

Wald's Identity

by Jeffery Hein

Dartmouth College, Math 100

1. Introduction

Given random variables X_1, X_2, X_3, \dots with common finite mean and a stopping rule τ which may depend upon the given sequence, we will be interested in finding conditions such that the following holds:

$$\mathbb{E}(X_1 + \dots + X_\tau) = \mathbb{E}(X_1)\mathbb{E}(\tau); \quad (1)$$

this is known as Wald's Identity. The beauty of Wald's Identity lies in its simplicity and intuitive nature: anyone could conjecture (1) after a few experiments. It wasn't until Wald established a framework in [2, 3] that our intuition was shown to be valid. Up to that point, intuition and experimentation suggested the validity of Wald's Identity, but intuition and experiment could only take us so far in the absence of rigor.

In this paper, we seek modest conditions for which Wald's Identity will hold, three in total. These conditions will by no means be the weakest conditions possible, but nonetheless they will provide us a starting point. To arrive at these conditions, we will consider three examples, each failing to achieve one condition while satisfying the remaining two. In doing so, we will justify each hypothesis and emphasize their mutual dependence.

2. Interpretations

As we'll see in the forthcoming examples, there are two important ways to interpret Wald's Identity: gambling and random walks.

In the gambling interpretation, we take each X_n as the incremental gain or loss of a gambler during his n -th game, so that $X_1 + \dots + X_n$ is the gambler's aggregate winnings up to and including the n -th game. We can then treat τ as the point at which the gambler decides to stop gambling. If the gambler plays a fair game, then $\mathbb{E}(X_1) = 0$ and so $\mathbb{E}(X_1 + \dots + X_\tau) = 0$ by Wald; in other words, no stopping strategy will give the gambler an advantage. The benefit of this interpretation is that it actually suggests one of the three conditions under which we will show that (1) holds: the gambler may not rely upon a stopping rule which depends upon the outcome of games which he has not yet played. This is intuitively obvious, no one would accept a bet against someone that can see infinitely far into the future.

In the random walk interpretation, we can treat $X_1 + \dots + X_n$ as our current position in the random walk after n steps, and τ as the duration of our random walk.

3. Hypothesis #1

In our first counterexample, we will consider the random walk interpretation.

3.1 Example

Let $X_n = \{\pm 1\}$ with equal probability, and interpret $P_n = X_1 + \cdots + X_n$ as our position after n steps during a simple random walk on the one-dimensional lattice of integers, wherein we began our walk at the origin. Our stopping rule will then be taken to be the first point at which $P_\tau = 1$; in other words, when our random walk reaches $+1$. Since random walks on the one-dimensional lattice are recurrent, we will reach $+1$ in finite time with probability 1. Having conditioned on $X_1 + \cdots + X_\tau = 1$, then the conditioned expectation $\mathbb{E}(X_1 + \cdots + X_\tau) = 1$ is evident; furthermore, having assumed that $X_n = \pm 1$ with equal probability, we easily see that $\mathbb{E}(X_1) = 0$. Under these observations, assuming Wald's Identity, we obtain an immediate contradiction.

3.2 Remarks

The contradiction above depends upon a curious fact about simple random walks on the unbounded one-dimensional lattice: there is a finite-time random walk between any two vertices, but the expected time of such a walk is infinite. In other words, we are guaranteed to reach any destination, but there is no guarantee that we will be alive when we arrive, no matter how close it is! With knowledge that $\mathbb{E}(\tau) = \infty$ in this example, our first hypothesis becomes evident:

Assume our stopping rule τ has finite expected value.

4. Hypothesis #2

We will employ a gambling interpretation for this example.

4.1 Example

A gambler bets on a sequence of fair coin flips and calls 'heads' each time, stopping upon his first win. He bets \$1 on the first coin flip and doubles his bet upon each successive coin flip. Employing this system, we can set $X_n = \{\pm 2^{n-1}\}$ with equal probability. If $W_n = X_1 + \cdots + X_n$ is the gambler's current winnings up to and including the n -th coin flip, taking τ to be the time at which the first head appears, it is easily seen that $W_\tau = 1$, and so $\mathbb{E}(X_1 + \cdots + X_\tau) = 1$. Since the gambler is merely waiting for the first occurrence of heads, this defines a Bernoulli trial and so $\mathbb{E}(\tau) = 2$. This observation tells us that our first hypothesis holds, and Wald's Identity says us that $1 = 2 \cdot \mathbb{E}(X_1)$. However, this is a contradiction in light of the fact that $\mathbb{E}(X_n) = 0$.

4.2 Remarks

This example emphasizes the fact that X_1, X_2, \dots having common finite mean is too weak an assumption for (1) to hold. Additionally, the assumption that $\mathbb{E}(\tau)$ is finite is not strong enough either. We are then forced to consider the distribution of our random variables. First, they are not identically distributed. Second, their expected absolute values take on all nonnegative powers of 2; in other words, $\mathbb{E}|X_n| = 2^{n-1}$. These observations lead us to our second hypothesis:

Assume our random variables X_1, X_2, \dots are identically distributed and

$$\mathbb{E}|X_n| = \mathbb{E}|X_1| < \infty.$$

We noted in the introduction that we would be making stronger assumptions than are needed in general, this is one such case. In general, it is possible to prove Wald's Identity without the assumption that X_1, X_2, \dots are identically distributed. We make the stronger assumption for two reasons: (1) the method of proof presented in this paper will require it, and (2) the chosen method of proof is more straightforward and transparent with this assumption.

