RETURNS TO THE ORIGIN FOR RANDOM WALKS ON $\mathbb{Z}$

Loucas A. Chrysafi and R. E. Bradley

Abstract. We present a combinatorial theorem which generalizes an identity of Feller and applies it to the study of returns to the origin for the symmetric random walk on $\mathbb{Z}$.

1. Introduction. An ambidextrous mathematician with a sore throat has a long day of lecturing ahead of her. She puts n throat lozenges into her left pocket and another n into her right pocket. When her throat feels irritated, she chooses a pocket at random and takes a lozenge from it. We assume that she is not aware of taking the last lozenge from a pocket and only notices that the pocket is empty when she next tries, unsuccessfully, to take a lozenge from it.

When she first realizes that one pocket is empty, how many lozenges remain in the other pocket? The answer is clearly a random variable with a range $r = 0, 1, \ldots, n$. In order to have precisely r of them in her right pocket when she finds that the left pocket is empty, she must have previously made $2n - r$ selections, exactly n of which were from the left pocket. This is a binomial probability and, assuming equal likelihood of choosing left and right, the quantity is

$$\binom{2n-r}{n}\frac{1}{2^{2n-r}}. \tag{1}$$

Of course, we must multiply the quantity (1) by $1/2$ to account for the event of choosing the left pocket in order to discover the empty state, but this is cancelled by the symmetric case when the right pocket is the first to be emptied.

This problem appears to originate with Feller [4] where it is called the *problem of Banach's matchboxes*, inspired by an address made by H. Steinhaus in honor of Banach, wherein the latter's smoking habits were mentioned. Mathematically, our presentation is nearly identical; the social setting is perhaps more appropriate for our times.

There is a connection between the lozenge problem and the random walk; indeed, the reader may already have observed that the case $r = 0$ in (1) is precisely the probability of having returned to the origin at the end of a $2n$-step symmetric random walk (SRW) in one dimension, a quantity which appears in an elementary proof of Polya's Theorem for one dimension; for example, Section 8 of Billingsley [1].

What is less obvious is that the same quantity is the probability that the walker does not return to the origin at all during the same $2n$-step SRW. In fact, the quantity (1) is the probability that a $2n$-step SRW returns to the origin exactly r times during the walk. This will be discussed in Section 3.

In this paper, we are primarily concerned with the random variable R, whose distribution is given by the formula (1). In an exercise in Feller [4], the reader is asked to find the expected value of such a random variable. As a hint, Feller asks the reader to use a relation involving the quantity (1). That relation is the special case, $k = 1$, of our Theorem 2.3.

In Section 2 of this paper, we prove this generalization of Feller's identity, which is a purely combinatorial result making no reference to random walks, throat lozenges, or smoking. In Section 3 we apply the result to the calculation of the mean and variance of R. Using a simple combinatorial argument we also prove the fact that the expected value of R is a sum of transition probabilities. The machinery of Section 2 allows us to calculate any moment of R, and the third, fourth, fifth, and sixth moments are presented without proof. In the final section, we make a conjecture about the asymptotic behavior of the moments of R as n tends to infinity.

2. The Main Theorem.

Definition. Suppose n and r are non-negative integers. If $0 \leq r \leq n$, let

$$P_{n,r} = \binom{2n-r}{n} 2^{r-2n}. \tag{2}$$

For $r > n$ we define $P_{n,r} = 0$.

Lemma 2.1. If $n \geq 1$ then

$$P_{n,0} = P_{n,1} = \frac{1}{2^{2n}} \binom{2n}{n}.$$

Proof. When $r = 0$, this follows from the definition. For $r = 1$, we observe that

$$\binom{2n}{n} = 2\binom{2n-1}{n}.$$

Definition. Given a real number x and an integer $k > 0$, define the *falling factorial* of x as follows $[x]_k = x(x-1)\cdots(x-k+1)$. We adopt the convention that $[x]_0 = 1$.

Definition. Given non-negative integers n and k with $k \geq 1$, let

$$F_{n,k}(\lambda) = \prod_{i=1}^{k}(2n + i - \lambda).$$

Lemma 2.2.

$$F_{n,k+1}(\lambda + 1) = (2n - \lambda)F_{n,k}(\lambda).$$

The proof is the result of factoring out the first term and re-indexing.

