Wald's Identities

13.1 Doob's martingale transform

Definition 13.1 (**Martingale difference sequence**). *Let $(\mathcal{F}_n)$ be a filtration and (Δ_n) be an adapted sequence of random variables to $(\mathcal{F}_n)$. Then (Δ_n) is a* ***Martingale difference sequence*** *if it satisfies* 1) $\Delta_n \in L^1(\mathcal{F}_n)$, 2) $\mathbb{E}(\Delta_{n+1} \mid \mathcal{F}_n) = 0$ *for all* $n \geqslant 0$.

Observe that it is possible to give an equivalent definition of Martingale in term of Martingale difference sequence as follows: Let $(\mathcal{F}_n)$ be a filtration and (Δ_n) be a Martingale difference sequence. Then the sequence (M_n) defined by

$$M_n = \sum_{i=0}^{n} \Delta_i, n \geqslant 1$$

is a Martingale. Notice that if 2) is changed to $\mathbb{E}(\Delta_{n+1} \mid \mathcal{F}_n) \geqslant 0$ ($\mathbb{E}(\Delta_{n+1} \mid \mathcal{F}_n) \leqslant 0$), then (M_n) is a sub-martingale (super-martingale).

Theorem 13.2 (Doob's martingale transform)**.** *Suppose $(M_n, \mathcal{F}_n)_{n\geqslant 0}$ is a martingale and $(H_n, \mathcal{F}_n)_{n\geqslant 0}$ is predictable. Define*

$$\Delta_n = M_n - M_{n-1}, \quad (H \cdot M)_n = \sum_{i=1}^{n} H_i \Delta_i$$

for $n \geqslant 1$. Then $((H \cdot M)_n, \mathcal{F}_n)$ is a martingale whenever $\mathbb{E}\,|(H \cdot M)_n| < \infty$ for all n.

Proof. Since $(H \cdot M)_{n+1} - (H \cdot M)_n = H_{n+1} \cdot \Delta_{n+1}$, we have for $n \geqslant 0$, $\mathbb{E}(H_{n+1}\Delta_{n+1} \mid \mathcal{F}_n) = H_{n+1}\,\mathbb{E}(\Delta_{n+1} \mid \mathcal{F}_n) = 0$. ■

Observe that if $H_n \geqslant 0$, then (M_n) is a sub-martingale (super-martingale) implies $(H \cdot M)_n$ is a sub-martingale (super-martingale).

Theorem 13.3. *If $(M_n, \mathcal{F}_n)$ is a martingale (sub/super-martingale) and T is a stopping time with respect to a filtration $(\mathcal{F}_n)$, then*

$$W_n = M_{n\wedge T}, n \geqslant 0$$

is a $(\mathcal{F}_n)-$martingale (sub/super-martingale).

Proof. Notice that

$$M_{n\wedge T} = \sum_{i=1}^{n\wedge T} \Delta_i = \sum_{i=1}^{n} \Delta_i \mathbb{1}_{T\geqslant i}.$$

In particular, $\mathbb{E}(M_{n\wedge T}) = \mathbb{E}(M_0)$ for all $n \geqslant 0$. ■

Note that, $M_{n\wedge T} \xrightarrow{a.s.} M_T$ as $n \to \infty$ when M_T is well-defined a.s.

Example 13.4. *Let (X_i) be an iid sequence of random variable which is identically distributed to simple random walk on $\mathbb{Z}$. Let $M_n = S_n = \sum_{i=1}^n X_i$. Then $\mathbb{E}(M_0) = 0$. Let $T = \inf\{n \mid S_n \geqslant 1\}$. Then*

$$M_T = S_T = 1 \text{ yields } \mathbb{E}(M_T) = 1 \neq \mathbb{E}(M_0).$$

This example shows that it is not always true that $\mathbb{E}(M_0) = \lim \mathbb{E}(M_{n\wedge T}) = \mathbb{E}(\lim M_{n\wedge T}) = \mathbb{E}(M_T)$.

Notice that if T is bounded by some k, then $n \wedge T = T$ for $n \geqslant k$ and $M_{n\wedge T} = M_T, n \geqslant k$.