5. Hypothesis #3

We've already discussed the third hypothesis in the interpretations section, and so here we will present an example of Wald's Identity failing when the stopping rule depends upon events $X_{\tau+1}, X_{\tau+2}, \dots$

5.1 Example

We know that a simple random walk on the interval $[0, m] \cap \mathbb{Z}$ is recurrent, i.e. a walk which begins at zero will eventually return to zero; such a walk is called an excursion. It is likewise evident that if we let X_1, X_2, X_3, \dots be the lengths of such excursions of a simple random walk on $[0, m] \cap \mathbb{Z}$, then this is an i.i.d. sequence of random variables. Furthermore, we know that the expected length of such an excursion, $\mathbb{E}(X_1) = 2m$, this is just the reciprocal of the stationary state distribution at zero, which we know to be $1/2m$. Since $X_n > 0$ for all n , then we plainly see that $\mathbb{E}|X_n| = \mathbb{E}(X_n)$ and so our two previously discussed hypotheses are satisfied. All that remains is to define our stopping rule. Given the sequence X_1, X_2, \dots we can ask whether or not the excursion associated to each X_n reached the point m , or in other words, whether the excursion covered the interval $[0, m] \cap \mathbb{Z}$. Let τ be such that $X_{\tau+1}$ is the first such excursion which covers the interval, so that $\mathbb{E}(\tau) = m - 1$. Wald's Identity then tells us that

$$\mathbb{E}(X_1 + \dots + X_\tau) = 2m(m - 1).$$

Fixing $m = 2$, we find that $\mathbb{E}(\tau) = 1$ and $\mathbb{E}(X_1 + \dots + X_\tau) = 4$; however, there is only one excursion which does not cover $\{0, 1, 2\}$: the excursion $0 \rightarrow 1 \rightarrow 0$ of length 2. This tells us that $\mathbb{E}(X_1 + \dots + X_\tau) = 2\tau$, which therefore fixes $\tau = 2$, and this contradicts our observation that $\mathbb{E}(\tau) = 1$.

5.2 Remarks

This example highlights the failure of Wald's Identity when we are allowed to define our stopping rule based on the outcome of future events. While the example above is not a fair game, we could let $Y_n = X_n - 2m$, so that $\mathbb{E}(Y_n) = 0$, thereby supporting our intuition that it is possible for clairvoyant gamblers to construct stopping rules to give themselves an advantage in an otherwise fair game. This is a reason that "insider trading" is not allowed on the stock market: if we assume that trading stocks is a fair

game, then insider trading gives brokers a means in which to gain an unfair advantage over competition. This of course leads us to our third and final hypothesis:

Assume our stopping rule τ is independent of $X_{\tau+1}, X_{\tau+2}, X_{\tau+3}, \dots$

6. Proof of Wald's Identity

Taking a quick look back at the first two examples, we can easily see that each example satisfies the other two hypotheses, yet fails the third. And so omission of any one hypothesis is insufficient to prove Wald's Identity, as we have provided counterexamples to each possibility. We will now turn our attention toward the proof that our hypotheses are sufficient to show that Wald's Identity holds. First, however, we will require two supplementary theorems of Kolmogorov which will aid us: The Strong Law of Large Numbers and The Converse to the Strong Law.

Theorem 6.1 (Strong Law of Large Numbers). *Let X_1, X_2, \dots be i.i.d. random variables with common finite mean with $\mathbb{E}|X_n| = \mathbb{E}|X_1| < \infty$. Then*

$$\mathbb{P}\left(\lim_{k \rightarrow \infty} \frac{X_1 + \dots + X_k}{k} = \mathbb{E}(X_1)\right) = 1.$$

Theorem 6.2 (Converse to SLLN). *Let X_1, X_2, \dots be i.i.d. random variables with common mean. If*

$$\mathbb{P}\left(\lim_{k \rightarrow \infty} \frac{X_1 + \dots + X_k}{k} \text{ exists}\right) = 1,$$

then $\mathbb{E}|X_n| = \mathbb{E}|X_1| < \infty$.

Theorem 6.3 (Wald's Identity). *Let X_1, X_2, \dots be i.i.d. random variables with common finite mean with $\mathbb{E}|X_n| = \mathbb{E}|X_1| < \infty$, and let τ be a stopping rule which is independent of $X_{\tau+1}, X_{\tau+2}, \dots$ for which $\mathbb{E}(\tau) < \infty$, then*

$$\mathbb{E}(X_1 + \dots + X_\tau) = \mathbb{E}(\tau)\mathbb{E}(X_1).$$

The proof we present here is a proof of David Blackwell from his 1946 paper [1].

