \tag{2.1}$$

Remark. Talagrand (1995) first obtained as an application of his isoperimetric inequality that for all $t > 0$,

$$P(L_n \geqslant M_n + t) \leqslant 2\exp\left(-\frac{t^2}{4(M_n + t)}\right), \qquad P(L_n \leqslant M_n - t) \leqslant 2\exp\left(-\frac{t^2}{4M_n}\right). \qquad (2.2)$$

where M_n denotes the median of L_n. Recently, Boucheron et al. (2000) used the log-Sobolev inequality to improve the Talagrand constants as follows:

$$P(L_n \geqslant EL_n + t) \leqslant \exp\left(-\frac{t^2}{2EL_n + 2t/3}\right), \qquad P(L_n \leqslant EL_n - t) \leqslant \exp\left(-\frac{t^2}{2EL_n}\right). \qquad (2.3)$$

Comparing (2.1) with (2.2) and (2.3), and noting $M_n/\sqrt{n} \to 2$ as $n \to \infty$, our elementary martingale argument provides the same order of bounds for L_n.

We also remark that Baik et al. (1999) proved that $\operatorname{Var} L_n \approx n^{1/3}$ and $(L_n - 2\sqrt{n})/n^{1/6}$ converges in distribution.

Proof. By Hammersley's equivalent formulation it suffices to show that the theorem holds for $L_n(U)$ instead of $L_n(\pi)$. 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 $+1, 0$, and -1. Moreover, since $E(\Delta_k | \mathscr{F}_{k-1}) = 0$ where $\mathscr{F}_{k-1} = \sigma(U_1, U_2, \ldots, U_{k-1})$, we have

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

Denote by A_j the event that any of the highest chains of $(U_1, \ldots, U_n)$ contains U_j. Since we are only concerned with the relative position of U_j in the n sample points instead of the order, by symmetry each A_j occurs with equal probability. Similarly, given $U_1, U_2, \ldots, U_{k-1}$, each A_j under the point configuration $U_1, \ldots, U_{k-1}, U_k, U_{k+1}', \ldots, U_n'$ occurs with equal probability for $j = k, k+1, \ldots, n$. Therefore, since $\Delta_k = +1$ implies that A_k happens, we have

$$P(\Delta_k = +1 | \mathscr{F}_{k-1}) \leqslant P(A_k | \mathscr{F}_{k-1})$$

$$= \frac{1}{n-k+1}\sum_{j=k}^{n} P(A_j | \mathscr{F}_{k-1})$$

$$= \frac{1}{n-k+1} E\left(\sum_{j=k}^{n} 1(A_j) | \mathscr{F}_{k-1}\right).$$

Note that $\sum_{j=k}^{n} 1(A_j)$ is just the number of points of $(U_k, U_{k+1}', \ldots, U_n')$, which are common to any highest chains of $(U_1, \ldots, U_{k-1}, U_k, U_{k+1}', \ldots, U_n')$ and hence if we collect the points U_j or U_j' with $1(A_j) = 1$, this itself forms a chain. Therefore, by the definition of chain we have

$$E\left(\sum_{j=k}^{n} 1(A_j) | \mathscr{F}_{k-1}\right) \leqslant EL_{n-k+1}(U_k, U_{k+1}, \ldots, U_n) \text{ a.s.}$$

Letting $p_k = 2EL_{n-k+1}(U_k, U_{k+1}, \ldots, U_n)/(n-k+1)$, we have

$$P(\Delta_k = +1|\mathscr{F}_{k-1}) \leqslant \frac{1}{n-k+1} E\left(\sum_{j=k}^n 1(A_j)|\mathscr{F}_{k-1}\right)$$

$$\leqslant \frac{1}{n-k+1} EL_{n-k+1}(U_k, U_{k+1}, \ldots, U_n)$$

$$= \frac{p_k}{2}.$$

Now we apply Theorem 1 to obtain

$$P(|L_n(U) - EL_n(U)| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{4\sum_{k=1}^n EL_k(U)/k + 2t/3}\right).$$

Since $EL_n(U)/\sqrt{n} \to 2$ as $n \to \infty$, $n^{-1/2}\sum_{k=1}^n EL_k(U)/k \to 4$ and hence the theorem follows. □

