The Annals of Applied Probability
1995, Vol. 5, No. 1, 44–48

ON RATES OF CONVERGENCE FOR COMMON SUBSEQUENCES AND FIRST PASSAGE TIME

BY WANSOO T. RHEE[1]

Ohio State University

We give a unified simple proof of recent results of Alexander concerning the rate of convergence of the mean length of the longest common subsequence of two random sequences and the rate of convergence of the expected value of certain passage times in percolation theory.

1. Longest common subsequence. Given a finite set A, we say that two sequences $\mathbf{x} = (x_i)_{i\le n}$ and $\mathbf{y} = (y_i)_{i\le n}$ in A have a common subsequence of length k provided for some $1 \le i_1 < \cdots < i_k \le n$ and $1 \le j_1 < \cdots < j_k \le n$, we have $x_{i_\ell} = y_{j_\ell}$ for all $1 \le \ell \le k$. The length of the longest common subsequence is denoted by $L(\mathbf{x}, \mathbf{y})$.

Consider now two independent A-valued i.i.d. sequences $\mathbf{X}_n = (X_i)_{i\le n}$ and $\mathbf{Y}_n = (Y_i)_{i\le n}$ with common law and the random variable $L_n = L(\mathbf{X}, \mathbf{Y})$. It follows from the standard superadditivity method that $\gamma = \lim_{n\to\infty} EL_n/n$ exists and depends only on the law of X_1.

THEOREM 1 (Alexander [2]). *For some universal constant K,*

$$n\gamma \ge EL_n \ge n\gamma - K(n\log n)^{1/2}. \tag{1}$$

Setting for simplicity $\gamma_n = EL_n$, we first observe that the inequality $\gamma n \ge \gamma_n$ follows from superadditivity. Our proof relies on the following two lemmas.

LEMMA 1. *For all n, $p \ge 0$,*

$$P(L_{4n} \ge 4p) \le (4n)^4 (P(L_{2n} \ge 2p))^{1/2}. \tag{2}$$

LEMMA 2. *For all $n > 0$, all t,*

$$P(|L_n - \gamma_n| \ge t) \le 2\exp(-t^2/2n). \tag{3}$$

We should observe that Lemma 2 is a straightforward consequence of Azuma's inequality; (see [1], [7], [8] and also [9], Section 6). Before we prove Lemma 1, we show how to deduce Theorem 1. Taking p such that $2p > \gamma_{2n}$,

Received January 1994; revised June 1994.
[1]Supported in part by NSF Grant DMS-93-03188.
AMS 1991 *subject classifications*. 60D05, 60G17.
Key words and phrases. Rate of convengence, Azuma's inequality, common subsequence, passage time.

(2) and (3) imply

$$P(L_{4n} \geq 4p) \leq 2(4n)^4 \exp\left(-\frac{(2p-\gamma_{2n})^2}{8n}\right).$$

A routine computation shows that this implies that

$$E(L_{4n}) = \gamma_{4n} \leq 2\gamma_{2n} + K(n \log n)^{1/2}. \tag{4}$$

(For simplicity, we denote by K a universal constant, not necessarily the same at each occurrence.) Using (4), for $2^k n$ rather than n, we get

$$\frac{\gamma_{2^{k+2}n}}{2^{k+2}n} \leq \frac{\gamma_{2^{k+1}n}}{2^{k+1}n} + \frac{K}{4}\left(\frac{\log 2^k n}{2^k n}\right)^{1/2}.$$

Summation of these inequalities for $0 \leq k \leq \ell$, and another routine computation shows that

$$\frac{\gamma_{2^{\ell+2}n}}{2^{\ell+2}n} \leq \frac{\gamma_{2n}}{2n} + K\left(\frac{\log n}{n}\right)^{1/2}.$$

Letting $\ell \to \infty$, we get $2n\gamma \leq \gamma_{2n} + K(n \log n)^{1/2}$. Thus (1) is proved for n even, and hence for all n by monotonicity of γ_n.