Theorem 2.3. Suppose n and r are nonnegative integers with $0 \leq r \leq n$. Then for any $k \geq 1$

$$[n-r]_k P_{n,r} = 2^{-k} F_{n,k}(r+k) P_{n,r+k}. \tag{3}$$

Proof. We proceed by induction. If $k = 1$, the right hand side of (3) is

$$\begin{aligned}
2^{-1}[(2n+1) - (r+1)]P_{n,r+1} &= 2^{-1+(r+1)-2n}(2n-r)\binom{2n-(r+1)}{n} \\
&= 2^{r-2n}(2n-r)\frac{(2n-r-1)!}{n!(n-r-1)!} \\
&= (n-r)2^{r-2n}\frac{(2n-r)!}{n!(n-r)!} \\
&= [n-r]_1 P_{n,r}.
\end{aligned}$$

Now suppose that equation (3) holds for k. We wish to show that

$$[n-r]_{k+1}P_{n,r} = 2^{-k-1}F_{n,k+1}(r+k+1)P_{n,r+k+1}. \tag{4}$$

We first observe that

$$(n-r-k)\binom{2n-(r+k)}{n} = (2n-r-k)\binom{2n-(r+k+1)}{n}.$$

Furthermore,

$$(2n-r-k)F_{n,k}(r+k) = (2n-r-k)\left[\prod_{i=1}^{k}(2n+i)-(r+k)\right]$$

$$= (2n-r-k)\left[\prod_{i=2}^{k+1}(2n+i)-(r+k+1)\right]$$

$$= F_{n,k+1}(r+k+1).$$

Combining this with Lemma 2.2, we have

$$(n-r-k)F_{n,k}(r+k)P_{n,r+k} = 2^{(r+k)-2n}F_{n,k}(r+k)(n-r-k)\binom{2n-(r+k)}{n}$$

$$= 2^{-1}2^{(r+k+1)-2n}F_{n,k+1}(r+k+1)\binom{2n-(r+k+1)}{n}$$

$$= 2^{-1}F_{n,k+1}(r+k+1)P_{n,r+k+1}.$$

We complete the argument by applying the inductive hypothesis to the left hand side of (4).

$$\begin{aligned}[n-r]_{k+1}P_{n,r} &= (n-r-k)[n-r]_kP_{n,r}\\ &= 2^{-k}(n-r-k)F_{n,k}(r+k)P_{n,r+k}\\ &= 2^{-k-1}F_{n,k+1}(r+k+1)P_{n,r+k+1},\end{aligned}$$

as desired.

3. Expected Number of Returns and Other Moments. Consider the symmetric random walk (SRW) on $\mathbb{Z}$. The usual definition is as follows: Let $X_1, X_2, X_3, \ldots$ be independent, identically distributed random variables each taking the values $+1$ and -1 with probability $1/2$. Let $S_0 = 0$ and define S_k for $k \geq 1$ inductively by $S_k = S_{k-1} + X_k$. By an *N-step symmetric random walk* we mean the ordered $(N+1)$-tuple $\langle S_0, S_1, \ldots, S_N \rangle$. By a *symmetric random walk* we mean the infinite tuple $\langle S_0, S_1, \ldots \rangle$.

Polya's Theorem guarantees that with probability 1, we have $S_k = 0$ infinitely often in a SRW. Since this can only occur for even values of k, we are interested in the random variable R which counts the *number of returns to the origin* in a simple random walk of length $2n$; that is, the number of indices k, $1 \leq k \leq 2n$ for which $S_k = 0$.

Theorem 3.1. Let R denote the number of returns to the origin in a $2n$-step SRW. Then
$$P[R = r] = P_{n,r}.$$
In particular,
$$\sum_{r=0}^{n} P_{n,r} = 1.$$

Proof. The proof is given in the second edition of Feller, Volume 1 [3] in an optional section on pages 81–83. In the third and final edition of Feller's masterpiece, it is relegated to the exercises at the end of chapter III.

Theorem 3.2. Let R be defined as in Theorem 3.1. Then the mean μ_n and variance σ^2 of R are as follows:

$$\mu_n = E[R] = (2n+1)\binom{2n}{n}2^{-2n} - 1 \tag{5}$$

$$E[R^2] = 2n - 3\mu_n \tag{6}$$
$$\text{and } \sigma^2 = 2n - \mu_n(3 + \mu_n). \tag{7}$$

Proof. Let $k = 1$ in Theorem 2.3. Then we have

$$(n-r)P_{n,r} = \frac{1}{2}\left[(2n+1)-(r+1)\right]P_{n,r+1}$$

$$= \frac{2n+1}{2}P_{n,r+1} - \frac{r+1}{2}P_{n,r+1}.$$

Summing both sides, we have

$$n\sum_{r=0}^{n} P_{n,r} - \sum_{r=0}^{n} rP_{n,r} = \frac{2n+1}{2}\sum_{r=0}^{n} P_{n,r+1} - \frac{1}{2}\sum_{r=0}^{n}(r+1)P_{n,r+1}.$$

