ON THE MEAN DURATION OF A BALL AND CELL GAME; A FIRST PASSAGE PROBLEM[1]

By Harry Dym and Eugene M. Luks

Massachusetts Institute of Technology and Tufts University

1. Introduction. In this paper we study the mean duration of the following r ball n cell game:

Each of r balls is placed at random into one of n cells. A ball is considered "captured" if (after all r balls have been distributed) it is the sole occupant of its cell. Captured balls are eliminated from further play. This completes the first "trial." The remaining balls are recovered and the process repeated (trials, 2, 3, 4, $\cdots$ etc.). The play continues until all balls have been captured. The number of trials required to achieve this state is called the duration of the game.

We first show, in Section 2, that the probability of exactly $r - t$ balls remaining in play (or of exactly t balls being "captured") is equal to

$$P_{r,r-t}(n) = \sum_{j=t}^{n} (-1)^{j-t}\binom{j}{t}\binom{n}{j}\binom{r}{j}j!\,[(n-j)^{r-j}/n^r].$$

Various bounds on the probability of this event are then derived for subsequent use. When the intent is clear the n dependency may be suppressed and the symbol P_{rt} used instead of $P_{rt}(n)$. The notation is suggestive of that used in the theory of Markov chains. Indeed, the r ball n cell game may be identified with an $r + 1$ state Markov chain, the states being the number of balls possibly in play at any stage: r, $r - 1$, $\cdots$, 2, 1, or 0. In keeping with conventional Markov chain terminology we shall refer to the quantities $P_{rt}(n)$ as transition probabilities. Note also that the mean duration of an r ball game is equal to the mean first passage time to state "0" from initial state "r."

Denoting the mean duration of the game by $M_n(r)$ (or simply $M(r)$ if there is no ambiguity) we proceed in Section 3 to derive the bounds:

$$r^{-1}[n/(n-1)]^{r-1} \leqq M_n(r) \leqq \sum_{j=1}^{r} [1 - P_{jj}(n)]^{-1} \leqq n^2[n/(n-1)]^{r-1}.$$

The principal result of this paper, namely that

$$M_n(r) = \sum_{j=1}^{r} j^{-1}[n/(n-1)]^{j-1} + O(1) \qquad (r \to \infty, n \quad \text{fixed})$$

appears in Section 4. There it is also shown that

$$M_n(r) = \sum_{j=1}^{r} [1 - P_{jj}(n)]^{-1} + O(1) \qquad (r \to \infty, n \quad \text{fixed}).$$

The arguments leading to this last result which appear in Section 4 (up to and including Theorem 1) are presented in a form suggested by the referee. They

Received 14 December 1964; revised 27 September 1965.

[1] This paper is a revised version of TM-04091 (1964) (Technical Document Number 64–639), the MITRE Corporation, Bedford, Massachusetts. This work reported herein was supported by the MITRE Corporation under Contract No. AF 19(628) 2390.

may, as he noted, be extended to a larger class of Markov chains than the ball and cell game. This is discussed briefly following Theorem 1 in Section 4.

Since

$$\sum_{i=1}^{r} i^{-1}[n/(n-1)]^{i-1} \sim [(n-1)/r][n/(n-1)]^{r} \qquad (r \to \infty, n \quad \text{fixed}),$$

it follows readily from the results cited that

$$M_n(r) \sim (n/r)[n/(n-1)]^{r-1} \qquad (r \to \infty, n \quad \text{fixed})$$

where, it is interesting to note, $r[(n-1)/n]^{r-1}$ is equal to the mean number of balls captured in the first trial. The difference between the mean and its asymptote, however, diverges to infinity exponentially fast. In fact it may be shown that for any fixed integer $t \geqq 0$,

$$\sum_{i=1}^{r} i^{-1}[n/(n-1)]^{i-1} \sim [(n-1)/r][n/(n-1)]^{r} \sum_{i=0}^{t} (n-1)^{i}/\binom{r-1}{i}$$

$$(r \to \infty, n \text{ fixed}).$$

2. The transition probabilities. Let A_i denote the event: cell number i, $i = 1, \cdots, n$, contains exactly one ball. Then

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_1 \cap A_2 \cap \cdots \cap A_k) = \binom{r}{k} k! [(n-k)^{r-k}/n^r]. \tag{2.1}$$

