A PROBABILISTIC PROOF THAT $\sum_{J=1}^{N} H_J^{(S)} = (N+1)H_N^{(S)} - H_N^{(S-1)}$

Reza Farhadian
Department of Statistics, Razi University, Kermanshah, Iran
farhadian.reza@yahoo.com

Received: 8/15/22, Revised: 2/23/23, Accepted: 4/6/23, Published: 4/24/23

Abstract

Let $H_n^{(s)}$ $(s \geq 0)$ be the nth generalized harmonic number. In this note, we provide a probabilistic proof for the familiar sum identity $\sum_{j=1}^{n} H_j^{(s)} = (n+1)H_n^{(s)} - H_n^{(s-1)}$.

1. Introduction

The harmonic number H_n is defined to be the partial sum of the harmonic series, i.e.,

$$H_n = \sum_{k=1}^{n} \frac{1}{k}.$$

The first 10 terms of the sequence of harmonic numbers are

$$1, \frac{3}{2}, \frac{11}{6}, \frac{25}{12}, \frac{137}{60}, \frac{49}{20}, \frac{363}{140}, \frac{761}{280}, \frac{7129}{2520}, \frac{7381}{2520}.$$

Since the harmonic series diverges, it is clear that H_n can get arbitrarily large, although it does so quite slowly. For instance, some large harmonic numbers are

$$\begin{aligned} H_{1000000} &\approx 14.3927, \\ H_{2000000} &\approx 15.0858, \\ H_{3000000} &\approx 15.4913, \\ H_{4000000} &\approx 15.7790. \end{aligned}$$

Harmonic numbers are also defined in a more generalized form. The generalized harmonic numbers $H_n^{(s)}$ of order s are defined by

$$H_n^{(s)} = \sum_{k=1}^{n} \frac{1}{k^s} \quad (s \geq 0).$$

DOI: 10.5281/zenodo.7859025

Generalized harmonic numbers are also denoted by $H_{n,s}$. In the special case of $s = 0$ we have $H_n^{(0)} = n$, and the special case of $s = 1$ gives $H_n^{(1)} = H_n$.

Harmonic numbers satisfy many interesting properties. Greene and Knuth [2, p.10] listed some commonly used identities. One of them is the following well-known sum identity:

$$\sum_{j=1}^{n} H_j = (n+1)H_n - n. \tag{1}$$

In general, this identity can be defined for generalized harmonic numbers as follows (see, e.g., [4, p. 853 and 856]):

$$\sum_{j=1}^{n} H_j^{(s)} = (n+1)H_n^{(s)} - H_n^{(s-1)} \quad (s \geq 1). \tag{2}$$

Since $H_n^{(1)} = H_n$ and $H_n^{(0)} = n$, the Identity (2) gives Identity (1) when $s = 1$.

In this note, we will give a probabilistic proof of Identity (2). To this end, in subsequent sections, the following notation will be used: $\mathbb{N}$ denotes the set of all natural numbers; $\Pr(X = x)$ denotes the probability mass function of a discrete random variable X; and $E[X]$ denotes the expected value of the random variable X.

2. Zipfian Distribution (Zipf's law)

A random variable X has the *Zipfian distribution* with parameters $s \geq 0$ and $n \in \mathbb{N}$, if its probability mass function is given by

$$\Pr(X = x) = \begin{cases} \frac{1}{x^s H_n^{(s)}} & \text{if } x \in \{1, 2, \ldots, n\}, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $H_n^{(s)}$ is the nth generalized harmonic number of order s.

Usually, the shorthand $X \sim Zipf(s, n)$ is used to indicate that the random variable X has Zipfian distribution with parameters s and n. In simple terms, the Zipfian distribution predicts that out of a population of n elements, the frequency of rank x is given by (3), where s is the value of the exponent characterizing the distribution. This prediction is called Zipf's law. For more details, see, for example, [1, p. 373].

3. Probabilistic Proof of Identity (2)

Proof of Identity (2). Let X be a random variable taking values among the non-negative integers. We recall the well-known tail sum formula for the expectation of

X (see, e.g., [3]), namely,

$$E[X] = \sum_{x=0}^{\infty} \Pr(X > x).$$

Since we have $E[X] = \sum_x x \Pr(X = x)$, the above identity may be written as

$$\sum_{x=0}^{\infty} x \Pr(X = x) = \sum_{x=0}^{\infty} \left(1 - \sum_{y=0}^{x} \Pr(X = y)\right). \tag{4}$$

If X has support in $\{0, 1, \ldots, n\}$, then Equation (4) gives

$$\begin{aligned} n + 1 &= \sum_{x=0}^{n} x \Pr(X = x) + \sum_{x=0}^{n} \sum_{y=0}^{x} \Pr(X = y) \\ &= \sum_{x=1}^{n} x \Pr(X = x) + \sum_{x=0}^{n} \sum_{y=0}^{x} \Pr(X = y). \end{aligned} \tag{5}$$

Now, suppose that X is a Zipfian random variable with parameters $n \in \mathbb{N}$ and $s \geq 0$. By using the probability mass function of X (see (3)) and applying (5), we have (note that $\Pr(X = 0) = 0$)

$$n + 1 = \sum_{x=1}^{n} \frac{x}{x^s H_n^{(s)}} + \sum_{x=1}^{n} \sum_{y=1}^{x} \frac{1}{y^s H_n^{(s)}} = \frac{1}{H_n^{(s)}} \left(\sum_{x=1}^{n} \frac{1}{x^{s-1}} + \sum_{x=1}^{n} \sum_{y=1}^{x} \frac{1}{y^s} \right);$$

that is,

$$(n + 1) H_n^{(s)} = \sum_{x=1}^{n} \frac{1}{x^{s-1}} + \sum_{x=1}^{n} \sum_{y=1}^{x} \frac{1}{y^s} = H_n^{(s-1)} + \sum_{x=1}^{n} H_x^{(s)}.$$

This completes the proof. □

Acknowledgements. The author would like to thank the Editor and the anonymous reviewer for their valuable suggestions.

References

[1] K. Funatsu, *New Fundamental Technologies in Data Mining*, IntechOpen, Rijeka, 2011.

[2] D.H. Greene and D.E. Knuth, *Mathematics for the Analysis of Algorithms*, Third ed., Birkhäuser, Boston, Basel and Berlin, 1990.

[3] A. Lo, Demystifying the integrated tail probability expectation formula, *Amer. Statist.* **73** (2019), 367–374.

[4] J. Spiess, Some identities involving harmonic numbers, *Math. Comp.* **55** (1990), 839–863.