PROOF OF LEMMA 1. *Step* 1. Given two sequences $\mathbf{x} = (x_i)_{i \leq 4n}$ and $\mathbf{y} = (y_i)_{i \leq 4n}$, and integers $1 \leq q_1 \leq q_2 \leq 4n$ and $1 \leq r_1 \leq r_2 \leq 4n$, we define

$$L(\mathbf{x}, \mathbf{y};\ q_1, q_2;\ r_1, r_2) = L(x_{q_1}, \dots, x_{q_2};\ y_{r_1}, \dots, y_{r_2}).$$

We prove that if $L(\mathbf{x}, \mathbf{y}) \geq 4p$, we can find q_1, q_2, r_1, r_2 such that

$$q_2 - q_1 + r_2 - r_1 = 2n, \qquad L(\mathbf{x}, \mathbf{y};\ q_1, q_2;\ r_1, r_2) \geq p. \tag{5}$$

Indeed for $i \leq 5$, we can find integers n_i, m_i, with $n_1 = 1 < n_2 < \cdots < n_5 = 4n$ and $m_1 = 1 < m_2 < \cdots < m_5 = 4n$ such that for $1 \leq i \leq 4$, we have

$$L(\mathbf{x}, \mathbf{y};\ n_i, n_{i+1};\ m_i, m_{i+1}) = p.$$

Because

$$\sum_{1 \leq i \leq 4} (n_{i+1} - n_i + m_{i+1} - m_i) = 8n,$$

we have for some $1 \leq i \leq 4$, $n_{i+1} - n_i + m_{i+1} - m_i \leq 2n$. We then choose $q_1 \leq n_i$, $n_{i+1} \leq q_2$ and $r_1 \leq m_i, m_{i+1} \leq r_2$ such that (5) holds.

Step 2. Because (crudely) there are at most $(4n)^4$ possible choices for q_1, q_2, r_1, r_2, we can find such numbers satisfying the first requirement of (5) and such that

$$P(L(\mathbf{X}_{4n}, \mathbf{Y}_{4n};\ q_1, q_2;\ r_1, r_2) \geq p) \geq (4n)^{-4} P(L(\mathbf{X}_{4n}, \mathbf{Y}_{4n}) \geq 4p).$$

Setting $q = q_2 - q_1$ and $r = r_2 - r_1$, we have

$$\begin{aligned} P(L(\mathbf{X}_q, \mathbf{Y}_r) \geq p) &= P(L(\mathbf{X}_r, \mathbf{Y}_q) \geq p) \\ &= P(L(\mathbf{X}_{4n}, \mathbf{Y}_{4n};\ q_1, q_2;\ r_1, r_2) \geq p). \end{aligned}$$

Step 3. The key point is now the reflection argument. Because $q + r = 2n$,

$$P(L(\mathbf{X}_{2n}, \mathbf{Y}_{2n}) \geq 2p) \geq P(L(\mathbf{X}_q, \mathbf{Y}_r) \geq p) P(L(\mathbf{X}_r, \mathbf{Y}_q) \geq p),$$

which concludes the proof. □

2. Passage time. Let there be a bond between each nearest neighbor pair of sites in $\mathbb{Z}^d$. To each bond b is attached a random passage time X_b. These passage times are i.i.d. of common law μ. We assume

$$\mu(\{0\}) < p_c, \tag{6}$$

$$\int t^2 \, d\mu(t) < \infty, \tag{7}$$

where p_c is the critical probability for Bernoulli bond percolation on $\mathbb{Z}^d$.

The passage time through a lattice path Π is the sum of the passage times through the bonds of Π. Denoting by $\mathbf{0}$ the origin and by $\mathbf{n}$ the point of coordinates $(n, 0, \ldots, 0)$. We are interested in the infimum T_n of the passage times through all the paths linking $\mathbf{0}$ to $\mathbf{n}$. We set $\gamma_n = ET_n$ so that $\gamma_{n+m} \leq \gamma_n + \gamma_m$ and hence $\gamma = \lim_{n\to\infty} \gamma_n / n$ exists.

THEOREM 2. *Under conditions* (6) *and* (7), *for some constant C independent of n, we have*

$$n\gamma \leq \gamma_n \leq n\gamma + C(n \log n)^{1/2}.$$

This theorem is closely related to a result of Alexander [3]. The reason for the gain of a factor $\sqrt{\log n}$ and the weaker integrability condition (7) lies in the use of Lemma 4 (which follows) rather than of the original result of Kesten [5].

The structure of the proof is identical to that of Theorem 1. It relies on two lemmas. For simplicity, we denote by C a number depending only on μ, d, that may vary at each occurrence.

LEMMA 3. *For each $t > 0$ and some constant C, we have*

$$P(T_{4n} \leq 4t) \leq Cn^C (P(T_{2n} \leq 2t))^{1/2} + C\exp(-n/C). \tag{8}$$

LEMMA 4. *For each n, each $t > 0$, we have*

$$P(T_n \leq \gamma_n - t) \leq C\exp(-t^2/Cn). \tag{9}$$

Once these are proved, the proof of Theorem 2 mimics the proof of Theorem 1. Lemma 4 follows from [9], Section 8.3, equation (8.2.9).

