Wald's Identity

Sinho Chewi

Fall 2017

1 Motivation

Let $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. random variables with finite mean. Let N be a random variable taking values in the positive integers, which we interpret as a *random time*. We are interested in computing the expectation of the random sum $X_1 + \cdots + X_N$. Notice that not only are the individual terms in the summation random, the number of terms is itself a random variable.

Random sums appear often in the study of stochastic processes. For example, in a Poisson process with interarrival times $X_1, X_2, X_3, \dots$ we may define N to be the smallest positive integer n such that $X_n < \tau$, for some fixed $\tau > 0$, in which case $\mathbb{E}[X_1 + \cdots + X_N]$ is the expected time we must wait until we observe an interarrival time of duration less than τ units of time. How can we calculate this expectation?

One might guess that $\mathbb{E}[X_1 + \cdots + X_N] = \mathbb{E}[X_1]\,\mathbb{E}[N]$, which conveys the intuition that we expect $\mathbb{E}[N]$ terms in the summation, each of which has an expected value $\mathbb{E}[X_1]$, so the entire sum is, in expectation, $\mathbb{E}[X_1]\,\mathbb{E}[N]$.

In fact, the intuition given above is not bad, but we must be careful when applying it. To start, let us try to mathematically justify our intuition. Using the law of iterated expectation, we might attempt to write

$$\mathbb{E}[X_1 + \cdots + X_N] = \mathbb{E}[\mathbb{E}(X_1 + \cdots + X_N \mid N)] = \mathbb{E}\big[N\,\mathbb{E}[X_1]\big] = \mathbb{E}[X_1]\,\mathbb{E}[N].$$

The above derivation is appealing, but it makes hidden assumptions. In particular, the intermediate step $\mathbb{E}(X_1 + \cdots + X_N \mid N) = N\,\mathbb{E}[X_1]$ is not justified unless we know that N is independent of $X_1, X_2, X_3, \dots$. To understand this, here is the same approach given above, but written out in more detail.

$$\begin{aligned}
\mathbb{E}[X_1 + \cdots + X_N] &= \sum_{n=1}^{\infty} \mathbb{E}[X_1 + \cdots + X_N \mid N = n]\mathbb{P}(N = n) \\
&= \sum_{n=1}^{\infty} \mathbb{E}[X_1 + \cdots + X_n \mid N = n]\mathbb{P}(N = n) \\
&\overset{(?)}{=} \sum_{n=1}^{\infty} \mathbb{E}[X_1 + \cdots + X_n]\mathbb{P}(N = n) = \mathbb{E}[X_1] \sum_{n=1}^{\infty} n\mathbb{P}(N = n) = \mathbb{E}[X_1]\,\mathbb{E}[N].
\end{aligned}$$

In the step labeled (?), the statement $\mathbb{E}[X_1 + \cdots + X_n \mid N = n] = \mathbb{E}[X_1 + \cdots + X_n]$ is assuming that the conditioning on $\{N = n\}$ does not change the distribution of $X_1, \dots, X_n$, but this assumption can be unfounded.

Example 1. Let $X_1, X_2, X_3, \ldots \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(1/2)$, and let N be the smallest non-negative integer n such that $X_{n+1} = 1$, that is, $N := \min\{n \in \mathbb{N} : X_{n+1} = 1\}$. Also, we define the random $X_1 + \cdots + X_N$ to be 0 when $N = 0$. Then, $\mathbb{E}[X_1] = 1/2$ and since N is a shifted geometric random variable, i.e., $N + 1 \sim \text{Geometric}(1/2)$, we can also compute $\mathbb{E}[N] = 1$. Thus, we have $\mathbb{E}[X_1]\,\mathbb{E}[N] = 1/2$, but the random sum $X_1 + \cdots + X_N$ is always 0. This is because X_{N+1} is the first of the random variables which is non-zero, by the definition of N. This example demonstrates that $\mathbb{E}[X_1]\,\mathbb{E}[N] = \mathbb{E}[X_1 + \cdots + X_N]$ need *not* always hold when N and $X_1, X_2, X_3, \ldots$ are not independent.

However, under the assumption that $X_1, X_2, X_3, \ldots$ and N are independent, we *can* prove the result we desire. We will write the proof as a warm-up for the main result discussed in the next section.

Theorem 1. *If $X_1, X_2, X_3, \ldots$ are i.i.d. with finite mean, and N is a random variable taking values in the positive integers, such that N is independent of $(X_1, X_2, X_3, \ldots)$ and $\mathbb{E}[N] < \infty$, then $\mathbb{E}[X_1 + \cdots + X_N] = \mathbb{E}[X_1]\,\mathbb{E}[N]$.*

Proof. Observe that $X_1 + \cdots + X_N = \sum_{n=1}^{\infty} X_n \mathbb{1}\{N \ge n\}$. Thus, [1]

$$\begin{aligned}\mathbb{E}[X_1 + \cdots + X_N] &= \mathbb{E}\Big[\sum_{n=1}^{\infty} X_n \mathbb{1}\{N \ge n\}\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n \mathbb{1}\{N \ge n\}] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[X_n]\,\mathbb{E}[\mathbb{1}\{N \ge n\}] = \mathbb{E}[X_1] \sum_{n=1}^{\infty} \mathbb{P}(N \ge n) = \mathbb{E}[X_1]\,\mathbb{E}[N].\end{aligned}$$

We have used the independence of N and $(X_1, X_2, X_3, \ldots)$, and the last step follows from the tail sum formula (Theorem 3). □

2 Stopping Times & Wald's Identity

Our goal now is to extend our results to situations when N is not necessarily independent of the sequence $X_1, X_2, X_3, \ldots$. After all, independence is quite a strong assumption, and it is often violated when N is actually defined in terms of the process $X_1, X_2, X_3, \ldots$.

