Wald's identities

Matija Vidmar

August 27, 2018

Setting. We are given a probability space $(\Omega, \mathcal{G}, \mathsf{P})$, a discrete random time $T \in \mathcal{G}/2^{\mathbb{N}_0 \cup \{\infty\}}$ and a sequence $X = (X_i)_{i\in\mathbb{N}}$ from $\mathcal{G}/\mathcal{B}_{[-\infty,\infty]}$.

Notation. We will denote by $\mathsf{P}[Y]$ the P-expectation of a $Y \in \mathcal{G}/\mathcal{B}_{[-\infty,\infty]}$, whenever it is well-defined. The symbol $\perp$ will be used to indicate independence under P.

We are first interested in the random sum[1]

$$S := \sum_{i=1}^{T} X_i,$$

where we set $S = 0$ on $A := \{\sum_{i=1}^T X_i^+ = \sum_{i=1}^T X_i^- = \infty\}$ (this is arbitrary, but in the end we will anyhow only deal with situations that will ensure $\mathsf{P}(A) = 0$). Off A, $\sum_{i=1}^T X_i$ is well-defined "omega-by-omega", either as a finite sum (when $T < \infty$; in particular we understand $\sum_1^0(\ldots) = 0$) or as a convergent series (when $T = \infty$) in $(-\infty,\infty]$ or $[-\infty,\infty)$ (so allowing convergence to ∞ or $-\infty$). It is of course the case that $S \in \mathcal{G}/\mathcal{B}_{[-\infty,\infty]}$.

We ask now what, and under which conditions, can be said about $\mathsf{P}[S]$? We notice, for instance, that if $X_i \geq 0$ a.s. and $\mathsf{P}[X_i] = \mathsf{P}[X_1]$ for all $i \in \mathbb{N}$, then:

(1) If T is deterministic or if all the X_i, $i \in \mathbb{N}$, are deterministic, i.e. if there exists a (necessarily unique) $n \in \mathbb{N}_0 \cup \{\infty\}$ such that $T = n$ a.s. or if there exists a (necessarily unique) $x \in [0,\infty]$ such that $X_i = x$ a.s. for all $i \in \mathbb{N}$, then $(\star)$ $\mathsf{P}[S] = \mathsf{P}[X_1]\mathsf{P}[T]$.

   [Is $(\star)$ true also when T and the X_i, $i \in \mathbb{N}$, are not deterministic?]

(2) Suppose X_i, $i \in \mathbb{N}$, are i.i.-Ber$(\frac{1}{2})$-d.; note that $\mathsf{P}[X_1] = \frac{1}{2}$. Set $U := \inf\{i \in \mathbb{N} : X_i = 1\}$. Then $U \sim \mathrm{geom}_{\mathbb{N}}(\frac{1}{2})$, hence $\mathsf{P}[U] = 2$.

   (a) Let $T = U$. Since $S = \mathbb{1}_{\{T<\infty\}}$, we have that $\mathsf{P}[S] = 1 = 2 \cdot \frac{1}{2} = \mathsf{P}[T]\mathsf{P}[X_1]$.

   [Is $(\star)$ maybe true always?]

   (b) Let $T = U - 1$. Because $\sum_{i=1}^{U-1} X_i = 0$, so $\mathsf{P}[\sum_{i=1}^{U-1} X_i] = 0 \neq 1 \cdot \frac{1}{2} = \mathsf{P}[U-1]\mathsf{P}[X_1]$.

   [So sometimes $(\star)$ is true and sometimes it is not.]

Reasonably innocuous conditions under which $\mathsf{P}[S] = \mathsf{P}[X_1]\mathsf{P}[T]$ obtains are the content of

Proposition 1 (Wald's 1st identity)**.** *Suppose (S1)* $(\mathsf{P}[X_1^+]\mathsf{P}[T]) \wedge (\mathsf{P}[X_1^-]\mathsf{P}[T]) < \infty$ *&* $\mathsf{P}[X_i^\pm] = \mathsf{P}[X_1^\pm]$ *for all* $i \in \mathbb{N}$*, and (S2)* $X_{i+1} \perp \{T \leq i\}$ *for all* $i \in \mathbb{N}_0$*. Then*

$$\mathsf{P}[S] = \mathsf{P}[X_1]\mathsf{P}[T]. \tag{1}$$

Remark 2. (S1) reduces to $\mathsf{P}[X_i] = \mathsf{P}[X_1]$ for all $i \in \mathbb{N}$ in case $X_i \geq 0$ a.s. for all $i \in \mathbb{N}$ (recall the convention $0 \cdot \infty = 0$), and in any event it implies that $\mathsf{P}(A) = 0$.

