

Martingales, Lévy's continuity theorem, and the martingale central limit theorem

Jordan Bell

jordan.bell@gmail.com

Department of Mathematics, University of Toronto

May 29, 2015

1 Introduction

In this note, any statement we make about filtrations and martingales is about filtrations and martingales indexed by the positive integers, rather than the nonnegative real numbers.

We take

$$\inf \emptyset = \infty,$$

and for $m > n$, we take

$$\sum_{k=m}^n = 0.$$

(Defined rightly, these are not merely convenient ad hoc definitions.)

2 Conditional expectation

Let (Ω, \mathcal{A}, P) be a probability space and let \mathcal{B} be a sub- σ -algebra of \mathcal{A} . For each $f \in L^1(\Omega, \mathcal{A}, P)$, there is some $g : \Omega \rightarrow \mathbb{R}$ such that (i) g is \mathcal{B} -measurable and (ii) for each $B \in \mathcal{B}$, $\int_B g dP = \int_B f dP$, and if $h : \Omega \rightarrow \mathbb{R}$ satisfies (i) and (ii) then $h(\omega) = g(\omega)$ for almost all $\omega \in \Omega$.¹ We denote some $g : \Omega \rightarrow \mathbb{R}$ satisfying (i) and (ii) by $E(f|\mathcal{B})$, called the **conditional expectation of f with respect to \mathcal{B}** . In other words, $E(f|\mathcal{B})$ is the unique element of $L^1(\Omega, \mathcal{B}, P)$ such that for each $B \in \mathcal{B}$,

$$\int_B E(f|\mathcal{B}) dP = \int_B f dP.$$

The map $f \mapsto E(f|\mathcal{B})$ satisfies the following:

1. $f \mapsto E(f|\mathcal{B})$ is positive linear operator $L^1(\Omega, \mathcal{A}, P) \rightarrow L^1(\Omega, \mathcal{B}, P)$ with norm 1.

¹Manfred Einsiedler and Thomas Ward, *Ergodic Theory: with a view towards Number Theory*, p. 121, Theorem 5.1.

2. If $f \in L^1(\Omega, \mathcal{A}, P)$ and $g \in L^\infty(\Omega, \mathcal{B}, P)$, then for almost all $\omega \in \Omega$,

$$E(gf|\mathcal{B})(\omega) = g(\omega)E(f|\mathcal{B})(\omega).$$

3. If \mathcal{C} is a sub- σ -algebra of \mathcal{B} , then for almost all $\omega \in \Omega$,

$$E(E(f|\mathcal{B})|\mathcal{C})(\omega) = E(f|\mathcal{C})(\omega).$$

4. If $f \in L^1(\Omega, \mathcal{B}, P)$ then for almost all $\omega \in \Omega$,

$$E(f|\mathcal{B})(\omega) = f(\omega).$$

5. If $f \in L^1(\Omega, \mathcal{A}, P)$, then for almost all $\omega \in \Omega$,

$$|E(f|\mathcal{B})(\omega)| \leq E(|f||\mathcal{B})(\omega).$$

6. If $f \in L^1(\Omega, \mathcal{A}, P)$ is independent of \mathcal{B} , then for almost all $\omega \in \Omega$,

$$E(f|\mathcal{B})(\omega) = E(f).$$

3 Filtrations

A **filtration** of a σ -algebra \mathcal{A} is a sequence \mathcal{F}_n , $n \geq 1$, of sub- σ -algebras of \mathcal{A} such that $\mathcal{F}_m \subset \mathcal{F}_n$ if $m \leq n$. We set $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

A sequence of random variables $\xi_n : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ is said to be **adapted to the filtration** \mathcal{F}_n if for each n , ξ_n is \mathcal{F}_n -measurable.

Let $\xi_n : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$, $n \geq 1$, be a sequence of random variables. The **natural filtration of \mathcal{A} corresponding to ξ_n** is

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n).$$

It is apparent that \mathcal{F}_n is a filtration and that the sequence ξ_n is adapted to \mathcal{F}_n .

4 Martingales

Let \mathcal{F}_n be a filtration of a σ -algebra \mathcal{A} and let $\xi_n : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ be a sequence of random variables. We say that ξ_n is a **martingale with respect to \mathcal{F}_n** if (i) the sequence ξ_n is adapted to the filtration \mathcal{F}_n , (ii) for each n , $\xi_n \in L^1(P)$, and (iii) for each n , for almost all $\omega \in \Omega$,

$$E(\xi_{n+1}|\mathcal{F}_n)(\omega) = \xi_n(\omega).$$

In particular,

$$E(\xi_1) = E(\xi_2) = \dots,$$

i.e.

$$E(\xi_m) = E(\xi_n), \quad m \leq n.$$

We say that ξ_n is a **submartingale with respect to** \mathcal{F}_n if (i) and (ii) above are true, and if for each n , for almost all $\omega \in \Omega$,

$$E(\xi_{n+1}|\mathcal{F}_n)(\omega) \geq \xi_n(\omega).$$

In particular,

$$E(\xi_1) \leq E(\xi_2) \leq \dots,$$

i.e.

$$E(\xi_m) \leq E(\xi_n), \quad m \leq n.$$

We say that ξ_n is a **supermartingale with respect to** \mathcal{F}_n if (i) and (ii) above are true, and if for each n , for almost all $\omega \in \Omega$,

$$\xi_n(\omega) \geq E(\xi_{n+1}|\mathcal{F}_n)(\omega).$$

In particular,

$$E(\xi_1) \geq E(\xi_2) \geq \dots,$$

i.e.

$$E(\xi_m) \geq E(\xi_n), \quad m \leq n.$$

If we speak about a martingale without specifying a filtration, we mean a martingale with respect to the natural filtration corresponding to the sequence of random variables.

