17. Stopping Times

Given a sequence of random variables $X_1, X_2, ...$ on a probability space $(\Omega, \mathcal{F}, P)$, consider the sub-σ-algebras $\mathcal{F}_n = \sigma(X_1, ..., X_n)$, $n \geq 1$. If we think of $X_1, X_2, \ldots$ as observations taken at times $1, 2, \ldots$, then $\mathcal{F}_n$ can be interpreted as the information available at time n.

Note that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$. Such an increasing sequence of sub-σ-algebras is known as a *filtration*, and the space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$ is called a *filtered probability space*.
(For the time being, we will assume that the filtration is indexed by $\mathbb{N}$, but one can consider more general index sets such as $[0, \infty)$ as well.)

If $X_1, X_2, ...$ satisfies $X_n \in \mathcal{F}_n$ for all n, we say that the sequence $\{X_n\}$ is *adapted* to the filtration $\{\mathcal{F}_n\}$. $\{\sigma(X_1, ..., X_n)\}$ is the smallest filtration with respect to which $\{X_n\}$ is adapted.

Definition. Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}}, P)$, a random variable $N : \Omega \to \mathbb{N} \cup \{\infty\}$ is said to be a *stopping time* if $\{N = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

The following proposition gives an equivalent definition of stopping times. When working with more general index sets than $\mathbb{N}$, this is the appropriate definition.

Proposition 17.1. *A random variable $N : \Omega \to \mathbb{N} \cup \{\infty\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$ is a stopping time if and only if $\{N \leq n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.*

Proof. Let $n \in \mathbb{N}$ be given. If N is a stopping time, then for each $m \leq n$, $\{N = m\} \in \mathcal{F}_m \subseteq \mathcal{F}_n$, so $\{N \leq n\} = \bigcup_{m=1}^{n} \{N = m\} \in \mathcal{F}_n$.

Conversely, if $\{N \leq m\} \in \mathcal{F}_m$ for all m, then $\{N \leq n-1\} \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$, so
$\{N = n\} = \{N \leq n\} \setminus \{N \leq n-1\} \in \mathcal{F}_n$. □

To motivate the definition, suppose that $X_1, X_2, ...$ represent one's winnings in successive games of roulette. Let N be a rule for when to stop gambling (in the sense that one quits playing after N games).
The requirement that $\{N = n\} \in \mathcal{F}_n$ means that the decision to stop playing after n games can only depend on the outcomes of the first n games.

For example, the random variable $N \equiv 6$ corresponds to the rule that one will play exactly six games, regardless of the outcomes.

The random variable $N = \inf\{n : \sum_{i=1}^{n} X_i < -10\}$ corresponds to quitting once one's losses exceed \$10.

The random variable $N = \inf\{n : \sum_{i=1}^{n} X_i \geq \sum_{i=1}^{m} X_i \text{ for all } m \in \mathbb{N}\}$, corresponding to quitting once one has attained the maximum amount they ever will, is not a stopping time since it depends on the future as well as the past and present.

The second rule is a canonical example of stopping times. It is the hitting time of $(-\infty, -10)$.

In general, the random variable $N = \inf\{n : S_n \in A\}$ is a stopping time known as the *hitting time of A*. To verify that N is indeed a stopping time, observe that $\{N = n\} = \{S_1 \in A^C, ..., S_{n-1} \in A^C, S_n \in A\} \in \mathcal{F}_n$.

Associated with each stopping time N is the *stopped σ-algebra* $\mathcal{F}_N$, which we think of as the information known at time N.

Formally $\mathcal{F}_N = \{A \in \mathcal{F} : A \cap \{N = n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}$. That is, on $N = n$, A must be measurable with respect to the information known at time n. It is worth noting that $\{N \le n\} \in \mathcal{F}_N$ for all $n \in \mathbb{N}$, hence N is $\mathcal{F}_N$-measurable. Our first result about $\mathcal{F}_N$ is

Theorem 17.1. *Suppose that $X_1, X_2, ...$ are i.i.d., $\mathcal{F}_n = \sigma(X_1, ..., X_n)$, and N is a stopping time for $\mathcal{F}_n$. Conditional on $\{N < \infty\}$, $\{X_{N+n}\}_{n\ge1}$ is independent of $\mathcal{F}_N$ and has the same distribution as the original sequence.*

Proof. Let $A \in \mathcal{F}_N$, $k \in \mathbb{N}$, $B_1, ..., B_k \in \mathcal{S}$ be given. Let μ denote the common distribution of the $X_i's$.