Next let us turn to the case of d-dimensional chains. Let $U_i = (X_i^1, X_i^2, \ldots, X_i^d)$, $i = 1, 2, \ldots, n$, be a sequence of iid uniform sample on the unit cube $[0,1]^d$. $U_{i_1}, U_{i_2}, \ldots, U_{i_k}$ is a monotone increasing chain of height k if

$$X_{i_j}^1 < X_{i_{j+1}}^1, X_{i_j}^2 < X_{i_{j+1}}^2, \ldots, X_{i_j}^d < X_{i_{j+1}}^d \quad \text{for } j = 1, 2, \ldots, k-1$$

Define $L_{n,d}(U)$ to be the maximum height of the chains in the sample $U_1, U_2, \ldots, U_n$. By an elementary calculation Bollobás and Winkler (1988) proved that there is a constant c_d such that

$$\frac{L_{n,d}(U)}{n^{1/d}} \to c_d \text{ in probability and in mean}$$

The exact values c_d for $d > 2$ are not known. However, there is a conjecture on c_d; $c_d = \sum_{k=0}^{d-1} 1/k!$. See Steele (1995) for details. With ingenuity and endeavor, Bollobás and Brightwell (1992) investigated the speed of convergence of $EL_{n,d}(U)/n^{1/d}$ to c_d and the concentration of $L_{n,d}(U)$ about its expectation by using the martingale inequality. Here is a better concentration bound. We skip its proof which is the same as that of Theorem 2.

Theorem 3. *Given any $\varepsilon > 0$, for all sufficiently large n and any $t > 0$*

$$P(|L_{n,d}(U) - EL_{n,d}(U)| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{4(dc_d + \varepsilon)n^{1/d} + 2t/3}\right).$$

3. The independence number

Given any complete graph K_n and $0 < p < 1$, we define a random graph $G(n,p)$ by selecting each edge in K_n with probability p independently. More specifically, first we name the vertices in K_n by k, $1 \leqslant k \leqslant n$. Let U_{ij}, $1 \leqslant i < j \leqslant m$, $m = \binom{n}{2}$, be the independent uniform weight assigned to the edge between i and j. Then, we keep the edges with $U \leqslant p$ and remove the edges with $U > p$. A subset $\mathscr{A}$ of vertices in K_n is independent in $G(n,p)$ if no two vertices from $\mathscr{A}$ are adjacent in $G(n,p)$. The independence number $\alpha(G(n,p))$ is the size of the largest independent set. We write $\alpha_n(p)$ for $\alpha(G(n,p))$ below.

For a fixed p, $\alpha_n(p)$ is remarkably concentrated; there exists $k = k(n, p)$ such that $\alpha_n(p) = k$ or $k + 1$ with high probability. See Shamir and Spencer (1987) for details. In this case, our method does not produce any new result.

For $p = s/n$ with $s > 0$ fixed, Bollobás and Thomason (1985) made very careful analysis of $\alpha_n(s/n)$. Let

$$f(s) = \sup \left\{ \beta > 0 \colon \lim_{n \to \infty} P\left(\alpha_n \left(\frac{s}{n} \right) > \beta n \right) = 1 \right\}.$$

In the range $0 < s \leqslant 1$, there is an explicit formula for $f(s)$. Although the formula is far from being pretty, it does enable one to compute particular values of $f(s)$. In the range $s > 1$, there is a rather weak low bound for $f(s)$; $f(s) \geqslant (s \log s - s + 1)/(s - 1)^2$. These results directly imply that in sparse random graphs the size of the largest independent sets is proportional to n, the size of the graph. So, it is natural to believe that the limit of the rate $E\alpha_n(s/n)/n$ exists. However, no rigorous proof is available. Frieze (1990) proved that by a large deviation inequality of Talagrand-type, for any fixed $\varepsilon > 0$ and for sufficiently large $n \geqslant n_0(\varepsilon)$ and $s \geqslant s_o(\varepsilon)$

$$\left| \alpha_n \left(\frac{s}{n} \right) - \gamma(s)n \right| \leqslant \frac{\varepsilon n}{s} \tag{3.1}$$

with high probability, and moreover

$$\left| E\alpha_n \left(\frac{s}{n} \right) - \gamma(s)n \right| \leqslant \frac{\varepsilon n}{s}, \tag{3.2}$$

where

$$\gamma(s) = \frac{2}{s}(\log s - \log \log s - \log 2 + 1).$$

However, his method cannot provide an explicit concentration inequality for the independence number. Here we apply Theorem 1 to obtain the following.