5 Stopping times

Let \mathcal{F}_n be a filtration of a σ -algebra \mathcal{A} . A **stopping time with respect to** \mathcal{F}_n is a function $\tau : \Omega \rightarrow \{1, 2, \dots\} \cup \{\infty\}$ such that for each $n \geq 1$,

$$\{\omega \in \Omega : \tau(\omega) = n\} \in \mathcal{F}_n.$$

It is straightforward to check that a function $\tau : \Omega \rightarrow \{1, 2, \dots\} \cup \{\infty\}$ is a stopping time with respect to \mathcal{F}_n if and only if for each $n \geq 1$,

$$\{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathcal{F}_n.$$

The following lemma shows that the time of first entry into a Borel subset of \mathbb{R} of a sequence of random variables adapted to a filtration is a stopping time.²

Lemma 1. *Let ξ_n be a sequence of random variables adapted to a filtration \mathcal{F}_n and let $B \in \mathcal{B}_{\mathbb{R}}$. Then*

$$\tau(\omega) = \inf\{n \geq 1 : \xi_n(\omega) \in B\}$$

is a stopping time with respect to \mathcal{F}_n .

²Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 55, Exercise 3.9.

Proof. Let $n \geq 1$. Then

$$\begin{aligned} \{\omega \in \Omega : \tau(\omega) = n\} &= \left(\bigcap_{k=1}^{n-1} \{\omega \in \Omega : \xi_k(\omega) \notin B\} \right) \cap \{\omega \in \Omega : \xi_n(\omega) \in B\} \\ &= \left(\bigcap_{k=1}^{n-1} A_k^c \right) \cap A_n, \end{aligned}$$

where

$$A_k = \{\omega \in \Omega : \xi_k(\omega) \in B\}.$$

Because the sequence ξ_k is adapted to the filtration \mathcal{F}_k , $A_k^c \in \mathcal{F}_k$ and $A_n \in \mathcal{F}_n$, and because \mathcal{F}_k is a filtration, the right-hand side of the above belongs to \mathcal{F}_n . \square

If ξ_n is a sequence of random variables adapted to a filtration \mathcal{F}_n and a stopping time τ with respect to \mathcal{F}_n , for $n \geq 1$ we define $\xi_{\tau \wedge n} : \Omega \rightarrow \mathbb{R}$ by

$$\xi_{\tau \wedge n}(\omega) = \xi_{\tau(\omega) \wedge n}(\omega), \quad \omega \in \Omega.$$

$\xi_{\tau \wedge n}$ is called **the sequence ξ_n stopped at τ** .³

Lemma 2. $\xi_{\tau \wedge n} : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ is a sequence of random variables adapted to the filtration \mathcal{F}_n .

Proof. Let $n \geq 1$ and let $B \in \mathcal{B}_{\mathbb{R}}$. Because

$$\{\omega : \xi_{\tau \wedge n}(\omega) \in B, \tau(\omega) > n\} = \{\omega : \xi_n(\omega) \in B, \tau(\omega) > n\}$$

and for any k ,

$$\{\omega : \xi_{\tau \wedge n}(\omega) \in B, \tau(\omega) = k\} = \{\omega : \xi_k \in B, \tau(\omega) = k\},$$

we get

$$\{\omega : \xi_{\tau \wedge n}(\omega) \in B\} = \{\omega : \xi_n(\omega) \in B, \tau(\omega) > n\} \cup \bigcup_{k=1}^n \{\omega : \xi_k(\omega) \in B, \tau(\omega) = k\}.$$

But

$$\{\xi_n \in B, \tau > n\} = \{\xi_n \in B\} \cap \{\tau > n\} \in \mathcal{F}_n$$

and

$$\{\xi_k \in B, \tau = k\} = \{\xi_k \in B\} \cap \{\tau = k\} \in \mathcal{F}_k,$$

and therefore

$$\{\xi_{\tau \wedge n} \in B\} \in \mathcal{F}_n.$$

In particular, $\{\xi_{\tau \wedge n} \in B\} \in \mathcal{A}$, namely, $\xi_{\tau \wedge n}$ is a random variable, and the above shows that this sequence is adapted to the filtration \mathcal{F}_n . \square

³Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 55, Exercise 3.10.

We now prove that a stopped martingale is itself a martingale with respect to the same filtration.⁴

Theorem 3. *Let \mathcal{F}_n be a filtration of a σ -algebra \mathcal{A} and let τ be a stopping time with respect to \mathcal{F}_n .*

1. *If ξ_n is a submartingale with respect to \mathcal{F}_n then so is $\xi_{\tau \wedge n}$.*
2. *If ξ_n is a supermartingale with respect to \mathcal{F}_n then so is $\xi_{\tau \wedge n}$.*
3. *If ξ_n is a martingale with respect to \mathcal{F}_n then so is $\xi_{\tau \wedge n}$.*

Proof. For $n \geq 1$, define

$$\alpha_n(\omega) = \begin{cases} 1 & \tau(\omega) \geq n \\ 0 & \tau(\omega) < n; \end{cases}$$

we remark that $\tau(\omega) \geq n$ if and only if $\tau(\omega) > n - 1$ and $\tau(\omega) < n$ if and only if $\tau(\omega) \leq n - 1$. For $B \in \mathcal{B}_{\mathbb{R}}$, (i) if $0, 1 \notin B$ then

$$\{\omega \in \Omega : \alpha_n(\omega) \in B\} = \emptyset \in \mathcal{F}_{n-1},$$