Hence, setting

$$S_0(r, n) = 1,$$
$$S_k(r, n) = \binom{n}{k} P(A_1 \cap A_2 \cap \cdots \cap A_k), \qquad k = 1, \cdots, n, \tag{2.2}$$

and using the formula for the realization of exactly t out of n events [1], p. 96, we get

$$P_{r,r-t}(n) = \sum_{j=t}^{n} (-1)^{j-t} \binom{j}{t} S_j(r, n). \tag{2.3}$$

Equation (2.3) may be "inverted" to yield

$$S_m(r, n) = \sum_{k=m}^{n} \binom{k}{m} P_{r,r-k}(n), \qquad m = 0, 1, \cdots, n. \tag{2.4}$$

Clearly (2.4) implies that $S_1(r, n)$ is equal to the average number of balls captured in a single trial, and also that

$$P_{r,r-t}(n) \leqq S_t(r, n), \qquad t = 0, 1, \cdots, n. \tag{2.5}$$

The observation that $1 - P_{rr}(n) = P(A_1 \cup A_2 \cup \cdots \cup A_n)$ leads to the bounds

$$1 - P_{rr}(n) \leqq nP(A_1) = S_1(r, n), \tag{2.6}$$

$$1 - P_{rr}(n) \geqq nP(A_1) - \binom{n}{2} P(A_1 \cap A_2) = S_1(r, n) - S_2(r, n), \tag{2.7}$$

$$1 - P_{rr}(n) \geqq P(A_1) \quad = S_1(r, n)/n \tag{2.8}$$

all of which will be utilized below.

3. Bounds on $M(r)$. It is well known that

$$M(r) = \sum_{t=1}^{r} P_{rt} M(t) + 1 \tag{3.1}$$

or equivalently that

$$M(r) = \sum_{t=1}^{r-1} [P_{rt}/(1 - P_{rr})] M(t) + (1 - P_{rr})^{-1}. \tag{3.2}$$

We shall use (3.2) to show inductively that

$$M(r) \leqq \sum_{i=1}^{r} (1 - P_{ii})^{-1} \tag{3.3}$$
$$(= \textstyle\sum_{i=1}^{r} (\text{mean no. of trials to leave state "}i\text{"})).$$

Clearly (3.3) holds if $r = 1$ for $M(1) = (1 - P_{11})^{-1}$. Suppose it is valid for $r < s$. A substitution into (3.2) then yields

$$\begin{aligned} M(s) &\leqq \sum_{t=1}^{s-1} [P_{st}/(1 - P_{ss})] \sum_{i=1}^{t} (1 - P_{ii})^{-1} + (1 - P_{ss})^{-1} \\ &= \sum_{i=1}^{s-1} (1 - P_{ii})^{-1} \sum_{t=i}^{s-1} [P_{st}/(1 - P_{ss})] + (1 - P_{ss})^{-1}. \end{aligned}$$

Since $\sum_{i=1}^{s-1} P_{st} \leqq 1 - P_{ss}$ (3.3) is valid for $r = s$ also. Now, using (2.8) in conjunction with (3.3) we get the upper bound

$$\begin{aligned} M(r) &\leqq n \sum_{i=1}^{r} i^{-1} [n/(n - 1)]^{i-1} \\ &\leqq n \sum_{i=1}^{r} [n/(n - 1)]^{i-1} \leqq n^2 [n/(n - 1)]^{r-1}. \end{aligned} \tag{3.4}$$

A lower bound for $M(r)$ may be deduced from (3.2) with the aid of (2.6):

$$M(r) \geqq (1 - P_{rr})^{-1} \geqq r^{-1} [n/(n - 1)]^{r-1}. \tag{3.5}$$

The latter bound, while not necessarily strong, does establish the exponential growth of $M(r)$ with r.