13.2 Wald's Identities

Theorem 13.5 (Wald's first identity). *Let $(X_n, n \geqslant 1)$ be a sequence of mean zero independent r.v.s with $\sup_{i\geqslant 1} \mathbb{E}\,|X_i| < \infty$. Consider the martingale*

$$M_n = S_n := X_1 + X_2 + \cdots + X_n, \quad \mathcal{F}_n := \sigma(X_1, \ldots, X_n), \quad n \geqslant 0.$$

Let T be a $(\mathcal{F}_n, n \geqslant 0)$ stopping time with $\mathbb{E}\,T < \infty$. Then

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0.$$

In particular, if X_i's are i.i.d. r.v.s with $\mathbb{E}\,X_1 = \mu$ and T is a $(\sigma(X_1, \ldots, X_n), n \geqslant 0)$ stopping time with $\mathbb{E}\,T < \infty$, then

$$\mathbb{E}\,S_T = \mu\,\mathbb{E}\,T.$$

Proof. We have $M_{T\wedge n} = \sum_{i=1}^n X_i \mathbb{1}_{T\geqslant i} \to M_T$ a.s. as $n \to \infty$. Moreover $\mathbb{E}(M_{T\wedge n}) = 0$ for $n \geqslant 0$. By the fact that $X_i \perp \mathcal{F}_{i-1}, \mathbb{1}_{T\geqslant i} = \mathbb{1}_{T>i-1} \in \mathcal{F}_{i-1}$ we have

$$\mathbb{E} \sum_{i=1}^\infty |X_i| \cdot \mathbb{1}_{T\geqslant i} = \sum_{i=1}^\infty \mathbb{E}(|X_i| \mathbb{1}_{T\geqslant i}) = \sum_{i=1}^\infty \mathbb{E}\,|X_i| \cdot \mathbb{E}(\mathbb{1}_{T\geqslant i}) \leqslant \sup_{i\geqslant 1} \mathbb{E}\,|X_i| \cdot \mathbb{E}\,T < \infty.$$

Using DCT with $|M_{T\wedge n}| \leqslant \sum_{i=1}^\infty |X_i| \cdot \mathbb{1}_{T\geqslant i}$ for all $n \geqslant 0$ we get the result. ∎

Theorem 13.6 (Wald's second identity). *Let $(X_n, n \geqslant 1)$ be a sequence of mean zero independent r.v.s with $\sup_{i\geqslant 1} \mathbb{E}\,X_i^2 < \infty$. Consider the martingale*

$$M_n = S_n^2 - \sum_{i=1}^n \mathbb{E}\,X_i^2, \quad \mathcal{F}_n := \sigma(X_1, \ldots, X_n), \quad n \geqslant 0$$

where $S_n := X_1 + X_2 + \cdots + X_n$. Let T be a $(\mathcal{F}_n, n \geqslant 0)$ stopping time with $\mathbb{E}\,T < \infty$. Then

$$\mathbb{E}(M_T) = \mathbb{E}(M_0) = 0.$$

In particular, if X_i's are i.i.d. r.v.s with $\mathbb{E}\,X_1 = 0, \mathrm{Var}(X_1) = \sigma^2$ and T is a $(\sigma(X_1, \ldots, X_n), n \geqslant 0)$ stopping time with $\mathbb{E}\,T < \infty$, then

$$\mathbb{E}\,S_T^2 = \sigma^2\,\mathbb{E}\,T.$$

Proof. Let $\sigma_i^2 = \mathbb{E}\, X_i^2$ for $i \geqslant 1$. We have

$$\mathbb{E}\, M_{T\wedge n} = \mathbb{E}\, S_{T\wedge n}^2 - \mathbb{E} \sum_{i=1}^{T\wedge n} \sigma_i^2 = 0 \text{ for all } n \geqslant 0.$$

Now

$$S_{T\wedge n} = \sum_{i=1}^{n} X_i \mathbb{1}_{T\geqslant i} \to S_T = \sum_{i=1}^{\infty} X_i \mathbb{1}_{T\geqslant i} \text{ a.s. as } n \to \infty,$$

X_i is orthogonal to $\mathcal{F}_{i-1}$ and $\mathbb{1}_{T\geqslant i} \in \mathcal{F}_{i-1}$ for $i \geqslant 1$. Thus

$$\mathbb{E}(X_i \mathbb{1}_{T\geqslant i} \cdot X_j \mathbb{1}_{T\geqslant j}) = 0$$

for all $i \neq j$ and for $0 \leqslant n < \infty$

$$\begin{aligned}\sup_{m\geqslant n} \|S_{T\wedge m} - S_{T\wedge n}\|_2^2 &= \sup_{m\geqslant n} \sum_{i=n+1}^{m} \|X_i \mathbb{1}_{T\geqslant i}\|_2^2 \\ &= \sup_{m\geqslant n} \sum_{i>n} \mathbb{E}(X_i^2 \mathbb{1}_{T\geqslant i}) = \sum_{i>n} \mathbb{E}\, X_i^2 \cdot \mathbb{E}(\mathbb{1}_{T\geqslant i}) \leqslant \sup_{i\geqslant 1} \mathbb{E}\, X_i^2 \cdot \sum_{i>n} \mathbb{E}(\mathbb{1}_{T\geqslant i}).\end{aligned}$$