For any $n \in \mathbb{N}$, we have

$$\begin{aligned} P(A, N = n, X_{N+1} \in B_1, ..., X_{N+k} \in B_k) &= P(A, N = n, X_{n+1} \in B_1, ..., X_{n+k} \in B_k) \\ &= P(A \cap \{N = n\}) \prod_{j=1}^{k} P(X_{n+j} \in B_j) = P(A \cap \{N = n\}) \prod_{j=1}^{k} \mu(B_j) \end{aligned}$$

since $A \cap \{N = n\} \in \mathcal{F}_n$ and $X_{n+1}, ..., X_{n+k}$ is independent of $\mathcal{F}_n$.

Summing over n gives

$$\begin{aligned} P(A, N < \infty, X_{N+1} \in B_1, \cdots, X_{N+k} \in B_k) &= \sum_{n=1}^{\infty} P(A, N = n, X_{N+1} \in B_1, ..., X_{N+k} \in B_k) \\ &= \sum_{n=1}^{\infty} P(A \cap \{N = n\}) \prod_{j=1}^{k} \mu(B_j) = P(A \cap \{N < \infty\}) \prod_{j=1}^{k} \mu(B_j), \end{aligned}$$

proving independence, and taking $A = \Omega$ shows that

$$\frac{P(N < \infty, X_{N+1} \in B_1, ..., X_{N+k} \in B_k)}{P(N < \infty)} = \prod_{j=1}^{k} \mu(B_j).$$

□

We now introduce the shift function the $\theta : \Omega \to \Omega$ which is defined coordinatewise by $(\theta\omega)_n = \omega_{n+1}$.

That is, applying θ results in dropping the first term and shifting all others one place to the left. Higher order shifts are defined by iterating θ:
$\theta^1 = \theta$ and $\theta^n = \theta \circ \theta^{n-1}$ for $n > 1$.

Thus θ^n acts on ω by dropping the first n terms and shifting the remaining terms n places to the left, so that $(\theta^n\omega)_i = \omega_{n+i}$.

We extend the shift function to stopping times by setting

$$\theta^N \omega = \begin{cases} \theta^n \omega \text{ on } \{N = n\} \\ \triangle \text{ on } \{N = \infty\} \end{cases}$$

where $\triangle$ is an extra point we add to Ω to make various natural constructions work out nicely.

Example 17.1 (Returns to zero)**.**

Suppose that $S = \mathbb{R}^d$ and let $\tau(\omega) = \inf\{n : \omega_1 + \ldots + \omega_n = 0\}$ where $\inf \emptyset = \infty$ and $\tau(\triangle) := \infty$.

Thus τ gives the first time the random walk visits 0.

Setting $\tau_2(\omega) = \tau(\omega) + \tau(\theta^\tau \omega)$, we see that on $\{\tau < \infty\}$,

$$\tau(\theta^\tau \omega) = \inf\{n : (\theta^\tau \omega)_1 + \ldots + (\theta^\tau \omega)_n = 0\} = \inf\{n : \omega_{\tau+1} + \ldots + \omega_{\tau+n} = 0\},$$

hence

$$\tau_2(\omega) = \tau(\omega) + \tau(\theta^\tau \omega) = \inf\{m > \tau : \omega_1 + ... + \omega_m = 0\}.$$

Because of the convention that $\tau(\triangle) = \infty$, this is well defined for all ω and gives the time of the second visit to zero.

The same reasoning shows that if we set $\tau_0 = 0$, then

$$\tau_n(\omega) = \tau_{n-1}(\omega) + \tau(\theta^{\tau_{n-1}} \omega)$$

is well-defined for all $n \in \mathbb{N}$ and gives the time of the *nth* visit to zero.