(ii) if $0, 1 \in B$ then

$$\{\omega \in \Omega : \alpha_n(\omega) \in B\} = \Omega \in \mathcal{F}_{n-1},$$

(iii) if $0 \in B$ and $1 \notin B$ then

$$\{\omega \in \Omega : \alpha_n(\omega) \in B\} = \{\omega \in \Omega : \alpha_n(\omega) = 0\} = \{\omega \in \Omega : \tau(\omega) \leq n - 1\} \in \mathcal{F}_{n-1},$$

and (iv) if $1 \in B$ and $0 \notin B$ then

$$\{\omega \in \Omega : \alpha_n(\omega) \in B\} = \{\omega \in \Omega : \alpha_n(\omega) = 1\} = \{\omega \in \Omega : \tau(\omega) > n - 1\} \in \mathcal{F}_{n-1},$$

Therefore $\{\alpha_n \in B\} \in \mathcal{F}_{n-1}$.

Set $\xi_0 = 0$, and we check that

$$\xi_{\tau \wedge n} = \sum_{k=1}^n \alpha_k(\xi_k - \xi_{k-1}).$$

It is apparent from this expression that if ξ_n is adapted to \mathcal{F}_n then $\xi_{\tau \wedge n}$ is adapted to \mathcal{F}_n , and that if each ξ_n belongs to $L^1(P)$ then each $\xi_{\tau \wedge n}$ belongs to $L^1(P)$. As each of $\alpha_1, \dots, \alpha_{n+1}$ is \mathcal{F}_n -measurable and is bounded,

$$E(\xi_{\tau \wedge (n+1)} | \mathcal{F}_n) = \sum_{k=1}^{n+1} E(\alpha_k(\xi_k - \xi_{k-1}) | \mathcal{F}_n) = \sum_{k=1}^{n+1} \alpha_k E(\xi_k - \xi_{k-1} | \mathcal{F}_n). \quad (1)$$

⁴Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 56, Proposition 3.2.

Suppose that ξ_n is a submartingale. By (1),

$$\begin{aligned} E(\xi_{\tau \wedge (n+1)} | \mathcal{F}_n) &= \sum_{k=1}^n \alpha_k (\xi_k - \xi_{k-1}) + \alpha_{n+1} E(\xi_{n+1} | \mathcal{F}_n) - \alpha_{n+1} \xi_n \\ &\geq \xi_{\tau \wedge n} + \alpha_{n+1} \xi_n - \alpha_{n+1} \xi_n \\ &= \xi_{\tau \wedge n}, \end{aligned}$$

which shows that $\xi_{\tau \wedge n}$ is a submartingale; the statement that $E(\xi_{\tau \wedge (n+1)} | \mathcal{F}_n) \geq \xi_{\tau \wedge n}$ means that $E(\xi_{\tau \wedge (n+1)} | \mathcal{F}_n)(\omega) \geq \xi_{\tau \wedge n}(\omega)$ for almost all $\omega \in \Omega$. \square

We now prove the **optional stopping theorem**.⁵

Theorem 4 (Optional stopping theorem). *Let \mathcal{F}_n be a filtration of a σ -algebra \mathcal{A} , let ξ_n be a martingale with respect to \mathcal{F}_n , and let τ be a stopping time with respect to \mathcal{F}_n . Suppose that:*

1. For almost all $\omega \in \Omega$, $\tau(\omega) < \infty$.
2. $\xi_\tau \in L^1(\Omega, \mathcal{A}, P)$.
3. $E(\xi_n 1_{\{\tau > n\}}) \rightarrow 0$ as $n \rightarrow \infty$.

Then

$$E(\xi_\tau) = E(\xi_1).$$

Proof. For each n , $\Omega = \{\tau \leq n\} \cup \{\tau > n\}$, and therefore

$$\xi_\tau = \xi_{\tau \wedge n} + \xi_\tau 1_{\{\tau > n\}} - \xi_n 1_{\{\tau > n\}} = \xi_{\tau \wedge n} + \sum_{k=n+1}^{\infty} \xi_k 1_{\{\tau=k\}} - \xi_n 1_{\{\tau > n\}}.$$

Theorem 3 tells us that $\xi_{\tau \wedge n}$ is a martingale with respect to \mathcal{F}_n , and hence

$$E(\xi_{\tau \wedge n}) = E(\xi_{\tau \wedge 1}) = E(\xi_1),$$

so

$$E(\xi_\tau) = E(\xi_1) + \sum_{k=n+1}^{\infty} E(\xi_k 1_{\{\tau=k\}}) - E(\xi_n 1_{\{\tau > n\}}). \quad (2)$$

But as $\xi_\tau \in L^1(P)$,

$$\int_{\Omega} (\xi_\tau)(\omega) dP(\omega) = \sum_{k=1}^{\infty} \int_{\{\tau=k\}} \xi_k(\omega) dP(\omega) = \sum_{k=1}^{\infty} E(\xi_k 1_{\{\tau=k\}}),$$

and the fact that this series converges means that $\sum_{k=n+1}^{\infty} E(\xi_k 1_{\{\tau=k\}}) \rightarrow 0$. With the hypothesis $E(\xi_n 1_{\{\tau > n\}}) \rightarrow 0$, as $n \rightarrow \infty$ we have

$$E(\xi_1) + \sum_{k=n+1}^{\infty} E(\xi_k 1_{\{\tau=k\}}) - E(\xi_n 1_{\{\tau > n\}}) \rightarrow E(\xi_1).$$

But (2) is true for each n , so we get $E(\xi_\tau) = E(\xi_1)$, proving the claim. \square

⁵Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 58, Theorem 3.1.

