SIAM REVIEW
Vol. 35, No. 2, pp. 294–297, June 1993

005

A COMBINATORIAL IDENTITY AND THE WORLD SERIES*

TAMÁS LENGYEL†

Abstract. In this note the author gives a simple probabilistic proof of a combinatorial identity by calculating the winning probability in the World Series. The winning probabilities and the expected length of the championship series are given by the applications of the identity and its generalization.

Key words. combinatorial identities, asymptotic identities, championship series

AMS(MOS) subject classifications. 60C05, 05A19, 05A15, 05A16, 90D60

1. Introduction. We give a probabilistic proof of a classical combinatorial identity and find the winning probabilities and the expected length of World Series–type games. There are Teams A and B to play no more than $2n-1$ games to decide a champion. Each single game ends in no tie. These rules apply to the World Series and the NBA play–offs as well with $n=4$. Suppose that in each game Team A (Team B) has a probability p ($q=1-p$) of winning and the outcomes of the games are independent. The winner is the first team to collect n victories. The length of the series is denoted by the random variable $W_n(p)$. The expected length, $EW_n(p)$, of the series can be determined by classical combinatorial or hypergeometric summation [2] if $p=1/2$. For short, let $EW_n = EW_n(1/2)$. The problem of determining EW_4 appears, for example, in [3, Question 3.8.14, p. 162] in the context of the usual *best* 4 *of* 7 series. In this note we use a simple method to calculate $EW_n(p)$. An asymptotic formula is given for $EW_n(p)$ in identity (7). Some related results have been proven in [5].

We start with the classical combinatorial identity [2, (5.20) p. 167]

$$\sum_{k=0}^{n} \binom{n+k}{k} 2^{-k} = 2^n \quad (n \geq 0) \tag{1}$$

and prove it by calculating the probabilities of winning a particular championship series. In this series the winner team must accumulate $n+1$ victories. Assume that the two teams are equally likely to win in each game, i.e. $p=q=1/2$. Let $p(n+1,k)$ denote the probability that Team A becomes the champion after winning the $n+k+1^{st}$ game where $0 \leq k < n+1$. In the first $n+k$ games, Teams A and B accumulate n and k victories, respectively, with probability $a(n,k)$. We have $a(n,k) = \binom{n+k}{k} 2^{-(n+k)}$. Clearly, $p(n+1,k) = \frac{1}{2}\, a(n,k)$, therefore Team A wins with probability

$$\sum_{k=0}^{n} p(n+1,k) = \sum_{k=0}^{n} \frac{1}{2} \binom{n+k}{k} 2^{-(n+k)}. \tag{2}$$

Either Team A or B wins; hence the winning probabilities must add up to 1, i.e., $\sum_{k=0}^{n} \binom{n+k}{k} 2^{-(n+k)} = 1$. ∎

*Received by the editors November 29, 1991; accepted for publication (in revised form) July 3, 1992.

†Department of Mathematics, Occidental College, 1600 Campus Dr., Los Angeles, CA 90041.

With little extra effort we can calculate the expected length of the series by using identity (1) again.

THEOREM 1. *The expected length, EW_n, of the series is* $2n\left(1-\binom{2n}{n}2^{-2n}\right)$ *if* $p=q=1/2$.

Proof. By the previous method, the series lasts $n+k$ games with probability $\binom{n+k-1}{k}2^{-(n+k-1)}$. We get

$$EW_n=\sum_{k=0}^{n-1}(n+k)\binom{n+k-1}{k}2^{-(n+k-1)} \tag{3}$$

for the expected length of the series. We shall use the identity $(n+k)\binom{n+k-1}{k}=n\binom{n+k}{k}$ to rewrite equation (3)

$$EW_n=2n\sum_{k=0}^{n-1}\binom{n+k}{k}2^{-(n+k)}=2n\left\{\sum_{k=0}^{n}\binom{n+k}{k}2^{-(n+k)}-\binom{2n}{n}2^{-2n}\right\}.$$

By identity (1), it follows that the expected value is $2n\left(1-\binom{2n}{n}2^{-2n}\right)$. ∎

For example, if $n=4$ then we get $8(1-\frac{70}{256})=5.8125$ games on the average. In general, EW_n is asymptotically $2n\{1-\frac{1}{\sqrt{\pi n}}\}$. Hence, at the end of the series, the winning team will collect n victories while the other one wins asymptotically $n-c\sqrt{n}$ games, on the average, where $c=2/\sqrt{\pi}$. Note that if the number N of games is fixed in advance then after N games the two teams will be asymptotically $c\sqrt{N/2}$ victories apart (cf. [1, Problem 35, p. 241] or [6, Putnam 1974, Problem A-4]), on the average. In fact, if $N\sim EW_n$ then the expected number of victories apart is asymptotically $c\sqrt{n}$ as we noted.

We generalize Theorem 1 for arbitrary p.