Since $\sum_{r=0}^{n} P_{n,r} = 1$ we have

$$\sum_{r=0}^{n} rP_{n,r} = n - \frac{2n+1}{2}\left[1 - P_{n,0} + P_{n,n+1}\right] + \frac{1}{2}\left[\sum_{r=0}^{n} rP_{n,r} + (n+1)P_{n,n+1}\right].$$

Since $P_{n,n+1} = 0$, we may solve this to obtain

$$2\sum_{r=0}^{n} rP_{n,r} = 2n - (2n+1)\left[1 - P_{n,0}\right] + \sum_{r=0}^{n} rP_{n,r}.$$

Thus,

$$\sum_{r=0}^{n} rP_{n,r} = 2n - (2n+1)\left[1 - P_{n,0}\right]$$

$$= (2n+1)P_{n,0} - 1.$$

Therefore, by Lemma 2.1

$$\mu_n = (2n+1)\binom{2n}{n}2^{-2n} - 1$$

as desired.

For the variance, we consider the case $k = 2$ in Theorem 2.3.

$$(n-r)(n-r-1)P_{n,r} = \frac{1}{4}\left([(2n+1)-(r+2)]\cdot[(2n+2)-(r+2)]\right)P_{n,r+2}$$

$$\begin{aligned}&4(n^2-2nr-n+r^2+r)P_{n,r}\\&=\left[(2n+1)(2n+2)-(4n+3)(r+2)+(r+2)^2\right]P_{n,r+2}.\end{aligned}$$

Summing both sides, we have

$$4n^2\sum_{r=0}^{n}P_{n,r}-8n\sum_{r=0}^{n}rP_{n,r}-4n\sum_{r=0}^{n}P_{n,r}+4\sum_{r=0}^{n}r^2P_{n,r}+4\sum_{r=0}^{n}rP_{n,r}$$

$$=(2n+1)(2n+2)\sum_{r=0}^{n}P_{n,r+2}-(4n+3)\sum_{r=0}^{n}(r+2)P_{n,r+2}+\sum_{r=0}^{n}(r+2)^2P_{n,r+2}.$$

Simplifying, we have

$$\begin{aligned}4n^2-8n\mu_n-4n+4E[R^2]+4\mu_n &= (2n+1)(2n+2)\left[1-P_{n,0}-P_{n,1}\right]\\&\quad-(4n+3)\left[\mu_n-P_{n,1}\right]+E[R^2]-P_{n,1}.\end{aligned}$$

By Lemma 2.1, $P_{n,0}=P_{n,1}$, so we may solve in the following way.

$$\begin{aligned}3E[R^2] &= 4n+8n\mu_n-4\mu_n+6n+2-2(4n^2+6n+2)P_{n,0}\\&\quad-(4n+3)\mu_n+(4n+3)P_{n,0}-P_{n,0}\\&=6n+9-9P_{n,0}[2n+1].\end{aligned}$$

Using the formula for μ_n, we have,

$$E[R^2]=2n+3-3(\mu_n+1)=2n-3\mu_n.$$

Hence,

$$\sigma^2 = E[R^2] - \mu_n^2 = 2n - \mu_n(3 + \mu_n),$$

as desired.

Using Theorem 2.3 for larger values of k and the techniques employed in the proof of Theorem 3.2 we may compute, at least in principle, any of the moments of R. We quote some of them in the following theorem.

Theorem 3.3. Let R be defined as in Theorem 3.1. Then

$$\begin{aligned}
E[R^3] &= \mu_n(4n + 13) - 8n \\
E[R^4] &= 12n^2 + 46n - 5\mu_n(8n + 15) \\
E[R^5] &= \mu_n(32n^2 + 388n + 541) - 148n^2 - 332n \\
E[R^6] &= 120n^3 + 1728n^2 + 2874n - 7\mu_n(96n^2 + 584n + 669).
\end{aligned}$$

The derivation for the third and fourth moment is given in [2] pages 18–27. The calculations are long but rely entirely on elementary methods. We leave the proof of the fifth and sixth moment to the interested and patient reader.

The general theory of Markov processes implies that the expected number of returns to the origin for any random walk is a sum of transition probabilities; specifically

$$\sum P[S_{2k} = 0]$$

where k ranges from 1 to n in the case of a finite walk and over all positive integers in an infinite walk. This is the key to a modern proof of Polya's Theorem as in [1]. In the case of the $2n$-step SRW, we can give a combinatorial proof of this fact. We first observe that

$$P[S_{2k} = 0] = \binom{2k}{k} 2^{-2k}$$

since a walker returns to the origin after $2k$ steps by choosing precisely k steps to the right and k steps to the left. Thus,

$$P[S_{2k} = 0] = P_{k,0}$$

by the definition of $P_{n,r}$. We shall show that

$$\mu_n = \sum_{k=1}^{n} P_{k,0},$$

but first we need the following lemma.