Suppose that η_n is a sequence of independent random variables each with the Rademacher distribution:

$$P(\eta_n = 1) = \frac{1}{2}, \quad P(\eta_n = -1) = \frac{1}{2}.$$

Let $\xi_n = \sum_{k=1}^n \eta_k$ and let $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$. Because

$$\xi_{n+1}^2 = (\xi_n + \eta_{n+1})^2 = \eta_{n+1}^2 + 2\eta_{n+1}\xi_n + \xi_n^2,$$

we have, as ξ_n is \mathcal{F}_n -measurable and belongs to $L^\infty(P)$ and as η_{n+1} is independent of the σ -algebra \mathcal{F}_n ,

$$\begin{aligned} E(\xi_{n+1}^2 - (n+1) | \mathcal{F}_n) &= E(\eta_{n+1}^2 + 2\eta_{n+1}\xi_n + \xi_n^2 - (n+1) | \mathcal{F}_n) \\ &= E(\eta_{n+1}^2) + 2\xi_n E(\eta_{n+1}) + \xi_n^2 - (n+1) \\ &= 1 + 0 + \xi_n^2 - (n+1) \\ &= \xi_n^2 - n. \end{aligned}$$

Therefore, $\xi_n^2 - n$ is a martingale with respect to \mathcal{F}_n .

Let K be a positive integer and let

$$\tau = \inf\{n \geq 1 : |\xi_n| = K\}.$$

Namely, τ is the time of first entry in the Borel subset $\{-K, K\}$ of \mathbb{R} , hence by Lemma 1 is a stopping time with respect to the filtration \mathcal{F}_n . With some work,⁶ one shows that (i) $P(\tau > 2Kn) \rightarrow 0$ as $n \rightarrow \infty$, (ii) $E(|\xi_\tau^2 - \tau|) < \infty$, and (iii) $E((\xi_n^2 - n)1_{\{\tau > n\}}) \rightarrow 0$ as $n \rightarrow \infty$. Then we can apply the optional stopping theorem to the martingale $\xi_n^2 - n$: we get that

$$E(\xi_\tau^2 - \tau) = E(\xi_1^2 - 1) = E(\xi_1^2) - 1 = E(\eta_1^2) - 1 = 0.$$

Hence

$$E(\tau) = E(\xi_\tau^2).$$

But $|\xi_\tau| = K$, so $\xi_\tau^2 = K^2$, hence

$$E(\tau) = E(K^2) = K^2.$$

6 Maximal inequalities

We now prove **Doob's maximal inequality**.⁷

⁶Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 59, Example 3.7.

⁷Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 68, Proposition 4.1.

Theorem 5 (Doob's maximal inequality). *Suppose that \mathcal{F}_n is a filtration of a σ -algebra \mathcal{A} , that ξ_n is a submartingale with respect to \mathcal{F}_n , and that for each n , $\xi_n \geq 0$. Then for each $n \geq 1$ and $\lambda > 0$,*

$$\lambda P \left(\max_{1 \leq k \leq n} \xi_k \geq \lambda \right) \leq E \left(\xi_n 1_{\{\max_{1 \leq k \leq n} \xi_k \geq \lambda\}} \right).$$

Proof. Define $\zeta_n(\omega) = \max_{1 \leq k \leq n} \xi_k(\omega)$, which is \mathcal{F}_n -measurable, and define $\tau : \Omega \rightarrow \{1, \dots, n\}$ by

$$\tau(\omega) = \min\{1 \leq k \leq n : \xi_k(\omega) \geq \lambda\}$$

if there is some $1 \leq k \leq n$ for which $\xi_k(\omega) \geq \lambda$, and $\tau(\omega) = n$ otherwise. For $1 \leq k \leq n$,

$$\{\tau = k\} = \left(\bigcap_{j=1}^{k-1} \{\xi_j < \lambda\} \right) \cap \{\xi_k \geq \lambda\} \in \mathcal{F}_k,$$

and for $k > n$,

$$\{\tau = k\} = \emptyset \in \mathcal{F}_k,$$

showing that τ is a stopping time with respect to the filtration \mathcal{F}_k .

For $k \geq 1$,

$$\xi_{k+1} - \xi_{\tau \wedge (k+1)} = \sum_{j=1}^k 1_{\{\tau=j\}} (\xi_{k+1} - \xi_{\tau \wedge (k+1)}) = \sum_{j=1}^k 1_{\{\tau=j\}} (\xi_{k+1} - \xi_j),$$

hence, because τ is a stopping time with respect to the filtration \mathcal{F}_k and because ξ_k is a submartingale with respect to this filtration,

$$\begin{aligned} E(\xi_{k+1} - \xi_{\tau \wedge (k+1)} | \mathcal{F}_k) &= \sum_{j=1}^k 1_{\{\tau=j\}} E((\xi_{k+1} - \xi_j) | \mathcal{F}_k) \\ &= \sum_{j=1}^k 1_{\{\tau=j\}} (E(\xi_{k+1} | \mathcal{F}_k) - \xi_j) \\ &\geq \sum_{j=1}^k 1_{\{\tau=j\}} (\xi_k - \xi_j) \\ &= \sum_{j=1}^{k-1} 1_{\{\tau=j\}} (\xi_k - \xi_j) \\ &= \xi_k - \xi_{\tau \wedge k}, \end{aligned}$$

from which we have that the sequence $\xi_k - \xi_{\tau \wedge k}$ is a submartingale with respect to the filtration \mathcal{F}_k . Therefore

$$E(\xi_k - \xi_{\tau \wedge k}) \geq E(\xi_1 - \xi_{\tau \wedge 1}) = E(\xi_1) - E(\xi_{\tau \wedge 1}) = E(\xi_1) - E(\xi_1) = 0,$$

and so $E(\xi_{\tau \wedge k}) \leq E(\xi_k)$. Because $\tau \wedge n = \tau$, this yields

$$E(\xi_\tau) \leq E(\xi_n).$$

We have

$$E(\xi_\tau) = E(\xi_\tau 1_{\{\zeta_n \geq \lambda\}}) + E(\xi_\tau 1_{\{\zeta_n < \lambda\}}).$$