Remark 3. If $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}_0}$ is a filtration on $(\Omega, \mathcal{G})$ w.r.t. which T is a stopping time and X has independent values (it means: $X_{i+1} \perp \mathcal{F}_i$ for all $i \in \mathbb{N}_0$) then (S2) certainly obtains. In particular this is the case if the X_i, $i \in \mathbb{N}$, are independent and T is a stopping time of X.

[1] In terms of motivation, one may e.g. think of X_i as the i-th claim, $i \in \mathbb{N}$, and of T as the cumulative number of claims in a given time period; then S is the cumulative claim amount.

Remark 4. If the X_i, $i \in \mathbb{N}$, are equally distributed, with $\mathsf{P}[X_1^+] \wedge \mathsf{P}[X_1^-] < \infty$, then for sure $\mathsf{P}[X_i^\pm] = \mathsf{P}[X_1^\pm]$ and (hence) $\mathsf{P}[X_1] = \mathsf{P}[X_i]$ for all $i \in \mathbb{N}$.

Remark 5. If $T \perp X$, or if just $T \perp X_i$ for all $i \in \mathbb{N}$, then (S2) obtains.

Remark 6. While we will not pursue this connection, (1) and the identities (2) and (3) to follow, are intimately connected with the theory of martingales and the optional sampling theorem.

Example 7. In (2b) we see that $U - 1$ cannot be a stopping time of (the natural filtration[2] of) X.[3]

Proof. Suppose first $X_i \geq 0$ for all $i \in \mathbb{N}$. We compute: $\mathsf{P}[S] = \mathsf{P}[\sum_{i=1}^T X_i] = \mathsf{P}[\sum_{i=1}^\infty X_i \mathbb{1}_{\{i \leq T\}}] =$ (we may interchange expectations and nonnegative sums) $\sum_{i=1}^\infty \mathsf{P}[X_i; i \leq T] = \sum_{i=1}^\infty \mathsf{P}[X_i; \Omega \backslash \{T \leq i-1\}] =$ (thanks to (S2)) $\sum_{i=1}^\infty \mathsf{P}[X_i]\mathsf{P}(\Omega \backslash \{T \leq i-1\}) =$ (because of (S1)) $\sum_{i=1}^\infty \mathsf{P}[X_1]\mathsf{P}(T \geq i) =$ (take $X_1 = 1$) $\mathsf{P}[X_1]\mathsf{P}[T]$. [We have silently used the conventions $0 \cdot \infty = 0$ and $\sum_1^0 \ldots = 0$.]

To handle the general case we reduce it to the nonnegative one. Clearly $X_{i+1}^\pm \perp \{T \leq i\}$ for all $i \in \mathbb{N}_0$. Then "the nonnegative case" with $X^\pm$ in lieu of X implies $\mathsf{P}[\sum_{i=1}^T X_i^\pm] = \mathsf{P}[X_1^\pm]\mathsf{P}[T]$. Because $(\mathsf{P}[X_1^+]\mathsf{P}[T]) \wedge (\mathsf{P}[X_1^-]\mathsf{P}[T]) < \infty$, linearity of the integral allows to conclude. □

The second moment of S also admits a nice expression under reasonably general assumptions.