Of course, this idea is applicable to stopping times in general:

If T is any stopping time, then setting $T_0 = 0$, we can define the iterates of T by

$$T_n(\omega) = T_{n-1}(\omega) + T\left(\theta^{T_{n-1}} \omega\right) \text{ for } n \geq 1.$$

Proposition 17.2. *In the above setting, if we assume that $P = \mu \times \mu \times \cdots$, then $P(T_n < \infty) = P(T < \infty)^n$.*

Proof. We argue by induction. The base case is trivial, so let us assume that the statement holds for $n - 1$. Applying Theorem 17.1 to T_{n-1}, we see that $\left\{X_{T_{n-1}+k}\right\}_{k \geq 1} = \left\{\theta^{T_{n-1}} X_k\right\}_{k \geq 1}$ is independent of $\mathcal{F}_{T_{n-1}}$ on $\{T_{n-1} < \infty\}$ and has the same distribution as $\{X_k\}_{k \geq 1}$.

Consequently, $\left\{T \circ \theta^{T_{n-1}} < \infty\right\}$ is independent of $\{T_{n-1} < \infty\}$ and has the same probability as $\{T < \infty\}$, hence

$$\begin{aligned} P(T_n < \infty) &= P(T_{n-1} < \infty, T \circ \theta^{T_{n-1}} < \infty) = P(T_{n-1} < \infty) P(T \circ \theta^{T_{n-1}} < \infty) \\ &= P(T < \infty)^{n-1} P(T < \infty) = P(T < \infty)^n \end{aligned}$$

and the result follows. □

Our next result about stopping times is the famous

Theorem 17.2 (Wald's equation)**.** *Let $X_1, X_2, ...$ be i.i.d. with $E|X_1| < \infty$. If N is a stopping time with $E[N] < \infty$, then $E[S_N] = E[X_1]E[N]$.*

Proof. First suppose that the $X_i's$ are nonnegative. Then

$$E[S_N] = \int S_N dP = \sum_{n=1}^{\infty} \int 1\{N=n\} S_n dP = \sum_{n=1}^{\infty} \sum_{m=1}^{n} \int 1\{N=n\} X_m dP$$
$$= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int 1\{N=n\} X_m dP = \sum_{m=1}^{\infty} \int 1\{N \geq m\} X_m dP$$

where interchanging the order of summation is justified by the nonnegativity assumption. Since $\{N \geq m\} = \{N \leq m-1\}^C \in \mathcal{F}_{m-1}$ and X_m is independent of $\mathcal{F}_{m-1}$, we have

$$E[S_N] = \sum_{m=1}^{\infty} \int 1\{N \geq m\} X_m dP = \sum_{m=1}^{\infty} E[1\{N \geq m\} X_m]$$
$$= \sum_{m=1}^{\infty} P(N \geq m) E[X_m] = E[X_1] \sum_{m=1}^{\infty} P(N \geq m) = E[X_1] E[N].$$

To prove the result for general X_i, we run the last argument in reverse to conclude that

$$\infty > E|X_1| E[N] = \sum_{m=1}^{\infty} P(N \geq m) E|X_m| = \sum_{m=1}^{\infty} \int 1\{N \geq m\} |X_m| dP$$
$$= \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int 1\{N=n\} |X_m| dP \geq \sum_{n=1}^{\infty} \int 1\{N=n\} |S_n| dP.$$

Since the double integrals converge absolutely, we can invoke Fubini to conclude that

$$E[X_1] E[N] = \sum_{m=1}^{\infty} P(N \geq m) E[X_m] = \sum_{m=1}^{\infty} \int 1\{N \geq m\} X_m dP = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \int 1\{N=n\} X_m dP$$
$$= \sum_{n=1}^{\infty} \sum_{m=1}^{n} \int 1\{N=n\} X_m dP = \sum_{n=1}^{\infty} \int 1\{N=n\} S_n dP = E[S_N]. \qquad \square$$

One consequence of Wald's equation is that one can gain no advantage in a fair or unfavorable game by employing a length-of-play strategy which does not depend on the possibility of infinitely many games, infinite payoff, or the ability to see the future.