If $\omega \in \{\zeta_n \geq \lambda\}$ then $(\xi_\tau)(\omega) \geq \lambda$, and if $\omega \in \{\zeta_n < \lambda\}$ then $\tau(\omega) = n$ and so $(\xi_\tau)(\omega) = \xi_n(\omega)$. Therefore

$$E(\xi_\tau) \geq E(\lambda \cdot 1_{\{\zeta_n \geq \lambda\}}) + E(\xi_n 1_{\{\zeta_n < \lambda\}}) = \lambda P(\zeta_n \geq \lambda) + E(\xi_n 1_{\{\zeta_n < \lambda\}}).$$

Therefore

$$\lambda P(\zeta_n \geq \lambda) + E(\xi_n 1_{\{\zeta_n < \lambda\}}) \leq E(\xi_n).$$

But $\xi_n = \xi_n 1_{\zeta_n < \lambda} + \xi_n 1_{\zeta_n \geq \lambda}$, hence

$$\lambda P(\zeta_n \geq \lambda) \leq E(\xi_n 1_{\{\zeta_n \geq \lambda\}}),$$

which proves the claim. \square

The following is **Doob's L^2 maximal inequality**, which we prove using Doob's maximal inequality.⁸

Theorem 6 (Doob's L^2 maximal inequality). *Suppose that \mathcal{F}_n is a filtration of a σ -algebra \mathcal{A} and that ξ_n is a submartingale with respect to \mathcal{F}_n such that for each $n \geq 1$, $\xi_n \geq 0$ and $\xi_n \in L^2(P)$. Then for each $n \geq 1$,*

$$E \left(\left| \max_{1 \leq k \leq n} \xi_k \right|^2 \right) \leq 4E(\xi_n^2).$$

Proof. Define $\zeta_n(\omega) = \max_{1 \leq k \leq n} \xi_k(\omega)$. It is a fact that if $\eta \in L^2(P)$ and $\eta \geq 0$ then

$$E(\eta^2) = 2 \int_0^\infty tP(\eta \geq t)dt.$$

Using this, Doob's maximal inequality, Fubini's theorem, and the Cauchy-

⁸Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 68, Theorem 4.1.

Schwarz inequality,

$$\begin{aligned}
E(\zeta_n^2) &= 2 \int_0^\infty tP(\zeta_n > t)dt \\
&\leq 2 \int_0^\infty E(\xi_n 1_{\{\zeta_n \geq t\}})dt \\
&= 2 \int_0^\infty \left(\int_{\{\zeta_n \geq t\}} \xi_n(\omega) dP(\omega) \right) dt \\
&= 2 \int_\Omega \left(\int_0^{\zeta_n(\omega)} dt \right) \xi_n(\omega) dP(\omega) \\
&= 2 \int_\Omega \zeta_n(\omega) \xi_n(\omega) dP(\omega) \\
&\leq 2(E(\zeta_n^2))^{1/2} (E(\xi_n^2))^{1/2}.
\end{aligned}$$

If $E(\zeta_n^2) = 0$ the claim is immediate. Otherwise, we divide this inequality by $(E(\zeta_n^2))^{1/2}$ and obtain

$$(E(\zeta_n^2))^{1/2} \leq 2(E(\xi_n^2))^{1/2},$$

and so

$$E(\zeta_n^2) \leq 4E(\xi_n^2),$$

proving the claim. □

7 Upcrossings

Suppose that ξ_n is a sequence of random variables that is adapted to a filtration \mathcal{F}_n and let $a < b$ be real numbers. Define

$$\tau_0 = 0,$$

and by induction for $m \geq 1$,

$$\sigma_m(\omega) = \inf\{k \geq \tau_{m-1}(\omega) : \xi_k(\omega) \leq a\}$$

and

$$\tau_m(\omega) = \inf\{k \geq \sigma_m(\omega) : \xi_k(\omega) \geq b\},$$

where $\inf \emptyset = \infty$. For each m , τ_m and σ_m are each stopping times with respect to the filtration \mathcal{F}_k . For $n \geq 0$ we define

$$U_n[a, b](\omega) = \sup\{m \geq 0 : \tau_m(\omega) \leq n\}.$$

For $x \in \mathbb{R}$, we write

$$x^- = \max\{0, -x\} = -\min\{0, x\}.$$

namely, the **negative part of x** .

We now prove the **upcrossings inequality**.

Theorem 7 (Upcrossings inequality). *If ξ_n , $n \geq 1$, is a supermartingale with respect to a filtration \mathcal{F}_n and $a < b$, then for each $n \geq 1$,*

$$(b - a)E(U_n[a, b]) \leq E((\xi_n - a)^-).$$

Proof. For $n \geq 1$ and $\omega \in \Omega$, and writing $N = U_n[a, b](\omega)$, for which $N \leq n$, we have