Proposition 8 (Wald's 2nd identity)**.** *Suppose that (V1) for all $i \in \mathbb{N}_0$, $(X_1, \ldots, X_i, T \wedge (i+1)) \perp X_{i+1}$ and (V2) $\mathsf{P}[T] < \infty$ & $\mathsf{P}[X_i] = 0$ & $\mathsf{P}[X_i^2] = \mathsf{P}[X_1^2] < \infty$ for all $i \in \mathbb{N}$. Then*

$$\mathsf{P}[S^2] = \mathsf{P}[X_1^2]\mathsf{P}[T]. \tag{2}$$

Remark 9. (V1) certainly obtains if X has independent values relative to some filtration $\mathcal{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$ to which it is adapted and w.r.t. which T is a stopping time. In particular this is the case if the X_i, $i \in \mathbb{N}$, are independent and T is a stopping time of X.

Example 10. Let $\{a, b\} \subset \mathbb{N}$. Suppose the X_i, $i \in \mathbb{N}$, are i.i.d. with values in $\{-1, 1\}$ and $\mathsf{P}(X_1 = 1) = 1 - \mathsf{P}(X_1 = -1) = \frac{1}{2}$ (independent equiprobable random signs). Set $W_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}_0$ so that $W = (W_n)_{n \in \mathbb{N}_0}$ is the simple symmetric random walk. Let $T = \inf\{n \in \mathbb{N} : W_n \in \{-b, a\}\}$. Then $S = W_T$ on $\{T < \infty\}$. We would like to (i.) establish that $T < \infty$ a.s., as well as determine (ii.) $\mathsf{P}[T]$ and (iii.) $\mathsf{P}(W_T = a, T < \infty)$, hence $\mathsf{P}(W_T = -b, T < \infty)$. Now certainly T is a stopping time of X but we do not know *a priori* that its mean is finite (or even that it is finite a.s.). For this reason we consider first, for each $N \in \mathbb{N}$, the bounded stopping time $T \wedge N$. For the latter in lieu of T, Wald's second identity writes as $\mathsf{P}[W_{T \wedge N}^2] = \mathsf{P}[X_1^2]\mathsf{P}[T \wedge N]$. Letting $N \uparrow \infty$ monotone convergence coupled with $|W_{T \wedge N}| \leq \max\{a, b\}$ renders $\mathsf{P}[T]\mathsf{P}[X_1^2] \leq \max\{a, b\}$. This and $\mathrm{var}(X_1) > 0$ entail that $\mathsf{P}[T] = \mathsf{P}[X_1^2] < \infty$ (in particular, $T < \infty$ a.s., which we will now assume is true with certainty without affecting any distributional results). Then Wald's first identity implies $\mathsf{P}[W_T] = \mathsf{P}[S] = \mathsf{P}[X_1]\mathsf{P}[T] = 0$. Since $\mathsf{P}[W_T] = a\mathsf{P}(W_T = a) - b\mathsf{P}(W_T = -b)$ and $\mathsf{P}(W_T = a) + \mathsf{P}(W_T = -b) = 1$, it means that $\mathsf{P}(W_T = a) = \frac{b}{a+b}$ and $\mathsf{P}(W_T = -b) = \frac{a}{a+b}$. Finally we use Wald's second identity (again, but now simply for T): $\mathsf{P}[W_T^2] = \mathsf{P}[X_1^2]\mathsf{P}[T]$. Since $\mathsf{P}[W_T^2] = \frac{b}{a+b}a^2 + \frac{a}{a+b}b^2 = ab$ and $\mathsf{P}[X_1^2] = 1$, we obtain $\mathsf{P}[T] = ab$.

Proof. Because $(\sum_{i=1}^T |X_i|)^2 \leq 2\sum_{i=1}^T X_i^2$, Wald's first identity for X^2 in lieu of X and monotonicity of the integral imply that $\mathsf{P}[(\sum_{i=1}^T |X_i|)^2] < \infty$. Then we may compute: $\mathsf{P}[S^2] = \mathsf{P}[(\sum_{i=1}^T X_i)^2] = \mathsf{P}[(\sum_{i=1}^\infty X_i \mathbb{1}_{\{i \leq T\}})^2] = \mathsf{P}[\sum_{i,j=1}^\infty X_i X_j \mathbb{1}_{\{i \leq T, j \leq T\}}] =$ (dominated convergence and linearity) $= \mathsf{P}[\sum_{i=1}^\infty X_i^2 \mathbb{1}_{\{i \leq T\}}] + \sum_{\substack{i,j=1 \\ i \neq j}}^\infty \mathsf{P}[X_i X_j; i \vee j \leq T]$

$$= \mathsf{P}\left[\sum_{i=1}^T X_i^2\right] + \sum_{\substack{i,j=1 \\ i \neq j}}^\infty \mathsf{P}[X_i X_j; i \vee j \leq T].$$