Proof. We begin by inductively defining a sequence of stopping times τ_1, τ_2, \dots from our initial sequence X_1, X_2, \dots of random variables. Let $\tau_1 = \tau$ and note that

$$\mathbb{P}(\tau_1 = n \mid X_{\tau+1}, X_{\tau+2}, \dots) = \mathbb{P}(\tau_1 = n)$$

by hypothesis for all n . Disregarding X_1, \dots, X_τ , we can apply our stopping rule to the remaining infinite sequence $X_{\tau+1}, X_{\tau+2}, \dots$, and so let τ_2 be the stopping time for this sequence, noting that

$$\mathbb{P}(\tau_2 = n \mid X_{\tau_1+\tau_2+1}, X_{\tau_1+\tau_2+2}, \dots) = \mathbb{P}(\tau_2 = n).$$

Repeating this process inductively, we obtain the desired sequence τ_1, τ_2, \dots . We can then easily verify that

$$\begin{aligned} \mathbb{P}(\tau_{n+1} = \alpha_{n+1} \mid \tau_1 = \alpha_1, \dots, \tau_n = \alpha_n) &= \frac{\mathbb{P}(\tau_{n+1} = \alpha_{n+1}, \tau_1 = \alpha_1, \dots, \tau_n = \alpha_n)}{\mathbb{P}(\tau_1 = \alpha_1, \dots, \tau_n = \alpha_n)} \\ &= \frac{\mathbb{P}(\tau_{n+1} = \alpha_{n+1})\mathbb{P}(\tau_1 = \alpha_1, \dots, \tau_n = \alpha_n)}{\mathbb{P}(\tau_1 = \alpha_1, \dots, \tau_n = \alpha_n)} \\ &= \mathbb{P}(\tau_{n+1} = \alpha_{n+1}), \end{aligned}$$

therefore τ_1, τ_2, \dots is an i.i.d. sequence of random variables.

Having established independence of the sequence of stopping times, we now define another sequence of random variables. Put

$$\begin{aligned} S_1 &= X_1 + \dots + X_{\tau_1} \\ S_2 &= X_{\tau_1+1} + \dots + X_{\tau_1+\tau_2} \\ S_3 &= X_{\tau_1+\tau_2+1} + \dots + X_{\tau_1+\tau_2+\tau_3} \\ &\vdots \end{aligned}$$

By reasoning similar to the sequence of τ 's, we see that S_1, S_2, \dots is an i.i.d. sequence of random variables. We would like to be able to apply the Strong Law of Large Numbers to this sequence, but we do not yet know that $\mathbb{E}|S_n| = \mathbb{E}|S_1| < \infty$, we merely know that they have common mean. If we were to show that we could apply the Converse to S_1, S_2, \dots , we would then be able to apply the Strong Law. To do this, we consider the following:

$$\lim_{k \rightarrow \infty} \frac{S_1 + \dots + S_k}{k} = \lim_{k \rightarrow \infty} \frac{S_1 + \dots + S_k}{N_1 + \dots + N_k} \cdot \lim_{k \rightarrow \infty} \frac{N_1 + \dots + N_k}{k} \quad (2)$$

where we have multiplied by 1 within the limit and separated the product. By definition of S_1, \dots, S_k , we observe that the leftmost term in (2) is equal to

$$\lim_{k \rightarrow \infty} \frac{X_1 + \dots + X_{\tau_1 + \dots + \tau_k}}{N_1 + \dots + N_k},$$

which is merely a subsequence of the sequence of averages of the X 's. By the Strong Law, we know that

$$\mathbb{P} \left(\lim_{k \rightarrow \infty} \frac{X_1 + \dots + X_k}{k} = \mathbb{E}(X_1) \right) = 1,$$

and so the subsequence above must also converge to a common limit with probability 1. Likewise, applying the Strong Law to the rightmost term in (2), we find that

$$\mathbb{P} \left(\lim_{k \rightarrow \infty} \frac{N_1 + \dots + N_k}{k} = \mathbb{E}(\tau) \right) = 1.$$

Together, these facts imply that

$$\mathbb{P} \left(\lim_{k \rightarrow \infty} \frac{S_1 + \dots + S_k}{k} = \mathbb{E}(\tau)\mathbb{E}(X_1) \right) = 1, \quad (3)$$

hence the limit exists with probability 1. By the Converse to SLLN, we find that $\mathbb{E}|S_n| = \mathbb{E}|S_1| < \infty$ and so we can apply the Strong Law to our sequence of sums S_1, S_2, \dots , obtaining

$$\mathbb{P}\left(\lim_{k \rightarrow \infty} \frac{S_1 + \dots + S_k}{k} = \mathbb{E}(S_1)\right) = 1. \quad (4)$$

Putting (3) and (4) together, we are left to conclude that

$$\mathbb{E}(X_1 + \dots + X_\tau) = \mathbb{E}(\tau)\mathbb{E}(X_1)$$

as desired. □

References

- [1] Blackwell, D., "On an equation of Wald," *Annals of Math. Stat.*, Vol. 17, (1946), pp. 84-87.
- [2] Wald, A., "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15, (1944), pp. 283-296.
- [3] Wald, A., "Some generalizations of the theory of cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 16, (1945), pp. 287-293.