One hears people advocate the policy of playing until they are ahead. Denoting the outcomes of successive games by $X_1, X_2, ...,$ this stopping rule is given by $\alpha = \inf\{n : X_1 + ... + X_n > 0\}$. If $E|X_1| < \infty$ and $E[X_1] \leq 0$, then $E[\alpha] < \infty$ would imply $0 < E[S_\alpha] = E[\alpha]E[X_1] \leq 0$, a contradiction. Thus the expected waiting time until one shows a profit on a sequence of independent and identical fair or unfavorable bets is infinite.

Some other amusing consequences involve variations on the following game:
Suppose that you are to roll a die repeatedly until a number of your choice appears. You are then awarded an amount of money equal to the sum of your rolls. Wald's equation shows that any number you choose will result in the same expected winnings.

Indeed the outcomes of each roll, $X_1, X_2, ...,$ are i.i.d. uniform over $\{1, 2, ..., 6\}$. The waiting time, N_i, until any number $i \in \{1, ..., 6\}$ appears is geometric with success probability $\frac{1}{6}$. Thus your expected winnings are $E[S_{N_i}] = E[N_i]E[X_1] = 6 \cdot \frac{1+2+...+6}{6} = 21$ regardless of the number you choose. There is no advantage in choosing six over one, say, in terms of expected winnings. (Of course there is an advantage in terms of things like maximizing your minimum potential winnings.)

We now consider an application of Wald's equation to the analysis of simple random walk.

Example 17.2 (Simple Random Walk)**.**

Let $X_1, X_2, ...$ be i.i.d. with $P(X_1 = 1) = P(X_1 = -1) = \frac{1}{2}$. Let $a < 0 < b$ be integers and set $N = \inf\{n : S_n \notin (a, b)\}$, the first time the walk exits (a, b).

We first note that for any $x \in (a, b)$, $P\left(x + S_{b-a} \notin (a, b)\right) \geq 2^{-(b-a)}$ since $b - a$ steps in the positive direction, say, will take us out of (a, b).
Iterating this inequality shows that $P\left(N > n(b-a)\right) \leq \left(1 - 2^{-(b-a)}\right)^n$, so $E[N] < \infty$.
(In fact, the exponential tail decay implies that N has moments of all orders.)

Applying Wald's equation shows that $bP(S_N = b) + aP(S_N = a) = E[S_N] = E[N]E[X_1] = 0$.
Since $P(S_N = a) = 1 - P(S_N = b)$, we have $(b - a)P(S_N = b) = -a$, hence

$$P(S_N = b) = \frac{-a}{b-a}, \quad P(S_N = a) = \frac{b}{b-a}.$$

Writing $T_a = \inf\{n : S_n = a\}$, $T_b = \inf\{n : S_n = b\}$ shows that $P(T_a < T_b) = P(S_N = a) = \frac{b}{b-a}$ for all integers $a < 0 < b$. Because $P(T_a < \infty) \geq P(T_a < T_b)$ for all $b > 0$, sending $b \to \infty$ shows that $P(T_a < \infty) = 1$ for all integral $a < 0$.

Symmetry implies that $P(T_x < \infty) = 1$ for all integral $b > 0$, and since the walk must pass through 0 to get from 1 to -1 or vice versa, we see that $P(T_0 < \infty) = 1$ as well.

However, Wald's equation shows that the expected time to visit any nonzero integer is infinite:
If $E[T_x] < \infty$ for $x \in \mathbb{Z} \setminus \{0\}$, we would have $x = E[S_{T_x}] = E[T_x]E[X_1] = 0$.

By conditioning on the first step, we see that the expected time for the walk to return to 0 is
$E\left[T_0\right] = E\left[\frac{1}{2}\left(1 + T_{-1}\right) + \frac{1}{2}\left(1 + T_1\right)\right] = \infty$.

To recap, simple random walk will visit any given integer in finite time with full probability, but the expected time to do so is infinite!