[2] It is the filtration $\mathcal{F}^X = (\mathcal{F}_n^X)_{n \in \mathbb{N}_0}$ given by $\mathcal{F}_n^X = \sigma(X_1, \ldots, X_n)$ for $n \in \mathbb{N}_0$ (in particular $\mathcal{F}_0^X = \{\emptyset, \Omega\}$).

[3] This is of course obvious in this case because $\{U - 1 = 0\}$ is non-trivial, indeed its probability is $\frac{1}{2}$, while X carries no information up to time 0. But in general it gives a useful trick for proving that some random time is not a stopping time.

The first expectation is equal to $\mathsf{P}[X_1^2]\mathsf{P}[T]$ on account of Wald's first identity (again with X^2 in lieu of X). Each term of the second sum is equal to zero as we now verify: let $i<j$ be natural numbers; then $\mathsf{P}[X_iX_j; i\lor j\leq T]=\mathsf{P}[X_iX_j;\Omega\backslash\{T\land j\leq (i-1)\lor(j-1)\}]=$ (because of (V1)) $\mathsf{P}[X_i;\Omega\backslash\{T\land j\leq(i-1)\lor(j-1)\}]\mathsf{P}[X_j]=$ (because of (V2)) 0. $\square$

Corollary 11. *Suppose that (I1) $X\perp T$ & the X_i, $i\in\mathbb{N}$, are independent, and (I2) $\mathsf{P}[T]<\infty$ & $\mathsf{P}[X_1]=\mathsf{P}[X_i]$ & $\mathsf{P}[X_i^2]=\mathsf{P}[X_1^2]<\infty$ for all $i\in\mathbb{N}$. Then $\mathrm{cov}(S,T)=\mathsf{P}[X_1]\mathrm{var}(T)$ and $\mathrm{var}(S)=\mathrm{var}(T)\mathsf{P}[X_1]^2+\mathrm{var}(X_1)\mathsf{P}[T]$.*

Proof. Applying Wald's second identity to the sequence $(X_i-\mathsf{P}[X_1])_{i\in\mathbb{N}}$ in lieu of the sequence X and combining with Wald's first identity yields the relation $\mathrm{var}(S)=-\mathrm{var}(T)\mathsf{P}[X_1]^2+2\mathrm{cov}(S,T)\mathsf{P}[X_1]+\mathrm{var}(X_1)\mathsf{P}[T]$. So it suffices to establish the statement concerning the covariance. But $\mathsf{P}[ST]=\mathsf{P}[\sum_{i=1}^T X_iT]=\sum_{k=0}^\infty k\mathsf{P}[\sum_{i=1}^k X_i;T=k]=\mathsf{P}[X_1]\sum_{k=0}^\infty k^2\mathsf{P}[T=k]=\mathsf{P}[X_1]\mathsf{P}[T^2]$, while by Wald's first identity $\mathsf{P}[S]\mathsf{P}[T]=\mathsf{P}[X_1]\mathsf{P}[T]^2$. $\square$

Finally in the multiplicative setting we have

Proposition 12 (Wald's 3rd identity)**.** *Let T be bounded by an $N\in\mathbb{N}_0$ and assume the X_i, $i\in\{1,\ldots,N\}$, are independent, take their values in $[0,\infty)$, with (M1) $\mathsf{P}[X_i]=1$ for all $i\in\mathbb{N}$ and (M2) $\mathbb{1}_{\{T=k\}}\prod_{i=1}^k X_i\perp\prod_{i=k+1}^N X_i$ for all $k\in\{0,\ldots,N\}$ holding true. Then with*

$$M:=\prod_{i=1}^T X_i,$$

one has

$$\mathsf{P}[M]=1, \tag{3}$$

where of course we understand $\prod_{i=1}^0(\ldots)=1$.[4]

Remark 13. (M2) certainly obtains if X has independent values relative to some filtration $\mathcal{F}=(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ to which it is adapted and w.r.t. which T is a stopping time.

