ON THE LENGTH OF THE LONGEST MONOTONE SUBSEQUENCE IN A RANDOM PERMUTATION

by

A. M. Frieze
Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213

Research Report No. 90-74 2

February 1990

On the length of the longest monotone subsequence in a random permutation

Alan Frieze*
Department of Mathematics, Carnegie-Mellon University,
Pittsburgh, U.S.A.

December 15, 1989

In this short note we prove a concentration result for the length L_n of the longest monotone increasing subsequence of a random permutation of the set $\{1, 2, ..., n\}$. It is known, Logan and Shepp [4], Vershik and Kerov [7] that

$$\lim_{n\to\infty} \frac{\mathbf{E}L_n}{\sqrt{n}} = 2 \qquad (1)$$

but less is known about the concentration of L_n around its mean. Our aim here is to prove the following.

Theorem 1 *Suppose that $\alpha > \frac{1}{3}$. Then there exists $\beta = \beta(\alpha) > 0$ such that for n sufficiently large*

$$\mathbf{Pr}(|L_n - \mathbf{E}L_n| \geq n^\alpha) \ \leq \ exp\{-n^\beta\}$$

□

*Supported by NSF grant CCR-8900112

Our main tool in the proof of this theorem is a simple inequality arising from the theory of martingales. It is often referred to as Azuma's inequality. See Bollobas [2] and McDiarmid [5] for surveys on its use in random graphs, probabilistic analiysis of algorithms etc., and Azuma [1] for the original result. A similar stronger inequality can be read out from Hoeffding [3]. We will use the result in the following form.

Suppose we have a random variable $Z = Z(U), U = U_1, U_2, ..., U_m$ where $U_1, U_2, ..., U_m$ are chosen independently from probability spaces $\Omega_1, \Omega_2, ..., \Omega_m$ i.e. $U \in \Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_m$. Assume next that Z is does not change by much if U does not change by much. More precisely write $U \simeq V$ for $U, V \in \Omega$ when U, V differ in at most one component i.e. $|\{i : U_i \neq V_i\}| = 1$. We state the inequality we need as a theorem.

Theorem 2 *Suppose Z above satisfies the following inequality;*

$$U \simeq V \textit{ implies } |Z(U) - Z(V)| \leq 1$$

then

$$\mathbf{Pr}(|Z - \mathbf{E}Z| \geq u) \leq 2exp\{-\frac{2u^2}{m}\},$$

for any real $u \geq 0$.

□

The value m is the *width* of the inequality and to obtain sharp concentration of measure we need $m = o((\mathbf{E}Z)^2)$.

Proof(of Theorem 1) Let $X = (X_1, X_2, ...X_n)$ be a sequence of independent uniform [0,1] random variables. We can clearly assume that L_n is the length of the longest monotone increasing subsequence of X.

Before getting on with the proof proper observe that although changing one X_i only changes L_n by at most 1, the width n is too large in relation to the mean $2\sqrt{n}$ for us to obtain a sharp concentration result. It therefore appears that to use the theorem in this case requires us to reduce the width by a more careful choice for Z.

For a set $I = \{i_1 < i_2 < ... < i_k\} \subseteq [n]$ we let $\lambda(I)$ denote the longest increasing subsequence of $X_{i_1}, X_{i_2}, ...X_{i_k}$. So for example $\lambda([n]) = L_n$.

Let $m = \lceil n^b \rceil, 0 < b < 1$ where a range for b will be given later. Let $\nu = \lfloor n/m \rfloor$ and $\mu = n - m\nu$ Let $I_1, I_2, ...I_m$ be the following partition of $[n] = \{1, 2, ..., n\}$: $I_j = \{k_j + 1, k_j + 2, ...k_{j+1}\}, j = 1, 2, ..., m$ where $k_j = j(\nu + 1)$ for $j = 0, 1, ...\mu$ and $k_j = j\nu + \mu$ for $j = \mu + 1, ..., m$. For $S \subseteq [m]$ we let $I_S = \bigcup_{j \in S} I_j$.

Let $\theta = n^a$ and $\epsilon = 2e^{-\theta}$. Define l by

$$l = max\{t : \mathbf{Pr}(L_n \leq t - 1) \leq \epsilon\}.$$

and

$$Z_n = max\{|S| : S \subseteq [m] \text{ and } \lambda(I_S) \leq l\}.$$

It follows from these definitions that

$$\mathbf{Pr}(Z_n = m) > \epsilon \tag{2}$$

Note next that for any $j \in [m]$, changing the value of $U_j = \{X_i : i \in I_j\}$, can only change the value of Z_n by at most one. We can thus apply Theorem 2 to obtain

$$\mathbf{Pr}(|Z_n - \mathbf{E}Z_n| \geq u) \leq 2exp\{-\frac{2u^2}{m}\} \tag{3}$$

Hence, putting $u = \sqrt{m\theta}$ in (3) and comparing with (2) we see that

$$\mathbf{E}Z_n \geq m - \sqrt{m\theta}.$$

Applying (3) once again with the same value for u we obtain

$$\mathbf{Pr}(Z_n \leq m - 2\sqrt{m\theta}) \leq \epsilon \tag{4}$$

We now need a crude probability inequality for L_s, where s is an arbitrary (large) positive integer.