We can compute the variance of random sums whose index of summation is a stopping time using

Theorem 17.3 (Wald's second equation)**.** *Let $X_1, X_2, ...$ be i.i.d. with $E[X_1] = 0$ and $E[X_1^2] = \sigma^2 < \infty$. If T is a stopping time with $E[T] < \infty$, then $E[S_T^2] = \sigma^2 E[T]$.*

Proof. We first note that for all $n \in \mathbb{N}$,

$$\begin{aligned} S_{T\wedge n}^2 &= \left(\sum_{i=1}^{T\wedge n} X_i\right)^2 = \left[\sum_{i=1}^{T\wedge(n-1)} X_i + X_n 1\left\{T \geq n\right\}\right]^2 \\ &= S_{T\wedge(n-1)}^2 + \left(2X_n S_{n-1}^2 + X_n^2\right) 1\left\{T \geq n\right\}. \end{aligned}$$

Since $\{T \geq n\} = \{T \leq n-1\}^C \in \mathcal{F}_{n-1}$ and X_n is independent of $\mathcal{F}_{n-1}$, taking expectations yields

$$\begin{aligned} E\left[S_{T\wedge n}^2\right] &= E\left[S_{T\wedge(n-1)}^2\right] + 2E\left[X_n\right]E\left[S_{n-1}^2 1\left\{T \geq n\right\}\right] + E\left[X_n^2\right]E\left[1\left\{T \geq n\right\}\right] \\ &= E\left[S_{T\wedge(n-1)}^2\right] + \sigma^2 P(T \geq n). \end{aligned}$$

By assumption, all expectations exist and are finite, so induction on n gives

$$E\left[S_{T\wedge n}^2\right] = \sigma^2 \sum_{k=1}^{n} P(T \geq k).$$

Now $E[T] = \sum_{k=1}^{\infty} P(T > k) = \lim_{n\to\infty} \sum_{k=1}^{n} P(T \geq k)$, so, since $S_{T\wedge n} \to S_T$ pointwise, if we can show that $S_{T\wedge n}$ is Cauchy in L^2, then it will follow that $E\left[S_T^2\right] = \lim_{n\to\infty} E\left[S_{T\wedge n}^2\right] = \sigma^2 E[T]$.

To this end, observe that for any $n > m$

$$E\left[(S_{T\wedge n} - S_{T\wedge m})^2\right] = E\left[\left(\sum_{k=m+1}^{T\wedge n} X_k\right)^2\right] = \sigma^2 \sum_{k=m+1}^{n} P(T \geq k) \leq \sigma^2 \sum_{k=m+1}^{\infty} P(T \geq k)$$

where the second equality follows from the exact same argument as above.

Since this goes to 0 as $m \to \infty$, $S_{T\wedge n}$ is indeed Cauchy in L^2 and the proof is complete. □

A consequence of Wald's second equation is

Theorem 17.4. *Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = 0$, $E[X_1^2] = 1$, and set $T_c = \inf\{n \geq 1 : |S_n| > c\sqrt{n}\}$. Then $E[T_c]$ is finite if and only if $c < 1$.*

Proof. If $E[T_c] < \infty$, then Wald's second equation implies $E[T_c] = E[T_c]E[X_1^2] = E[S_{T_c}^2]$.
However, $E[S_{T_c}^2] > E\left[\left(c\sqrt{T_c}\right)^2\right] = c^2 E[T_c]$ by construction, so when $c \geq 1$, the assumption that $E[T_c] < \infty$ leads to the contradiction $E[T_c] = E[S_{T_c}^2] > c^2 E[T_c]$.

Thus it remains only to show that $E[T_c] < \infty$ when $c \in [0, 1)$.