The random time N defined in Example 1 is quite strange: it "peeks" into the future in that the event $\{N = n\}$, for any positive integer n, depends on the *next* random variable X_{n+1}. Intuitively, if N carries information about the future, then conditioning on $\{N = n\}$ may change the distributions of $X_1, \ldots, X_n$. However, if the event $\{N = n\}$ does not carry any new information besides that which is already conveyed through $X_1, \ldots, X_n$, then conditioning on $\{N = n\}$ should not disturb $X_1, \ldots, X_n$ too much. In other words, if we remove from consideration random times N which require clairvoyance, then our desired identity for $\mathbb{E}[X_1 + \cdots + X_N]$ holds.

To convert our intuition into mathematical precision, we will make a definition. The random variable N is called a **stopping time** if N takes on values in the positive integers

[1] We will not address here the subtle technical point about the validity of exchanging the expectation and the infinite summation: $\mathbb{E}[\sum_{n=1}^{\infty} X_n \mathbb{1}\{N \ge n\}] = \sum_{n=1}^{\infty} \mathbb{E}[X_n \mathbb{1}\{N \ge n\}]$. Instead, take our word that under our assumptions, we have committed no fouls.

and for each positive integer n, the event $\{N \le n\}$ is completely determined by $X_1, \ldots, X_n$. In other words, the random variable $\mathbb{1}\{N \le n\}$ is a function of $X_1, \ldots, X_n$.

The definition says that if N is a stopping time, then after observing $X_1, \ldots, X_n$, we can know whether or not the random time N has already happened. In contrast, the random time N in Example 1 is *not* a stopping time, because the event $\{N \le n\}$ also depends on knowledge of X_{n+1}.

Note: Stopping times actually play a major role in the study of stochastic processes, beyond the specific application we present here.

Theorem 2 (Wald's Identity). *If $X_1, X_2, X_3, \ldots$ be i.i.d. with finite mean, and N is a stopping time with $\mathbb{E}[N] < \infty$, then $\mathbb{E}[X_1 + \cdots + X_N] = \mathbb{E}[X_1]\,\mathbb{E}[N]$.*

Proof. As before, we start with $X_1 + \cdots + X_N = \sum_{n=1}^{\infty} X_n \,\mathbb{1}\{N \ge n\}$. Thus,

$$\begin{aligned}\mathbb{E}[X_1 + \cdots + X_N] &= \mathbb{E}\Big[\sum_{n=1}^{\infty} X_n \,\mathbb{1}\{N \ge n\}\Big] = \sum_{n=1}^{\infty} \mathbb{E}[X_n \,\mathbb{1}\{N \ge n\}] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{E}(X_n \,\mathbb{1}\{N \ge n\} \mid X_1, \ldots, X_{n-1})].\end{aligned}$$

Now, note that $\mathbb{1}\{N \ge n\} = 1 - \mathbb{1}\{N \le n-1\}$ is a function of $X_1, \ldots, X_{n-1}$ because the event $\{N \le n-1\}$ is completely determined by $X_1, \ldots, X_{n-1}$; therefore, we can pull the indicator outside of the conditioning.

$$\begin{aligned}&= \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}\{N \ge n\}\,\mathbb{E}(X_n \mid X_1, \ldots, X_{n-1})] = \sum_{n=1}^{\infty} \mathbb{E}\big[\mathbb{1}\{N \ge n\}\,\mathbb{E}[X_n]\big] \\ &= \mathbb{E}[X_1] \sum_{n=1}^{\infty} \mathbb{P}(N \ge n) = \mathbb{E}[X_1]\,\mathbb{E}[N]\end{aligned}$$

where we have again used the tail sum formula (Theorem 3). □

Now, it remains to prove:

Theorem 3 (Tail Sum Formula). *If N is a random variable taking values in $\mathbb{N}$, then*

$$\mathbb{E}[N] = \sum_{n=1}^{\infty} \mathbb{P}(N \ge n).$$

Proof. The expectation of N is:

$$\begin{aligned}\mathbb{E}[N] &= \mathbb{P}(N = 1) + 2\mathbb{P}(N = 2) + 3\mathbb{P}(N = 3) + \cdots \\ &= \begin{cases}\mathbb{P}(N = 1) + \\ \mathbb{P}(N = 2) + \mathbb{P}(N = 2) + \\ \mathbb{P}(N = 3) + \mathbb{P}(N = 3) + \mathbb{P}(N = 3) + \cdots .\end{cases}\end{aligned}$$

Now, sum down the columns to obtain $\mathbb{E}[N] = \mathbb{P}(N \ge 1) + \mathbb{P}(N \ge 2) + \mathbb{P}(N \ge 3) + \cdots$. □