Thus, $(S_{T\wedge n})$ is L^2-Cauchy and $\mathbb{E}\, S_{T\wedge n}^2 \to \mathbb{E}\, S_T^2$. Similar argument and DCT shows that $\mathbb{E} \sum_{i=1}^{T\wedge n} \sigma_i^2 \to \mathbb{E} \sum_{i=1}^{T} \sigma_i^2$. Thus we have the result. ■

Theorem 13.7 (Wald's third identity). *Let $(X_n, n \geqslant 1)$ be a sequence of i.i.d. r.v.s with $\mathbb{E}(e^{\theta X_1}) = \phi(\theta) < \infty$. Let T be a $(\sigma(X_1, X_2, \ldots, X_n))-$stopping time. If T is a.s. bounded or*

$$\phi(\theta)^{-n} \cdot e^{\theta S_n} \cdot \mathbb{1}_{T\geqslant n} \leqslant K \text{ for all } n \geqslant 0,$$

then

$$\mathbb{E}\left(\phi(\theta)^{-T} \cdot e^{\theta S_T}\right) = 1.$$

Proof. Since $M_n := \phi(\theta)^{-n} \cdot e^{\theta S_n}$ is a martingale w.r.t. the filtration $\mathcal{F}_n := \sigma(X_1, \ldots, X_n), n \geqslant 0$, we have $\mathbb{E}(M_{T\wedge n}) = \mathbb{E}(M_{T\wedge 0}) = 1$. The results follows in the first case by the fact that $T \leqslant K$ a.s. implies $T \wedge K = T$ and in the second case by DCT. ■

We provide some examples to illustrate how to use Wald's identities.

Example 13.8 (**SSRW (Simple Symmetric Random Walk) on $\mathbb{Z}$).** *Let (X_n) be a sequence of i.i.d. r.v.s with $\mathbb{P}(X_1 = +1) = \mathbb{P}(X_1 = -1) = 1/2, \mathbb{E}\, X_1 = 0, \mathbb{E}\, X_1^2 = 1$. Define $S_0 = 0, S_n := S_{n-1} + X_n, n \geqslant 1$. Let $a, b \in \mathbb{Z}$ with $a, b > 0$. Define*

$$T = \inf\{n \geqslant 0 \mid S_n = -a \text{ or } S_n = b\}.$$

Then $\mathbb{E}\, T < \infty$, by using second moment martingale to get $\mathbb{E}(T \wedge n) = \mathbb{E}\, S_{T\wedge n}^2 \leqslant a^2 + b^2$ for all $n \geqslant 0$.

Moreover, by Wald's first identity we have $\mathbb{E}(S_T) = 0$, i.e.,

$$a\, \mathbb{P}(S_T = -a) = b\, \mathbb{P}(S_T = b) = b(1 - \mathbb{P}(S_T = a)).$$

So

$$\mathbb{P}(S_T = -a) = \frac{b}{a+b}, \mathbb{P}(S_T = b) = \frac{a}{a+b}.$$

Example 13.9 (**SRW (Simple Random Walk) with drift).** *Fix $p \in (0,1), p \neq 1/2$. Let (X_n) be a sequence of i.i.d. r.v.s with $\mathbb{P}(X_1 = +1) = p, \mathbb{P}(X_1 = -1) = q = 1 - p, \mu := \mathbb{E}\, X_1 = 2p - 1, \mathbb{E}\, X_1^2 = 1$. Define $S_0 = 0, S_n := S_{n-1} + X_n, n \geqslant 1$. Let $a, b \in \mathbb{Z}$ with $a, b > 0$. Define*

$$T = \inf\{n \geqslant 0 \mid S_n = -a \text{ or } S_n = b\}.$$

By Wald's first identity, we have

$$\mathbb{E}(T \wedge n) = \mu^{-1}\, \mathbb{E}\, S_{T\wedge n} \leqslant |\mu|^{-1} \cdot (a+b) \text{ for all } n \geqslant 0$$

and thus $\mathbb{E}\, T < \infty$. Choose $\theta = \log(q/p) \in \mathbb{R}$ so that $e^\theta = q/p$ and $\mathbb{E}\, e^{\theta X_1} = 1$. We also have $e^{\theta S_n} \mathbb{1}_{\{T \geqslant n\}} \leqslant e^{|\theta| \cdot (a+b)}$ for all $n \geqslant 0$. By Wald's Third Identity we have

$$1 = \left(\frac{q}{p}\right)^{-a} \mathbb{P}(S_T = -a) + \left(\frac{q}{p}\right)^{b} (1 - \mathbb{P}(S_T = -a)).$$