Example 14. We assume the setting of Example 10, but insist $a=b$. Let also $\gamma\in[1,\infty)$. We are interested in $\mathsf{P}[\gamma^{-T};T<\infty]$. It was already shown in Example 10 that $T<\infty$ a.s and we will assume that $T<\infty$ with certainty, without affecting any distributional results. For $\theta\in(0,\infty)$, set $\psi(\theta):=\mathsf{P}[\theta^{X_1}]=\frac{\theta+\theta^{-1}}{2}$. Then, for any $N\in\mathbb{N}$, Wald's third identity for $T\land N$ in lieu of T and for the sequence $(\theta^{X_i}/\psi(\theta))_{i\in\mathbb{N}}$ in lieu of X produces $\mathsf{P}[\theta^{W_{T\land N}}\psi(\theta)^{-T\land N}]=1$. As $N\uparrow\infty$, we have that $W_{T\land N}\to W_T$. By bounded convergence we obtain $\mathsf{P}[\theta^{W_T}\psi(\theta)^{-T}]=1$. The process $-X$ has the same law as X, while the time T is invariant under this transformation. So the same argument with $-X$ in lieu of X gives $\mathsf{P}[\theta^{-W_T}\psi(\theta)^{-T}]=1$. Adding the two equalities yields $\mathsf{P}[\psi(\theta)^{-T}]=\frac{2}{\theta^a+\theta^{-a}}$. Letting now $\gamma\in[1,\infty)$ and taking $\theta=\gamma-\sqrt{\gamma^2-1}$ (so that $\psi(\theta)=\gamma$), we obtain finally $\mathsf{P}[\gamma^{-T}]=\frac{2}{(\gamma-\sqrt{\gamma^2-1})^a+(\gamma-\sqrt{\gamma^2-1})^{-a}}$. (As a check, for $a=1$ we obtain $\mathsf{P}[\gamma^{-T}]=\gamma^{-1}$ as we should.)

Proof. We compute $\mathsf{P}[\prod_{i=1}^T X_i]=\sum_{k=0}^N\mathsf{P}[\prod_{i=1}^k X_i;T=k]=$ (because the X_i, $i\in\mathbb{N}$, are independent and since $\mathsf{P}[X_i]=1$ for all $i\in\mathbb{N}$, which together imply that $\mathsf{P}[\prod_{i=k+1}^N X_i]=1$ for all $k\in\{0,\ldots,N\}$) $\sum_{k=0}^N\mathsf{P}[\prod_{i=1}^k X_i;T=k]\mathsf{P}[\prod_{i=k+1}^N X_i]=$ (because of (M2)) $\sum_{k=0}^N\mathsf{P}[\prod_{i=1}^N X_i;T=k]=\mathsf{P}[\prod_{i=1}^N X_i]=1$. $\square$

[4] In terms of motivation one may e.g. think of X_i as the multiplicative increment of a risky asset/gamble (normalized by its mean) during period i, $i\in\mathbb{N}$, and that we hold/play for T time periods. Then $M-1$ is the cumulative relative (normalized) gain/loss until termination.

References

[1] Miller, H. D. A Generalization of Wald's Identity with Applications to Random Walks. Ann. Math. Statist. 32 (1961), no. 2, 549–560. doi:10.1214/aoms/1177705060. https://projecteuclid.org/euclid.aoms/1177705060.

[2] Novikov A.A. (1992) Wald Identities (No. 35). In: Shiryayev A.N. (eds) Selected Works of A. N. Kolmogorov. Mathematics and Its Applications (Soviet Series), vol 26. Springer, Dordrecht.

[3] Bellman, Richard Ernest, On a Generalization of the Fundamental Identity of Wald. Santa Monica, CA: RAND Corporation, 1956. https://www.rand.org/pubs/papers/P777.html.

[4] Blom, Gunnar. A Generalization of Wald's Fundamental Identity. Ann. Math. Statist. 20 (1949), no. 3, 439–444. doi:10.1214/aoms/1177729996. https://projecteuclid.org/euclid.aoms/1177729996.