Lemma 3.4. Let μ_n be defined as in Theorem 3.2. If $n \geq 1$ then

$$\mu_n = \mu_{n-1} + P_{n,0}.$$

Proof. By Theorem 3.2

$$\begin{aligned}
\mu_{n-1} &= (2n-1)\binom{2n-2}{n-1}2^{2-2n} - 1 \\
&= 4(2n-1)\frac{(2n-2)!}{(n-1)!(n-1)!}2^{-2n} - 1 \\
&= 4(2n-1)\frac{n^2(2n)!2^{-2n}}{2n(2n-1)(n!)^2} - 1 \\
&= \frac{2n(2n)!}{(n!)^2}2^{-2n} - 1 \\
&= 2nP_{n,0} - 1 = \mu_n - P_{n,0}.
\end{aligned}$$

Therefore,

$$\mu_n = \mu_{n-1} + P_{n,0}$$

as desired.

Theorem 3.5. If $n \geq 1$ then

$$\mu_n = \sum_{k=1}^{n} P_{k,0}.$$

Proof. By repeated applications of Lemma 3.4 we get

$$\begin{aligned}\mu_n &= \mu_{n-1} + P_{n,0}\\ &= \mu_{n-2} + P_{n-1,0} + P_{n,0}\\ &= \mu_{n-3} + P_{n-2,0} + P_{n-1,0} + P_{n,0}\\ &= \cdots = \mu_0 + \sum_{k=1}^{n} P_{k,0}.\end{aligned}$$

But $\mu_0 = 0$, therefore,

$$\mu_n = \sum_{k=1}^{n} P_{k,0}.$$

4. Asymptotic Results.

Proposition 4.

$$\binom{2n}{n} 2^{-2n} \sim (\pi n)^{-\frac{1}{2}}. \tag{8}$$

Proof. Using Stirling's formula

$$n! \sim (2\pi n)^{\frac{1}{2}} n^n e^{-n}$$

we have

$$\begin{aligned}\binom{2n}{n} 2^{-2n} &\sim \left[\frac{(2\pi)^{\frac{1}{2}}(2n)^{2n+\frac{1}{2}} e^{-2n}}{2\pi n^{2n+1} e^{-2n}}\right] 2^{-2n}\\ &\sim (2\pi)^{-\frac{1}{2}} n^{-\frac{1}{2}} 2^{\frac{1}{2}} = (\pi n)^{-\frac{1}{2}}\end{aligned}$$

as desired.

Relation (8), combined with the results of Section 3, implies that

$$E[R^k] \sim cn^{\frac{k}{2}}$$

for $k = 1, 2, \ldots 6$, where c depends only on the power k.

It is clear from the details of Theorem 3.3 that the methods employed for calculating $E[R^k]$ can be extended for any value of k, giving a polynomial in n with coefficients that are either integers or integral multiples of μ_n. More precisely, that

$$E[R^{2k-1}] = \mu_n(a_{k-1}n^{k-1} + \cdots + a_0) - (b_{k-1}n^{k-1} + \cdots + b_1 n) \quad k = 1, 2, 3, \ldots$$

and

$$E[R^{2k}] = (\alpha_k n^k + \cdots + \alpha_1 n) - \mu_n(\beta_{k-1}n^{k-1} + \cdots + \beta_0) \quad k = 1, 2, 3, \ldots$$

where $a_{k's}, b_{k's}, \alpha_{k's}$ and $\beta_{k's}$ are positive integers.

A proof of this conjecture remains elusive, as does a closed form for the coefficients. Nevertheless, the details of the construction give one a degree of confidence in asserting the conjecture that

$$E[R^k] \sim c_k n^{\frac{k}{2}}$$

for all positive integers k.

References

1. P. Billingsley, *Probability and Measure,* 2nd ed., Wiley, New York, 1986.
2. L. Chrysafi, *Returns to the Origin for Random Walks,* doctoral thesis, UMI Dissertation Services, 1997.
3. W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. 1, 2nd ed., Wiley, New York, 1957.
4. W. Feller, *An Introduction to Probability Theory and Its Applications,* Vol. 1, 3rd ed., Wiley, New York, 1968.

Loucas A. Chrysafi
Department of Mathematics
SUNY Farmingdale
Farmingdale, NY 11735
email: chrysala@farmingdale.edu

R. E. Bradley
Department of Mathematics & Computer Science
Adelphi University
Garden City, NY 11530-4213
email: bradley@panther.adelphi.edu