Thus, we have

$$\mathbb{P}(S_T = -a) = \frac{p^b q^a - q^{a+b}}{p^{a+b} - q^{a+b}} \text{ and } \mathbb{P}(S_T = b) = \frac{p^{a+b} - p^b q^a}{p^{a+b} - q^{a+b}}.$$

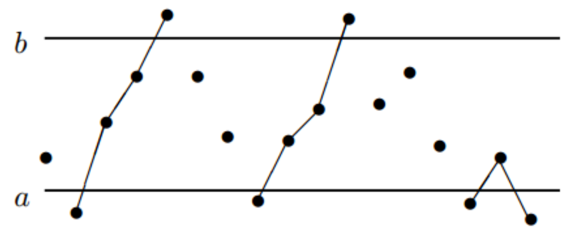
Using Wald's third identity with general $\theta \in \mathbb{R}$, one can calculate the probabilities $\mathbb{P}(T = k), k \geqslant 0$ (see the homework exercise). The first identity can be generalized under the assumption $\sup_{n\geqslant 1} \mathbb{E}(|M_n - M_{n-1}| \mid \mathcal{F}_{n-1}) \leqslant K < \infty$ a.s.

13.3 Martingale Convergence Theorem

Let $(M_n, \mathcal{F}_n)_{n\geqslant 0}$ be a submartingale and $a < b$. We define $N_0 = -1$ and

$$N_1 = \inf\{i > N_0 \mid M_i \leqslant a\}, \qquad N_2 = \inf\{i > N_1 \mid M_i \geqslant b\},$$
$$N_3 = \inf\{i > N_2 \mid M_i \leqslant a\}, \qquad N_4 = \inf\{i > N_3 \mid M_i \geqslant b\},$$
$$\text{In general, } N_{2k-1} = \inf\{i > N_{2k-2} \mid M_i \leqslant a\}, \qquad N_{2k} = \inf\{i > N_{2k-1} \mid M_i \geqslant b\} \text{ for } k \geqslant 1.$$

First we claim that.



Lemma 13.10. *For $i \geqslant 1$, N_i is a stopping time w.r.t. the filtration $(\mathcal{F}_n)_{n\geqslant 0}$.*

Proof. The proof is by induction. Clearly, N_1 is a stopping time. Suppose $N_1, \ldots, N_{i-1}$ are stopping time. If i is even, then

$$\{N_i \leqslant n\} = \bigcup_{j=1}^{n} \{N_{i-1} = j-1\} \cap \{j \leqslant N_i \leqslant n\} = \bigcup_{j=1}^{n-1} \{N_{i-1} = j-1\} \cap \bigcup_{k=j}^{n} \{M_k \geqslant b\} \in \mathcal{F}_n.$$

Similarly, when i is odd,

$$\{N_i \leqslant n\} = \bigcup_{j=1}^{n} \{N_{i-1} = j-1\} \cap \{j \leqslant N_i \leqslant n\} = \bigcup_{j=1}^{n-1} \{N_{i-1} = j-1\} \cap \bigcup_{k=j}^{n} \{M_k \leqslant a\} \in \mathcal{F}_n. \qquad \blacksquare$$

Define the upcrossing random variable

$$U_n(a,b) = \sup\{k \mid N_{2k} \leqslant n\}.$$

For fixed $m \geqslant 1$ and any $k \geqslant 1$,

$$\{N_{2k-1} < m \leqslant N_{2k}\} = \{N_{2k-1} \leqslant m-1 \text{ and } N_{2k} > m-1\} \in \mathcal{F}_{m-1}.$$

Thus $H_n = \cup_{k\geqslant 1}\{N_{2k-1} < n \leqslant N_{2k}\}, n \geqslant 0$ is a predictable sequence and

$$(H \cdot M)_n \geqslant (b-a)U_n(a,b).$$

Theorem 13.11 (Upcrossing Inequality). *For any $a < b$ and any submartingale $(M_n, \mathcal{F}_n)_{n\geqslant 0}$, we have,*

$$(b-a)\,\mathbb{E}(U_n(a,b)) \leqslant \mathbb{E}(M_n - a)^+ - \mathbb{E}(M_0 - a)^+ \text{ for all } n \geqslant 1.$$