$$\begin{aligned} & \sum_{m=1}^n (\xi_{\tau_m \wedge n}(\omega) - \xi_{\sigma_m \wedge n}(\omega)) \\ = & \sum_{m=1}^N (\xi_{\tau_m \wedge n}(\omega) - \xi_{\sigma_m \wedge n}(\omega)) + \xi_{\tau_{N+1} \wedge n}(\omega) - \xi_{\sigma_{N+1} \wedge n}(\omega) \\ & + \sum_{m=N+2}^n (\xi_{\tau_m \wedge n}(\omega) - \xi_{\sigma_m \wedge n}(\omega)) \\ = & \sum_{m=1}^N (\xi_{\tau_m}(\omega) - \xi_{\sigma_m}(\omega)) + \xi_n(\omega) - \xi_{\sigma_{N+1} \wedge n}(\omega) + \sum_{m=N+1}^n (\xi_n(\omega) - \xi_n(\omega)) \\ = & \sum_{m=1}^N (\xi_{\tau_m}(\omega) - \xi_{\sigma_m}(\omega)) + 1_{\{\sigma_{N+1} \leq n\}}(\omega)(\xi_n(\omega) - \xi_{\sigma_{N+1}}(\omega)) \\ \geq & \sum_{m=1}^N (b - a) + 1_{\{\sigma_{N+1} \leq n\}}(\omega)(\xi_n(\omega) - \xi_{\sigma_{N+1}}(\omega)). \end{aligned}$$

Because $\xi_{\sigma_{N+1}}(\omega) \leq a$, we have

$$(b - a)N \leq 1_{\{\sigma_{N+1} \leq n\}}(\omega)(a - \xi_n(\omega)) + \sum_{m=1}^n (\xi_{\tau_m \wedge n}(\omega) - \xi_{\sigma_m \wedge n}(\omega)).$$

One proves that⁹

$$1_{\{\sigma_{N+1} \leq n\}}(\omega)(a - \xi_n(\omega)) \leq \min\{0, a - \xi_n(\omega)\} = (\xi_n(\omega) - a)^-.$$

Thus

$$(b - a)E(U_n[a, b]) \leq E((\xi_n - a)^-) + \sum_{m=1}^n E(\xi_{\tau_m \wedge n} - \xi_{\sigma_m \wedge n}).$$

Using that ξ_n is a supermartingale, for each $1 \leq m \leq n$,

$$E(\xi_{\tau_m \wedge n} - \xi_{\sigma_m \wedge n}) \leq 0.$$

Therefore

$$(b - a)E(U_n[a, b]) \leq E((\xi_n - a)^-).$$

□

⁹I am not this “one”. I have not sorted out why this inequality is true. In every proof of the upcrossings inequality I have seen there are pictures and things like this are asserted to be obvious. I am not satisfied with that reasoning; one should not have to interpret an inequality visually to prove it.

8 Doob's martingale convergence theorem

We now use the upcrossings inequality to prove **Doob's martingale convergence theorem**.¹⁰

Theorem 8 (Doob's martingale convergence theorem). *Suppose that ξ_n , $n \geq 1$, is a supermartingale with respect to a filtration \mathcal{F}_n and that*

$$M = \sup_n E(|\xi_n|) < \infty.$$

Then there is some $\xi \in L^1(\Omega, \mathcal{A}, P)$ such that for almost all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)$$

and with $E(|\xi|) \leq M$.

Proof. For any $a < b$ and $n \geq 1$, the upcrossings inequality tells us that

$$E(U_n[a, b]) \leq \frac{E(\xi_n - a)^-}{b - a} \leq \frac{E(|\xi_n - a|)}{b - a} \leq \frac{E(|\xi_n| + |a|)}{b - a} \leq \frac{M + |a|}{b - a}.$$

For each $\omega \in \Omega$, the sequence $U_n[a, b](\omega) \in [0, \infty)$ is nondecreasing, so by the monotone convergence theorem,

$$E\left(\lim_{n \rightarrow \infty} U_n[a, b]\right) = \lim_{n \rightarrow \infty} E(U_n[a, b]) \leq \frac{M + |a|}{b - a}.$$

This implies that

$$P\left(\omega \in \Omega : \lim_{n \rightarrow \infty} U_n[a, b](\omega) < \infty\right) = 1.$$

Let

$$A = \bigcap_{a, b \in \mathbb{Q}, a < b} \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} U_n[a, b](\omega) < \infty \right\}.$$

This is an intersection of countably many sets each with measure 1, so $P(A) = 1$.

Let

$$B = \{\omega \in \Omega : \liminf_n \xi_n(\omega) < \limsup_n \xi_n(\omega)\}.$$

If $\omega \in B$, then there are $a, b \in \mathbb{Q}$, $a < b$, such that

$$\liminf_n \xi_n(\omega) < a < b < \limsup_n \xi_n(\omega).$$

It follows from this $\lim_{n \rightarrow \infty} U_n[a, b](\omega) = \infty$. Thus $\omega \notin A$, so $B \cap A = \emptyset$, and because $P(A) = 1$ we get $P(B) = 0$.

¹⁰Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 71, Theorem 4.2.

We define $\xi : \Omega \rightarrow \mathbb{R}$ by

$$\xi(\omega) = \begin{cases} \lim_{n \rightarrow \infty} \xi_n(\omega) & \omega \notin B \\ 0 & \omega \in B, \end{cases}$$

which is Borel measurable. Furthermore, since $|\xi| = \liminf_n |\xi_n|$ almost everywhere, by Fatou's lemma we obtain

$$\begin{aligned} E(|\xi|) &= E(\liminf_n |\xi_n|) \\ &\leq \liminf_n E(|\xi_n|) \\ &\leq \sup_n E(|\xi_n|) \\ &= M. \end{aligned}$$

