ON THE EXPECTATION OF THE MAXIMUM WAITING TIME

TAMÁS F. MÓRI
Dept. of Probability Theory and Statistics, Eötvös Loránd University
1088 Budapest, Múzeum krt. 6 – 8.

(Received September 23, 1985)

1. Introduction

Let $Z_1, Z_2, \ldots$ be a sequence of i. i. d. random variables distributed uniformly on the set $S = \{1, 2, \ldots, d\}$. For any given sequence $A \in S^n$ let $\tau(A)$ denote the waiting time until A occurs as a run in the sequence $Z_1, Z_2, \ldots$. A great number of papers have been published on problems related to these waiting times since de Moivre's work [4], which can be considered as the origin of the systematic study of the longest pure head run in a sequence of fair coin tossings.

In [5] the following question was put and left unanswered. What is the asymptotics to the number of experiments tried until every possible head-and-tail pattern of length n appears as a connected subsequence of the outcomes in a coin-tossing game? This random variable can easily be expressed in terms of the above waiting times: it is simply $\max \tau(A)$ when A runs over the whole S^n and $d = 2$. In the following we shall denote it by W_n.

In a recent paper [2] A. Benczúr gives two-sided estimations for $\mathrm{E}(W_n)$. He proves that

$$2^n(\log 2^n - \log n + c + o(1)) \leqq E(W_n) \leqq 2^{n+2}(\log 2^n + c + o(1)),$$

where c is the Euler constant.

In a couple of interesting papers D. J. Aldous has dealt with the time taken by random walks on finite groups to visit every state. In [1] he proved that if the time taken for the transition probabilities to approach the uniform distribution is small compared with the cardinality G of the group, then the time taken for the random walk to visit every state is approximately $RG \log G$, where R is the mean number of returns to the initial state in the short term. Though the problem he investigated seems to be closely related to our subject, his theorems cannot be applied directly. On observing the sequence $Z_1, Z_2, \ldots$ one can readily define a homogeneous Markov chain with S^n as state space in a natural way, but no group structure on S^n will turn this Markov chain into a random walk in the sense of Aldous. We can, however, utilize some of Aldous' proof.

The aim of the present note is to prove the following estimation.

Theorem.

$$E(W_n) \sim d^n \log d^n$$

as $n \to \infty$.

The proof will be carried out in two steps contained in Sections 2 and 3.

2. Upper bound for $E(W_n)$

Since $W_n = \max\{\tau(A): A \in S^n\}$, it follows that

$$P(W_n > x) \leqq \sum_A P(\tau(A) > x). \tag{1}$$

Asymptotic exponentiality (and independence) of the waiting times $\tau(A)$ was proved in [6]. Here we need a little more: we must estimate the departure from exponentiality. We can refer to [5] where this was done, too, on the basis of the stability in the well-known characterization of the exponential distribution by its lack-of-memory property. As a special case of [5, Lemma 3] we obtain

$$P(\tau(A) > x) \leqq \frac{1-4b}{1-5b} \exp\left(-\frac{x}{E(\tau(A))}\right), \tag{2}$$

where $b = nd^{-n}$.

The expectation $E(\tau(A))$ can be calculated by using Conway's leading number algorithm (see [3] for an elegant proof): let $A = (a_1, a_2, \ldots, a_n) \in S^n$, then

$$E(\tau(A)) = \sum_{i=1}^{n} \varepsilon_i d^i,$$

where $\varepsilon_i = 1$ if the two sequences $(a_1, \ldots, a_i)$ and $(a_{n-i+1}, \ldots, a_n)$ are identical, and $\varepsilon_i = 0$ otherwise. It is easy to see that $d^n \leqq E(\tau(A)) < \frac{d}{d-1} d^n$ and the average of $E(\tau(A))$ over S^n is $d^n + n - 1$, hence for most A' s in S^n $E(\tau(A))$ cannot differ from d^n significantly. Let ε be a fixed positive number and

$$H = \{A \in S^n : E(\tau(A)) > d^n(1+\varepsilon)\},$$

then $|H| < \frac{n}{\varepsilon}$. From (1) it follows that

$$E(W_n) = \int_0^\infty P(W_n > x)dx \leqq \int_0^\infty \min\{\sum_A P(\tau(A) > x), 1\}\, dx \leqq$$

$$\leqq \delta + \sum_A \int_\delta^\infty P(\tau(A) > x)\, dx,$$

where $\delta = (1+\varepsilon)d^n \log d^n$. By applying (2) we obtain

$$\int_{\delta}^{\infty} P(\tau(A) > x)\, dx \leqq \frac{1-4b}{1-5b} \int_{\delta}^{\infty} \exp\left(-\frac{x}{E(\tau(A))}\right) dx =$$

$$= \frac{1-4b}{1-5b} E(\tau(A)) \exp\left(-\frac{\delta}{E(\tau(A))}\right).$$

For $A \notin H$ this can be majorized by $\frac{1-4b}{1-5b}(1+\varepsilon)$, while for $A \in H$ by $\frac{1-4b}{1-5b} \frac{d}{d-1} d^{n/d}$. Thus

$$E(W_n) \leqq (1+\varepsilon)\, d^n \log d^n + \frac{1-4b}{1-5b}\left[(1+\varepsilon)\, d^n + \frac{d}{d-1} \frac{n}{\varepsilon} d^{n/d}\right].$$

Choosing $\varepsilon = n^{-2}$ we arrive at the upper bound

$$E(W_n) \leqq d^n(\log d^n + 1 + o(1)). \tag{3}$$

3. Lower bound for $\mathbf{E(W_n)}$

Denote by $H(i)$ the set of n-sequences not observed as a run in $Z_1, Z_2, \ldots$ $\ldots, Z_{ni}$ and let $H(0) = S^n$. Consider the random variables $t_i = |H(i)|$ and the stoppig time $U = \inf\{i : t_i = 0\}$. Then obviously $t_i \leqq t_{i-1} \leqq t_i + n$ and $W_n > n(U-1)$.