Theorem 4. *Let $\alpha_n = \alpha_n(s/n)$. Then, for any fixed $\varepsilon > 0$ and for all sufficiently large $n \geqslant n_1(\varepsilon)$ and $s \geqslant s_1(\varepsilon)$*

$$P(|\alpha_n - E\alpha_n| \geqslant t) \leqslant 2 \exp\left(-\frac{t^2}{4(\varepsilon + (1 - \varepsilon)(\gamma(\varepsilon s) + 1/s))n + 2t/3} \right).$$

Remark. For large s this provides a much better tail bound than the bound obtained by simply applying Azuma's inequality. Boucheron et al. (2000) also obtained similar bounds for α_n by using the log-Sobolev inequality.

Proof. We use the vertex-expose martingale argument. First, we name the vertices in K_n by k, $1 \leqslant k \leqslant n$. Let U_{ij}, $1 \leqslant i < j \leqslant m$, $m = \binom{n}{2}$, be the uniform weight assigned to the edge between i and j and let $\{U'_{ij}\}$ be an independent copy of $\{U_{ij}\}$. For each $1 \leqslant k \leqslant n$, let $G'_k(n, s/n)$ be the random graph generated by replacing the weight U adjacent to the vertex k by U', and write $\alpha(1, \ldots, k - 1, k', k + 1, \ldots, n)$ for the corresponding independence number. Then, we let

$$\Delta_k = \alpha(1, \ldots, k - 1, k, k + 1, \ldots, n) - \alpha(1, \ldots, k - 1, k', k + 1, \ldots, n).$$

It is easy to see that Δ_k takes only $+1, 0$, and -1. Moreover, since $E(\Delta_k|\mathscr{F}_{k-1}) = 0$ where $\mathscr{F}_k = \sigma(U_{ij}, 1 \leqslant i < j \leqslant k)$, we have

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

Denote by A_k the event that any of the largest independent sets of $G(n, s/n)$ contains the vertex k. By symmetry, given U_{ij}, $1 \leqslant i < j \leqslant k - 1$, each A_j, $k \leqslant j \leqslant n$, occurs equally likely. More precisely, for $k \leqslant j \leqslant n$

$$P(A_k|\mathscr{F}_{k-1}) = P(A_j|\mathscr{F}_{k-1}).$$

Thus, we have

$$P(\Delta_k = +1|\mathscr{F}_{k-1}) \leqslant P(A_k|\mathscr{F}_{k-1})$$
$$= \frac{1}{n - k + 1} \sum_{j=k}^{n} P(A_j|\mathscr{F}_{k-1})$$
$$= \frac{1}{n - k + 1} E\left(\sum_{j=k}^{n} 1(A_j)|\mathscr{F}_{k-1}\right).$$

Note that $\sum_{j=k}^{n} 1(A_j)$ is just the number of vertices which are common to any largest independent sets of $G(n, s/n)$ and hence if we collect the vertices j with $1(A_j) = 1$, this forms an independent set of the random graph on $\{k, k + 1, \ldots, n\}$ where each edge appears with probability s/n. Therefore, by the definition of the independent set we have

$$E\left(\sum_{j=k}^{n} 1(A_j)|\mathscr{F}_{k-1}\right) \leqslant E\alpha_{n-k+1}\left(\frac{s}{n}\right) \quad \text{a.s.}$$

Now a direct application of Theorem 1 gives

$$P(|\alpha_n - E\alpha_n| \geqslant t) \leqslant 2 \exp\left(-\frac{t^2}{4 \sum_{k=1}^{n} E\alpha_k(s/n)/k + 2t/3}\right). \tag{3.3}$$