Proof. Define $Y_n := \phi(M_n) = (M_n - a)^+ + a, n \geqslant 0$ where $\phi(x) = (x-a)^+ + a$ is a non-decreasing convex function. By conditional Jensen's inequality, we have

$$\mathbb{E}(Y_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(\phi(M_{n+1}) \mid \mathcal{F}_n) \geqslant \phi(\mathbb{E}(M_{n+1} \mid \mathcal{F}_n)) \geqslant \phi(M_n) = Y_n.$$

Thus $(Y_n, \mathcal{F}_n)_{n\geqslant 0}$ is a submartingale. Moreover, upcrossings for M_n and Y_n over the interval $[a,b]$ are the same. Thus $(H \cdot Y)_n \geqslant (b-a)U_n(a,b)$ implies that

$$(b-a)\,\mathbb{E}(U_n(a,b)) \leqslant \mathbb{E}(H \cdot Y)_n.$$

Now, we have $Y_n - Y_0 = (H \cdot Y)_n + ((1-H)\cdot Y)_n$ and $\mathbb{E}((1-H)\cdot Y)_n \geqslant \mathbb{E}((1-H)\cdot Y)_0 = 0$. Thus,

$$\mathbb{E}(H \cdot Y)_n \leqslant \mathbb{E}(Y_n - Y_0). \qquad \blacksquare$$

Theorem 13.12 (Martingale Convergence Theorem (MGCT)). *If* $(M_n, \mathcal{F}_n)_{n\geqslant 0}$ *is a sub-martingale with* $\sup_n \mathbb{E}\, M_n^+ < \infty$, *then*

$$M_n \to M_\infty \text{ a.s.}$$

for some $M_\infty \in L^1(\mathcal{F})$.

Proof. Let $K = \sup_n \mathbb{E}\, M_n^+$. Fix $a < b$. We have

$$(b-a)\,\mathbb{E}\, U_n(a,b) \leqslant \mathbb{E}(M_n - a)^+ \leqslant \mathbb{E}\, M_n^+ + a \leqslant K + a.$$

Note that, $U_n(a,b) \uparrow U(a,b)$ as $n \to \infty$ where $U(a,b)$ is the total number of upcrossings of the interval $[a,b]$. Thus, $\mathbb{E}\, U(a,b) < \infty$ and in particular, $U(a,b) < \infty$ a.s. Thus,

$$\mathbb{P}\left(\cup_{a<b,a,b\in\mathbb{Q}} \{U(a,b) = \infty\}\right) = 0,$$

which implies that

$$\begin{aligned}\mathbb{P}\left(\liminf M_n = \limsup M_n\right) &= 1 - \mathbb{P}\left(\liminf M_n < \limsup M_n\right)\\ &= 1 - \mathbb{P}\left(\cup_{a<b,a,b\in\mathbb{Q}} \{\liminf M_n < a < b < \limsup M_n\}\right)\\ &= 1 - P\left(\cup_{a<b,a,b\in\mathbb{Q}} \{U(a,b) = \infty\}\right) = 1.\end{aligned}$$

In Particular,

$$M_n \to M_\infty \text{ a.s.}$$

for some r.v. M_∞. Now $M_n^+ \to M_\infty^+$ and $M_n^- \to M_\infty^-$ a.s. By Fatou's Lemma, we have

$$\mathbb{E}\, M_\infty^+ = \mathbb{E}(\liminf M_n^+) \leqslant \liminf \mathbb{E}\, M_n^+, \quad \mathbb{E}\, M_\infty^- = \mathbb{E}(\liminf M_n^-) \leqslant \liminf \mathbb{E}\, M_n^-.$$

Moreover,

$$\mathbb{E}\, M_n^- = \mathbb{E}(M_n^+ - M_n) = \mathbb{E}\, M_n^+ - \mathbb{E}\, M_n \leqslant \sup \mathbb{E}\, M_n^+ - \mathbb{E}\, M_0$$

which implies that $\mathbb{E}\, M_\infty^+, \mathbb{E}\, M_\infty^- < \infty$. ■

Note that, for a submartingale $(X_n, \mathcal{F}_n)_{n\geqslant 1}$

$$\sup_n \mathbb{E}\, X_n^+ < \infty \Leftrightarrow \sup_n \mathbb{E}\, |X_n| < \infty.$$

Corollary 13.13. *A positive super-martingale converges a.s. to an integrable r.v.*

Proof. If $(X_n)_{n\geqslant 0}$ is a positive super-martingale, then $(-X_n)_{n\geqslant 0}$ is a negative sub-martingale, *i.e.,* $\sup_n \mathbb{E}((-X_n)^+) = 0 < \infty$ and thus we can apply MGCT to $-X_n$ to get the desired result. ■