PROOF OF LEMMA 3. *Step* 1. By a result of Kesten [5], it follows from (6) that there exist constants C_1, C_2, independent of n, such that with probability greater than or equal to $1 - e^{-n/C_1}$ there exists a path of length at most $C_2 n$ from $\mathbf{0}$ to $4\mathbf{n}$ with passage time T_{4n}.

Step 2. Consider a path Π from $\mathbf{0}$ to $4\mathbf{n}$ with passage time θ. For $0 \le i \le 4$, consider the first (resp. last) site $\mathbf{b}_i$ (resp. $\mathbf{a}_i$) on Π that has its first coordinate equal to in. Obviously, there exists $0 \le i \le 3$ such that the sum of the passage times on the bonds of Π between $\mathbf{a}_i$ and $\mathbf{b}_{i+1}$ is at most $\theta/4$. We observe also that if Π has length less than or equal to $C_2 n$, then there are at most $C_3 n^d$ choices for each $\mathbf{a}_i$ and each $\mathbf{b}_{i+1}$, where C_3 depends on C_2, d only.

Step 3. Given two sites $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^d$, let us denote by $H(\mathbf{a}, \mathbf{b}, t)$ the event that there exists a path Π from $\mathbf{a}$ to $\mathbf{b}$, with passage time at most t, such that each site visited by Π, except $\mathbf{a}$ and $\mathbf{b}$, has a first coordinate larger than the first coordinate a_1 of $\mathbf{a}$ and smaller than the first coordinate b_1 of $\mathbf{b}$.

It follows from Steps 1 and 2 that for any $t > 0$, we can find $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^d$, with $b_1 - a_1 = n$ and

$$P(H(\mathbf{a}, \mathbf{b}, t)) \ge C_3^{-10} n^{-10d} P(T_{4n} \le 4t) - \exp(-n/C_1).$$

Step 4. By translation invariance, $P(H(\mathbf{a}, \mathbf{b}, t)) = P(H(\mathbf{0}, \mathbf{b} - \mathbf{a}, t))$. Because the first coordinate of $\mathbf{b} - \mathbf{a}$ is n, we also have by reflection

$$P(T_{2n} \le 2t) \ge P(H(\mathbf{0}, \mathbf{b} - \mathbf{a}, t))^2.$$

This concludes the proof. □

3. Conclusion. The proofs we have given are rather straightforward and are significantly simpler than the original proofs. Both, however, rely on a reflection argument. This is a very serious limitation. Results of Kesten [6] and sharper recent results of Alexander [4] cover cases where no reflections seem relevant, and in particular the "off axis" case of first passage percolation. These results are considerably deeper than those presented here, whose appeal lies in simplicity rather than depth.

REFERENCES

[1] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tōhoku Math. J.* **19** 357–367.

[2] ALEXANDER, K. (1993). A note on some rates of convergence in first-passage percolation. *Ann. Appl. Probab.* **3** 81–90.

[3] ALEXANDER, K. (1993). The rate of convergence of the mean length of a longest common subsequence. Preprint.

[4] ALEXANDER, K. (1994). Fluctuations in the boundary of the wet region for first passage percolation in two and three dimensions. Preprint.

[5] KESTEN, H. (1986). Aspects of first-passage percolation. *Ecole d'Eté de Probabilités de Saint Flour XIV. Lecture Notes in Math.* **1180** 125–264. Springer, New York.

[6] KESTEN, H. (1993). On the speed of convergence in first-passage percolation. *Ann. Appl. Probab.* **3** 296–338.

[7] MCDIARMID, C. (1989). The method of bounded difference sequences. *Surveys in Combinatorics. London Math. Soc. Lecture Notes* **141** 148–188. Cambridge Univ. Press.

[8] RHEE, W. and TALAGRAND, M. (1987). Martingale inequalities and NP-complete problems. *Math. Oper. Res.* **12** 177–181.

[9] TALAGRAND, M. (1993). Isoperimetric inequalities and concentration of measure in product spaces. Preprint.

DEPARTMENT OF MANAGEMENT SCIENCES
OHIO STATE UNIVERSITY
301 HAGERTY HALL
1775 COLLEGE ROAD
COLUMBUS, OHIO 43210-1399