Given $\varepsilon > 0$, in view of (3.2), for $k \geqslant n_0(1)$, $\varepsilon s \geqslant s_0(1)$ we have

$$\left|E\alpha_k\left(\frac{\varepsilon s}{k}\right)) - \gamma(\varepsilon s)k\right| \leqslant \frac{k}{s}.$$

Thus, since $\alpha_k(s/n) \leqslant k$, it easily follows that for all sufficiently large n and s

$$\sum_{k=1}^{n} \frac{E\alpha_k(\frac{s}{n})}{k} = \sum_{k \leqslant \varepsilon n} \frac{E\alpha_k(s/n)}{k} + \sum_{k > \varepsilon n} \frac{E\alpha_k(s/n)}{k}$$
$$\leqslant \varepsilon n + \sum_{k > \varepsilon n} \frac{E\alpha_k(\varepsilon s/k)}{k}$$
$$\leqslant \varepsilon n + (1 - \varepsilon)\left(\gamma(\varepsilon s) + \frac{1}{s}\right) n.$$

Now the theorem follows from (3.3). □

References

Aldous, D.J., Diaconis, P., 1999. Longest increasing subsequences: From patience sorting to the Baik–Deift–Johansson theorem. Bull. Amer. Math. Soc. 36, 413–432.

Azuma, K., 1967. Weighted sums of certain dependent variables. Tôhoku Math. J. 3, 357–367.

Baik, J., Deift, P., Johansson, K., 1999. On the distribution of the length of the longest increasing subsequence of random permutations. J. Amer. Math. Soc. 12, 1119–1178.

Bollobás, B., Brightwell, G., 1992. The height of a random partial order: concentration of measure. Ann. Appl. Probab. 2, 1009–1018.

Bollobás, B., Thomason, A.G., 1985. Random graphs of small order. In: Karoński, M., Ruciński, A. (Eds.), Random graphs'83, Proceedings, Poznań, 1983. North-Holland, Amsterdam, New York, pp. 47–97.

Bollobás, B., Winkler, P.M., 1988. The longest chain among random points in Euclidean space. Proc. Amer. Math. Soc. 103, 347–353.

Boucheron, S., Lugosi, G., Massart, P., 2000. A sharp concentration inequality with applications. Random Struct. Alg. 16, 277–292.

Deift, P., 2000. Integrable systems and combinatorical theory. Notices AMS. 47, 631–640.

Frieze, A.M., 1990. On the independence number of random graphs. Discrete Math. 81, 171–175.

Godbole, A., Hitczenko, P., 1998. Beyond the method of bounded differences. DIAMCS ser., Discrete Mathematics and Theoretical Computer Science 41, 43–57.

Hammersley, J.M., 1972. A few seedlings of research. Proceedings of the Sixth Berkeley Symposium on Mathematical and Statistical Probability, University of California Press, Berkeley, CA, pp. 345–394.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13–20.

Janson, S., Luczak, T., Rucinski, A., 2000. Random Graphs. Wiley, New York.

Luczak, M.J., McDiarmid, C., 2001. Bisecting sparse random graphs. Random Struct. Algebra 18, 31–38.

McDiarmid, C., 1989. On the Method of Bounded Differences. Surveys in Combinatorics. Cambridge University Press, Cambridge, pp. 148–188.

McDiarmid, C., 1997. Concentration. In: Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B. (Eds.), Probabilistic methods for algorithmic discrete mathematics. Springer, New York, pp. 195–248.

Shamir, E., Spencer, J., 1987. Sharp concentration of chromatic numbers on random graphs $G_{n,p}$. Combinatorica 7, 124–129.

Steele, J.M., 1995. Variations on the long increasing subsequence theme of Erdős and Szekeres. In: Aldous, D., Diaconis, P., Steele, J.M. (Eds.), Discrete probability and algorithms, Volumes in Mathematics and its applications, Vol. 72. Springer, New York, pp. 111–131.

Steele, J.M., 1997. Probability Theory and Combinatorial Optimization. SIAM, Philadelphia.

Talagrand, M., 1995. Concentration of measure and isoperimetric inequalities in product spaces. Publ. Math. del's IHES 81, 73–205.

Vu, V.H., 2001. Concentration of non-Lipschitz functions and applications, preprint.