4. The limiting behavior of $M_n(r)$ ($r \to \infty$, n fixed). It follows easily from (3.2) that

$$\begin{aligned} &(1 - P_{jj})^{-1} - [M(j) - M(j - 1)] \\ &\qquad = (1 - P_{jj})^{-1} \sum_{s=0}^{j-2} [M(j - 1) - M(s)] P_{js} \\ &\qquad \leqq C_j M(j - 1), \end{aligned}$$

where

$$C_j = (1 - P_{jj})^{-1} \sum_{s=0}^{j-2} P_{js}. \tag{4.1}$$

Hence, utilizing this inequality in (3.3) we see that

$$\begin{aligned} 0 &\leqq \sum_{j=1}^{r} (1 - P_{jj})^{-1} - M(r) \\ &= \sum_{j=2}^{r} \{(1 - P_{jj})^{-1} - [M(j) - M(j - 1)]\} \leqq \sum_{j=2}^{r} C_j M(j - 1). \end{aligned}$$

This implies

THEOREM 1.

$$M(r) = \sum_{j=1}^{r} (1 - P_{jj})^{-1} + O(1) \qquad (r \to \infty)$$

whenever $\sum_{j=2}^{\infty} C_j M(j-1) \leqq \sum_{j=2}^{\infty} C_j \sum_{k=1}^{j-1} (1 - P_{kk})^{-1} < \infty$.

It should be noted that Theorem 1 is valid for any Markov chain (with states 0, 1, 2, $\cdots$) whose transition probabilities, P_{ij}, satisfy the conditions:
(1) $P_{jj} < 1, j = 1, 2, \cdots$, and
(2) $P_{ij} = 0$ if $j > i$.
In the case at hand we find, upon substituting (2.5) and (2.8) into (4.1),

$$\begin{aligned} C_j &= (1 - P_{jj})^{-1} \sum_{k=0}^{j-2} P_{jk} \leqq [n/S_1(j, n)] \sum_{k=2}^{n} S_k(j, n) \\ &= [(n-1)/j][n/(n-1)]^j \sum_{k=2}^{n} \binom{n}{k}\binom{j}{k} k!(n-k)^{j-k}/n^j \\ &\leqq [(n-1)/j][n/(n-1)]^j[(n-2)/n]^j j^n \sum_{k=2}^{n} \binom{n}{k} \\ &\leqq (n-1)[(n-2)/(n-1)]^j 2^n j^{n-1}, \end{aligned}$$

which when combined with (3.4) yields

$$C_j M(j-1) \leqq (n-1)^3 2^n j^{n-1}[1 - (n-1)^{-2}]^j.$$

Hence

$$\sum_{j=2}^{\infty} C_j M(j-1) \leqq \sum_{j=2}^{\infty} (n-1)^3 2^n j^{n-1}[1 - (n-1)^{-2}]^j < \infty$$

and Theorem 1 implies

(4.2) $$M_n(r) = \sum_{j=1}^{r} (1 - P_{jj})^{-1} + O(1) \qquad (r \to \infty, n \text{ fixed}).$$

Now, it follows by (2.6), (2.7) and (2.8) that

$$\begin{aligned} 0 &\leqq (1 - P_{jj})^{-1} - [S_1(j, n)]^{-1} = (S_1 - 1 + P_{jj})/(1 - P_{jj})S_1 \\ &\leqq nS_2/S_1^2 \leqq n[1 - (n-1)^{-2}]^j 2 \end{aligned}$$

and thus, since $\sum_{j=1}^{\infty} n[1 - (n-1)^{-2}]^j < \infty$, we can, recalling the formula for $S_1(j, n)$, write

(4.3) $$M_n(r) = \sum_{j=1}^{r} j^{-1}[n/(n-1)]^{j-1} + O(1) \quad (r \to \infty, n \text{ fixed}).$$

The above derivation also shows that

$$M_2(r) = \sum_{j=1}^{r} 2^{j-1}/j.$$

REMARK 1. Consider the difference between the mean and the approximating sum:

$$E_n(r) = M_n(r) - \sum_{j=1}^{r} j^{-1}[n/(n-1)]^{j-1}.$$

$E_2(r) = 0$. Direct computation indicates that $|E_3(r)| \leqq 0.25$ and that $E_3(r)$ approaches about 0.042 as r approaches infinity. Sample estimates of $E_n(r)$

were also obtained for the cases $4 \leqq n \leqq 12$ by simulating the game on a digital computer. The estimates for $E_n(r)$ were, at least in the range $r \leqq 5n$, always less than one in magnitude.

REFERENCE

[1] FELLER, W. (1959). An Introduction to Probability Theory and Its Applications, **1** (2nd ed.). Wiley, New York.