To this end, we let $\tau_n = T_c \wedge n$ and note that $S_{\tau_n - 1}^2 \leq c^2(\tau_n - 1) \leq c^2\tau_n$, so Theorem 17.3 and Cauchy-Schwarz give

$$\begin{aligned} E[\tau_n] = E[S_{\tau_n}^2] &= E\left[S_{\tau_n-1}^2 + 2S_{\tau_n-1}X_{\tau_n} + X_{\tau_n}^2\right] \\ &\leq c^2 E[\tau_n] + 2c\sqrt{E[\tau_n]E[X_{\tau_n}^2]} + E[X_{\tau_n}^2]. \end{aligned}$$

To complete the proof, we show

Lemma 17.1. *If $X_1, X_2, \ldots$ are i.i.d. with $E[X_1^2] < \infty$ and T is a stopping time with $E[T] = \infty$, then*

$$\lim_{n\to\infty} \frac{E[X_{T\wedge n}^2]}{E[T\wedge n]} = 0.$$

It will then follow that if $E[T_c] = \infty$ for $c \in [0, 1)$, then for any $\varepsilon \in (0, (1-c)^2)$, there is an $N \in \mathbb{N}$ so that $E[X_{\tau_n}^2] < \varepsilon E[\tau_n]$ whenever $n \geq N$, giving the contradiction

$$E[\tau_n] \leq c^2 E[\tau_n] + 2c\sqrt{E[\tau_n]E[X_{\tau_n}^2]} + E[X_{\tau_n}^2] < \left(c^2 + 2c\sqrt{\varepsilon} + \varepsilon\right) E[\tau_n] = (c + \sqrt{\varepsilon})^2 E[\tau_n] < E[\tau_n].$$

To prove Lemma 17.1, we observe that

$$E[X_{T\wedge n}^2] = E[X_{T\wedge n}^2; X_{T\wedge n}^2 \le \varepsilon\,(T\wedge n)] + \sum_{j=1}^{n} E[X_{T\wedge n}^2; X_{T\wedge n}^2 > \varepsilon j, T\wedge n = j]$$
$$\le \varepsilon E[T\wedge n] + \sum_{j=1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j].$$

Now, since $E[X_j^2; X_j^2 > \varepsilon j] \to 0$ as $j \to \infty$ (by the DCT), we can choose N large enough that $\sum_{j=1}^{n} E[X_j^2; X_j^2 > \varepsilon j] < n\varepsilon$ whenever $n > N$.

Then for all $n > N$, we have

$$\sum_{j=1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j] = \sum_{j=1}^{N} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j] + \sum_{j=N+1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j]$$
$$\le N E[X_1^2] + \sum_{j=N+1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j].$$

To bound the latter sum, we compute

$$\sum_{j=N+1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j] \le \sum_{j=N+1}^{n} E[X_j^2; T\wedge n \ge j, X_j^2 > \varepsilon j]$$
$$= \sum_{j=N+1}^{n} E\left[X_j^2 1\{T\wedge n \ge j\}\, 1\left\{X_j^2 > \varepsilon j\right\}\right]$$
$$= \sum_{j=N+1}^{n} P(T\wedge n \ge j) E\left[X_j^2; X_j^2 > \varepsilon j\right]$$
$$= \sum_{j=N+1}^{n} \sum_{k=j}^{\infty} P(T\wedge n = k) E\left[X_j^2; X_j^2 > \varepsilon j\right]$$
$$\le \sum_{k=N+1}^{\infty} \sum_{j=1}^{k} P(T\wedge n = k) E\left[X_j^2; X_j^2 > \varepsilon j\right]$$
$$\le \sum_{k=N+1}^{\infty} P(T\wedge n = k) k\varepsilon \le \varepsilon E[T\wedge n].$$

It follows that $n > N$ implies

$$E[X_{T\wedge n}^2] \le \varepsilon E[T\wedge n] + \sum_{j=1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j]$$
$$\le \varepsilon E[T\wedge n] + N E[X_1^2] + \sum_{j=N+1}^{n} E[X_j^2; T\wedge n = j, X_j^2 > \varepsilon j]$$
$$\le \varepsilon E[T\wedge n] + N E[X_1^2] + \varepsilon E[T\wedge n] = 2\varepsilon E[T\wedge n] + N E[X_1^2].$$

Since $E[T\wedge n] \to E[T] = \infty$ (by the MCT), we see that $\limsup_{n\to\infty} \frac{E[X_{T\wedge n}^2]}{E[T\wedge n]} \le 2\varepsilon$ and the lemma follows because $\varepsilon > 0$ is arbitrary. □