□

9 Uniform integrability

Let $\xi : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ be a random variable. It is a fact¹¹ that $\xi \in L^1$ if and only if for each $\epsilon > 0$ there is some M such that

$$\int_{\{|\xi| > M\}} |\xi| dP < \epsilon.$$

(One's instinct might be to try to use the Cauchy-Schwarz inequality to prove this. This doesn't work.) Thus, if ξ_n is a sequence in $L^1(\Omega, \mathcal{A}, P)$ then for each $\epsilon > 0$ there are M_n such that, for each n ,

$$\int_{\{|\xi_n| > M_n\}} |\xi_n| dP < \epsilon.$$

A sequence of random variables ξ_n is said to be **uniformly integrable** if for each $\epsilon > 0$ there is some M such that, for each n ,

$$\int_{\{|\xi_n| > M\}} |\xi_n| dP < \epsilon.$$

If a sequence ξ_n is uniformly integrable, then there is some M such that for each n ,

$$\int_{\{|\xi_n| > M\}} |\xi_n| dP < 1,$$

and so

$$E(|\xi_n|) = \int_{\{|\xi_n| \leq M\}} |\xi_n| dP + \int_{\{|\xi_n| > M\}} |\xi_n| dP < \int_{\{|\xi_n| \leq M\}} M dP + 1 \leq M + 1.$$

¹¹Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 73, Exercise 4.3.

The following lemma states that the conditional expectations of an integrable random variable with respect to a filtration is a uniformly integrable martingale with respect to that filtration.¹²

Lemma 9. *Suppose that $\xi \in L^1(\Omega, \mathcal{A}, P)$ and that \mathcal{F}_n is a filtration of \mathcal{A} . Then $E(\xi|\mathcal{F}_n)$ is a martingale with respect to \mathcal{F}_n and is uniformly integrable.*

We now prove that a uniformly integrable supermartingale converges in L^1 .¹³

Theorem 10. *Suppose that ξ_n is a supermartingale with respect to a filtration \mathcal{F}_n , and that the sequence ξ_n is uniformly integrable. Then there is some $\xi \in L^1(\Omega, \mathcal{A}, P)$ such that $\xi_n \rightarrow \xi$ in L^1 .*

Proof. Because the sequence ξ_n is uniformly integrable, there is some M such that for each $n \geq 1$,

$$E(|\xi_n|) \leq M + 1.$$

Thus, because ξ_n is a supermartingale, Doob's martingale convergence theorem tells us that there is some $\xi \in L^1(\Omega, \mathcal{A}, P)$ such that for almost all $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega).$$

Because ξ_n is uniformly integrable and converges almost surely to ξ , the Vitali convergence theorem¹⁴ tells us that $\xi_n \rightarrow \xi$ in L^1 . \square

The above theorem shows in particular that a uniformly integrable martingale converges to some limit in L^1 . The following theorem shows that the terms of the sequence are equal to the conditional expectations of this limit with respect to the natural filtration.¹⁵

Theorem 11. *Suppose that a sequence of random variables ξ_n is uniformly integrable and is a martingale with respect to its natural filtration*

$$\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n).$$

Then there is some $\xi \in L^1(\Omega, \mathcal{A}, P)$ such that $\xi_n \rightarrow \xi$ in L^1 and such that for each $n \geq 1$, for almost all $\omega \in \Omega$,

$$\xi_n(\omega) = E(\xi|\mathcal{F}_n)(\omega).$$

¹²Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 75, Exercise 4.5.

¹³Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 76, Theorem 4.3.

¹⁴V. I. Bogachev, *Measure Theory*, volume I, p. 268, Theorem 4.5.4; <http://individual.utoronto.ca/jordanbell/notes/L0.pdf>, p. 8, Theorem 9.

¹⁵Zdzisław Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, p. 77, Theorem 4.4.

Proof. By Theorem 10, there is some $\xi \in L^1(\Omega, \mathcal{A}, P)$ such that $\xi_n \rightarrow \xi$ in L^1 . The hypothesis that the sequence ξ_n is a martingale with respect to \mathcal{F}_n tells us that for that for $n \geq 1$ and for any $m \geq n$,

$$E(\xi_m | \mathcal{F}_n) = \xi_n,$$

and so for $A \in \mathcal{F}_n$,

$$\int_A \xi_m dP = \int_A E(\xi_m | \mathcal{F}_n) dP = \int_A \xi_n dP.$$

Thus

$$\begin{aligned} \left| \int_A (\xi_n - \xi) dP \right| &= \left| \int_A (\xi_m - \xi) dP \right| \\ &\leq \int_A |\xi_m - \xi| dP \\ &\leq E(|\xi_m - \xi|). \end{aligned}$$

But $E(|\xi_m - \xi|) \rightarrow 0$ as $m \rightarrow \infty$. Since m does not appear in the left-hand side, we have

$$\left| \int_A (\xi_n - \xi) dP \right| = 0,$$

and thus

$$\int_A \xi_n dP = \int_A \xi dP.$$

But $E(f | \mathcal{F}_n)$ is the unique element of $L^1(\Omega, \mathcal{F}_n, P)$ such that for each $A \in \mathcal{F}_n$,

$$\int_A E(f | \mathcal{F}_n) dP = \int_A f dP,$$

and because ξ_n satisfies this, we get that $\xi_n = E(f | \mathcal{F}_n)$ in L^1 , i.e., for almost all $\omega \in \Omega$,

$$\xi_n(\omega) = E(f | \mathcal{F}_n)(\omega),$$

proving the claim. □

10 Lévy's continuity theorem

For a metrizable topological space X , we denote by $\mathcal{P}(X)$ the set of Borel probability measures on X . The **narrow topology on $\mathcal{P}(X)$** is the coarsest topology such that for each $f \in C_b(X)$, the map

$$\mu \mapsto \int_X f d\mu$$

is continuous $\mathcal{P}(X) \rightarrow \mathbb{C}$.

