Wald’s Identity and Geometric Expectation

A geometric random variable \( N \) is the number of independent Bernoulli trials until the first success. If each trial has a probability \( p \) of success, then the probability that \( k \) trials are needed is \( P(N = k) = (1-p)^{k-1}p \) for \( k = 1, 2, 3, \ldots \).

Calculating the expected value \( E[N] \) involves evaluating the power series \( \sum_{k=1}^{\infty} k(1-p)^{k-1}p \), which can be identified as the derivative of a geometric series. Hong [1] surveys eight different ways to calculate this expected value.

We offer a shorter derivation that formalizes the intuition that the expectation should be the inverse of the success probability. It relies on Wald’s identity for the sum of a random number \( N \) of i.i.d. random variables \( X_1, X_2, \ldots \):

\[
E\left[ \sum_{i=1}^{N} X_i \right] = E[N] E[X_1]. \tag{1}
\]

Most probability students learn a special case of this identity, when \( N \) is independent of \( X_1, X_2, \ldots \) [3, p. 340], but the identity holds more generally, as long as \( N \) is a stopping time—that is, \( \{N = n\} \) is independent of \( X_{n+1}, X_{n+2}, \ldots \) for all \( n \) [2, p. 105]. In other words, the decision to stop at time \( n \) can only depend on the values seen so far; it cannot depend on future values. With stopping times and Wald’s identity, the geometric expectation is immediate.

**Proposition.** \( E[N] = 1/p \).

**Proof.** Let \( X_1, X_2, \ldots \) be i.i.d. Bernoulli\((p)\) random variables representing the trials. The geometric random variable \( N \) is a stopping time and can be formally defined in terms of the \( X_i \)’s as \( N := \inf\{n : X_n = 1\} \).

By definition, \( \sum_{i=1}^{N} X_i = 1 \) (almost surely), since every term in the sum is 0 except for the last, which is 1. So the expectation of the random sum is 1 as well. But we can also expand the expectation using (1). Equating the two expressions,

\[
1 = E\left[ \sum_{i=1}^{N} X_i \right] = E[N] E[X_1]. \tag{2}
\]

Solving for \( E[N] \), we see that \( E[N] = 1/E[X_1] = 1/p \).

Similar arguments appear in renewal theory. For example, Ross [2, p. 104-106] uses (1) to calculate the expected number of coin flips to get 10 heads.

**REFERENCES**


—Submitted by Dennis L. Sun, California Polytechnic State University