First we show that

$$E(t_i - t_{i+1} \mid H(i-1)) \leqq nd^{-n}. \tag{4}$$

Indeed,

$$E(t_i - t_{i+1} \mid H(i-1)) =$$

$$= \mathrm{E}\Big(\sum_{A \in H(i)} I(\text{A is a run in } \{Z_j, n(i-1)+2 \leqq j \leqq n(i+1)\} \mid H(i-1)\Big) =$$

$$\leqq \mathrm{E}\Big(\sum_{A \in H(i-1)} I(\text{A is a run in } \{Z_j, n(i-1)+2 \leqq j \leqq n(i+1)\} \mid H(i-1)\Big),$$

where I stands for the indicator function. Since the process $\{Z_j, j \geqq n(i-1)+2\}$ is independent of $H(i-1)$, we obtain

$$E(t_i - t_{i+1} \mid H(i-1)) \leqq \sum_{A \in H(i-1)} P(A \text{ is a run in } Z_1, \ldots, Z_{2n-1}) \leqq t_{i-1} nd^{-n}.$$

Consider the sequence

$$Y_i = ind^{-n} + \log(t_i + 2n).$$

Using (4) we get

$$E(Y_{i+1}-Y_i|H(i-1)) =$$
$$= nd^{-n}-E(\log(t_i+2n)-\log(t_{i+1}+2n)|H(i-1)) \geqq \tag{5}$$
$$\geqq nd^{-n}-E\left(\frac{t_i-t_{i+1}}{t_{i+1}+2n}\,\middle|\,H(i-1)\right) \geqq nd^{-n}\left(1-\frac{t_{i-1}}{t_{i+1}+2n}\right) \geqq 0.$$

For any positive integer k

$$Y_{(U+1)\wedge k} = Y_1+\sum_{i=1}^{k-1}(Y_{i+1}-Y_i)I(U\geqq i).$$

Taking expectation we have

$$E(Y_{(U+1)\wedge k}) = E(Y_1)+E\left(\sum_{i=1}^{k-1}E[(Y_{i+1}-Y_i)I(U\geqq i)|H(i-1)]\right).$$

On the right hand side $I(U\geqq i) = I(H(i-1)\neq\emptyset)$ can be picked out from the conditional expectation, thus (5) gives

$$E(Y_{(U+1)\wedge k})\geqq E(Y_1)\geqq nd^{-n}+\log(d^n+n).$$

Here we used that $t_1\geqq d^n-n$.

On the other hand, the Jensen inequality implies

$$E(Y_{(U+1)\wedge k})\leqq E((U+1)nd^{-n})+E(\log(t_{(U+1)\wedge k}+2n))\leqq$$
$$\leqq nd^{-n}E(U+1)+\log E(t_{(U+1)\wedge k}+2n).$$

Let k tend increasingly to infinity and apply the monotone convergence theorem to the second term:

$$nd^{-n}+\log(d^n+n)\leqq nd^{-n}E(U+1)+\log 2n.$$

Hence

$$E(U)\geqq\frac{d^n}{n}\log\left(\frac{d^n+2n}{2n}\right),$$

and we obtain the following lower bound:

$$E(W_n)\geqq d^n\log\frac{d^n}{2n}-n.$$

From here and inequality (3) our theorem immediately follows.

4. Remarks

1. According to Theorem 3.1 of [3], the expected time till any of r competing runs appears can be calculated by solving a system of $r+1$ linear equations. This fact offers a method for computing the exact value of $E(W_n)$,

since the maximum of a set can be expressed in terms of the partial minimums assigned to its subsets. This sieve formula, however, seems to be hardly evaluable.

2. Author is able to prove that

$$\lim_{n\to\infty} P(d^{-n}W_n - \log d^n < x) = e^{-e^{-x}}.$$

The proof of this limit theorem is based on the large deviation results of [5] combined with the graph-sieve of Rényi and requires delicate technique. Details will be published elsewhere. If, in addition, one could verify the uniform integrability of the sequence $d^{-n}W_n - \log d^n$, the following equality would hold:

$$E(W_n) = d^n(\log d^n + c + o(1)),$$

where c is the Euler constant.

REFERENCES

[1] *Aldous D. J.:* On the time taken by random walks on finite groups to visit every state. *Z. Wahrsch. Verw. Gebiete* **62** (1983), 361–374.

[2] *Benczúr A.:* On the expected time of the first occurence of every k bit long patterns in the symmetric Bernoulli process. *Acta Math. Hungar.* **47** (1986), 233–238.

[3] *Li S. R.:* A martingale approach to the study of occurence of sequence patterns in repeated experiments. *Ann. Probab.* **8** (1980), 1171–1176.

[4] *de Moivre A.*: Doctrine of Chances, 2nd ed., London, 1738.

[5] *Móri T. F.:* Large deviation results for waiting times in repeated experiments. *Acta Math. Hungar.* **45** (1985), 213–221.

[6] *Móri T. F.* and *Székely G. J.:* Asymptotic independence of pure head stopping times. *Statistics & Probab. Letters* **2** (1984), 5–8.