Lemma 1

$$\mathbf{Pr}(L_s \geq 2e\sqrt{s}) \leq e^{-2e\sqrt{s}}$$

Proof Let $s_0 = \lceil e\sqrt{s} \rceil$. Then

$$\begin{aligned} \mathbf{Pr}(L_s \geq s_0) &\leq \binom{s}{s_0} / s_0! \\ &\leq \left(\frac{se^2}{s_0{}^2}\right)^{s_0} \\ &\leq e^{-2e\sqrt{s}} \end{aligned}$$

□

Let now $s = \lceil 2\sqrt{m\theta} \rceil$ and let $\mathcal{E}$ denote the event

$$\{\exists S \subseteq [m] : |S| = s \text{ and } \lambda(S) \geq 6\sqrt{\frac{sn}{m}}\}.$$

Now if $|S| = s$ then $|I_S| = (1 + o(1))(sn/m)$ and so on applying the lemma above we get

$$\begin{aligned}
\mathbf{Pr}(\mathcal{E}) &\le \binom{m}{s} e^{-2e\sqrt{sn/m}} \\
&\le exp\{s \ln m - 2e\sqrt{\frac{sn}{m}}\} \\
&\le \epsilon_1 = exp\{e(n^{\frac{a+b}{2}} \ln m - 2n^{\frac{1}{2}+\frac{a}{4}-\frac{b}{4}})\}
\end{aligned}$$

Now if $Z_n > m - 2\sqrt{m\theta}$ and $\mathcal{E}$ does not occur then

$$L_n \le l + 6\sqrt{\frac{sn}{m}}.$$

So

$$\mathbf{Pr}(L_n > l + 6\sqrt{\frac{sn}{m}}) \le \epsilon + \epsilon_1. \tag{5}$$

Putting $l_0 = l + 3\sqrt{\frac{sn}{m}}$ and combining this with the definition of l we have

$$\mathbf{Pr}(|L_n - l_0| \ge 3\sqrt{\frac{sn}{m}}) \le 2\epsilon + \epsilon_1 \tag{6}$$

The theorem follows by choosing any a, b, α, β such that

$$a + 3b < 2$$

and

$$\beta < \frac{1}{2} + \frac{a}{4} - \frac{b}{4} < \alpha$$

□

We remark finally that Steele [6] has generalised (1) in the following way: let now k be a fixed positive integer and given a random permutation let $L_{k,n}$ denote the length of the longest subsequence which can be decomposed into

$k+1$ successive monotone sequences, alternately increasing and decreasing. The monotone case above corresponds to k=0. In analogy to (1) Steele proves

$$\lim_{n\to\infty}\frac{\mathbf{E}L_{k,n}}{\sqrt{n}} = 2\sqrt{k+1}.$$

Theorem 1 generalises easily to include this problem. In fact we only need to change L_n to $L_{k,n}$ throughout. In order to avoid complicating the proof of Lemma 1, it suffices to prove

$$\mathbf{Pr}(L_s \geq 2(k+1)e\sqrt{s}) \leq e^{-2e\sqrt{s}}.$$

This follows from Lemma 1 since if the 'up and down' sequence is of length at least $2(k+1)e\sqrt{s}$ then one of the monotone pieces is at least $2e\sqrt{s}$ in length.

References

[1] K.Azuma, *Weighted sums of certain dependent random variables*, Tokuku Mathematics Journal 19, (1967) 357-367.

[2] B.Bollobas, *Sharp concentration of measure phenomena in random graphs*, to appear.

[3] W.Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association 27, (1963) 13-30.

[4] B.F.Logan and L.A.Shepp, *A variational problem for Young tableaux*, Advances in Mathematics 26 (1977) 206-222.

[5] C.J.H.McDiarmid, *On the method of bounded differences*, Surveys in Combinatorics, 1989, Invited papers at the Twelfth British Combinatorial Conference, Edited by J.Siemons, Cambridge University Press, 148-188.

[6] J.M.Steele, *Long unimodal subsequences: a problem of F.R.K.Chung*, Discrete Mathematics 33 (1981) 223-225.

[7] A.M.Vershik and C.V.Kerov, *Asymptotics of the Plancherel measure of the symmetric group and a limiting form for Young Tableau*, Dokl. Akad. Nauk. U.S.S.R. 233 (1977) 1024-1027.