THEOREM 2. *For the probabilities,* P_n^A *and* P_n^B*, of winning the championship by Teams A and B, respectively, and the expected length,* $EW_n(p)$*, of the series*

$$P_n^A=p+\frac{(p-q)}{2}\sum_{k=1}^{n-1}\binom{2k}{k}(pq)^k, \tag{4}$$

$$P_n^B=q+\frac{(q-p)}{2}\sum_{k=1}^{n-1}\binom{2k}{k}(pq)^k, \tag{5}$$

and

$$EW_n(p)=2n\left\{1-\binom{2n}{n}(pq)^n-\frac{(p-q)^2}{4pq}\sum_{k=1}^{n}\binom{2k}{k}(pq)^k\right\}. \tag{6}$$

Remark. It follows from identity (6) that

$$\lim_{n\to\infty}\frac{1}{2n}\Big(EW_n(1/2)-EW_n(p)\Big)=\frac{|p-q|\,(1-|p-q|)}{4pq}, \tag{7}$$

since $\sum_{k=0}^{\infty}\binom{2k}{k}x^k=1/\sqrt{1-4x}$, for $|x|<1/4$ (see [2]). Similarly, identities (4) and (5) imply that for $p>q$, $\lim_{n\to\infty}P_n^A=1$ and for $p<q$, $\lim_{n\to\infty}P_n^B=1$.

If $p = q = 1/2$ then $P_n^A = P_n^B = 1/2$. Notice that the limit in identity (7) can be expressed as $1 - \frac{1}{2p}$ if $p \geq 1/2$.

Proof. The probability that the series ends at the end of the $n + k^{th}$ game is $\binom{n+k-1}{k}\left(p^n q^k + q^n p^k\right)$ and

$$P_n^A = p^n \sum_{k=0}^{n-1} \binom{n+k-1}{k} q^k \quad \text{and} \quad P_n^B = q^n \sum_{k=0}^{n-1} \binom{n+k-1}{k} p^k.$$

Now $P_n^A + P_n^B = 1$ which gives a generalization of identity (1). Obviously, $P_1^A = p$ and $P_1^B = q$. Similarly to the proof of Theorem 1,

$$EW_n(p) = np^n \sum_{k=0}^{n-1} \binom{n+k}{k} q^k + nq^n \sum_{k=0}^{n-1} \binom{n+k}{k} p^k. \tag{8}$$

Let EW_n^A denote the first part of the sum on the right hand side of equation (8). The simple relation between P_n^A and EW_n^A is given by

$$EW_n^A = \frac{n}{p} P_{n+1}^A - n \binom{2n}{n} (pq)^n$$

which implies that

$$EW_n(p) = n \left\{ \frac{P_{n+1}^A}{p} + \frac{P_{n+1}^B}{q} \right\} - 2n \binom{2n}{n} (pq)^n. \tag{9}$$

We obtain the relation (4) for P_n^A by taking the difference $P_{n+1}^A - P_n^A$. This step is motivated by a similar one used in the solution to Problem 44, "winning an unfair game,"in [4]. The difference expresses the advantage of Team A in the terms of the probability of winning the championship if two extra games are played. The status of Team A as a winner or loser does not differ at the end of the extended series from that in the original series if Team A has already won more than n games (winner) or less than $n-1$ games (loser) in the first $2n-1$ games. The difference comes from two sources. The probability that Team A becomes the champion after winning only $n-1$ of the $2n-1$ games and then winning the next two games will increase the winning probability. On the other hand, Team A will finish as the loser if it has accumulated exactly n victories and then it falls behind by losing the last two games. We use the identity $\binom{2n}{n} = 2\binom{2n-1}{n-1}$ and obtain

$$\begin{aligned} P_{n+1}^A - P_n^A = p^2 \binom{2n-1}{n-1} p^{n-1} q^n - q^2 \binom{2n-1}{n-1} p^n q^{n-1} &= (p-q) \binom{2n-1}{n-1} (pq)^n \\ &= \frac{p-q}{2} \binom{2n}{n} (pq)^n. \end{aligned}$$

This yields relation (4) and identity (5) follows similarly. Since $\left(1 - (p/q + q/p)/2\right) = -\frac{(p-q)^2}{2pq}$, the formula (9) implies identity (6). ∎

REFERENCES

[1] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. I, 3^{rd} edition, Wiley, New York, 1968.

[2] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1988.
[3] R. J. Larsen and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, Prentice-Hall, 1986.
[4] F. Mosteller, *Fifty Challenging Problems in Probability with Solutions*, Addison-Wesley, Reading, MA, 1965.
[5] P. Tannenbaum, *Handicapping a championship series with elementary symmetric polynomials*, UMAP Journal, (10)1989, pp. 115–135.
[6] *The William Putnam Mathematical Competition, Problems and Solutions: 1965-1984,* Eds. G. L. Alexanderson, L. F. Klosinski, and L. C. Larson, MAA, Washington, 1980.