A subset \mathcal{H} of $\mathcal{P}(X)$ is called **tight** if for each $\epsilon > 0$ there is a compact subset K_ϵ of X such that if $\mu \in \mathcal{H}$ then $\mu(X \setminus K_\epsilon) < \epsilon$, i.e. $\mu(K_\epsilon) > 1 - \epsilon$. (An element μ of $\mathcal{P}(X)$ is called tight when $\{\mu\}$ is a tight subset of $\mathcal{P}(X)$.)

For a Borel probability measure μ on \mathbb{R}^d , we define its **characteristic function** $\tilde{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$ by

$$\tilde{\mu}(u) = \int_{\mathbb{R}^d} e^{ix \cdot u} d\mu(x), \quad u \in \mathbb{R}^d.$$

$\tilde{\mu}$ is bounded by 1 and is uniformly continuous. Because $\mu(\mathbb{R}^d) = 1$,

$$\tilde{\mu}(0) = 1.$$

Lemma 12. *Let $\mu \in \mathcal{P}(\mathbb{R})$. For $\delta > 0$,*

$$\mu\left(\left\{x \in \mathbb{R} : |x| \geq \frac{2}{\delta}\right\}\right) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \tilde{\mu}(u)) du;$$

in particular, the right-hand side of this inequality is real.

Proof. Using Fubini's theorem and the fact that all real t , $1 - \frac{\sin t}{t} \geq 0$,

$$\begin{aligned} \int_{-\delta}^{\delta} (1 - \tilde{\mu}(u)) du &= \int_{-\delta}^{\delta} \left(\int_{\mathbb{R}} (1 - e^{ixu}) d\mu(x) \right) du \\ &= \int_{\mathbb{R}} \left(\int_{-\delta}^{\delta} 1 - e^{iux} du \right) d\mu(x) \\ &= \int_{\mathbb{R}} \left(u - \frac{e^{iux}}{ix} \right)_{-\delta}^{\delta} d\mu(x) \\ &= \int_{\mathbb{R}} \left(2\delta - \frac{e^{i\delta x}}{ix} + \frac{e^{-i\delta x}}{ix} \right) d\mu(x) \\ &= 2\delta \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x} \right) d\mu(x) \\ &\geq 2\delta \int_{|\delta x| \geq 2} \left(1 - \frac{\sin(\delta x)}{\delta x} \right) d\mu(x) \\ &\geq 2\delta \int_{|\delta x| \geq 2} \left(1 - \frac{1}{|\delta x|} \right) d\mu(x) \\ &\geq 2\delta \int_{|\delta x| \geq 2} \frac{1}{2} d\mu(x) \\ &= \delta \mu(\{x \in \mathbb{R} : |\delta x| \geq 2\}). \end{aligned}$$

□

The following lemma gives a condition on the characteristic functions of a sequence of Borel probability measures on \mathbb{R} under which the sequence is tight.¹⁶

¹⁶Krishna B. Athreya and Soumendra N. Lahiri, *Measure Theory and Probability Theory*, p. 329, Lemma 10.3.3.

Lemma 13. *Suppose that $\mu_n \in \mathcal{P}(\mathbb{R})$ and that $\tilde{\mu}_n$ converges pointwise to a function $\phi : \mathbb{R} \rightarrow \mathbb{C}$ that is continuous at 0. Then the sequence μ_n is tight.*

Proof. Write $\phi_n = \tilde{\mu}_n$. Because $|\phi_n| \leq 1$, for each $\delta > 0$, by the dominated convergence theorem we have

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi_n(t)) dt \rightarrow \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) dt.$$

On the other hand, that ϕ is continuous at 0 implies that for any $\epsilon > 0$ there is some $\eta > 0$ such that when $|t| < \eta$, $|\phi(t) - 1| < \epsilon$, and hence for $\delta < \eta$,

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) dt \leq 2 \sup_{|t| \leq \delta} |1 - \phi(t)| \leq 2\epsilon,$$

thus

$$\frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) dt \rightarrow 0, \quad \delta \rightarrow 0.$$

Let $\epsilon > 0$. There is some $\delta > 0$ for which

$$\left| \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) dt \right| < \epsilon.$$

Then there is some n_δ such that when $n \geq n_\delta$,

$$\left| \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi_n(t)) dt - \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi(t)) dt \right| < \epsilon,$$

whence

$$\left| \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi_n(t)) dt \right| < 2\epsilon.$$

Lemma 12 then says

$$\mu_n \left(\left\{ x \in \mathbb{R} : |x| \geq \frac{2}{\delta} \right\} \right) \leq \frac{1}{\delta} \int_{-\delta}^{\delta} (1 - \phi_n(t)) dt < 2\epsilon.$$

Furthermore, any Borel probability measure on a Polish space is tight (**Ulam's theorem**).¹⁷ Thus, for each $1 \leq n < n_\delta$, there is a compact set K_n for which $\mu_n(\mathbb{R} \setminus K_n) < \epsilon$. Let

$$K_\epsilon = K_1 \cup \dots \cup K_{n_\delta-1} \cup \left\{ x \in \mathbb{R} : |x| \leq \frac{2}{\delta} \right\},$$

which is a compact set, and for any $n \geq 1$,

$$\mu_n(\mathbb{R} \setminus K_\epsilon) < 2\epsilon,$$

showing that the sequence μ_n is tight. □

¹⁷Alexander S. Kechris, *Classical Descriptive Set Theory*, p. 107, Theorem 17.11.

For metrizable spaces X_1, \dots, X_d , let $X = \prod_{i=1}^d X_i$ and let $\pi_i : X \rightarrow X_i$ be the projection map. We establish that if \mathcal{H} is a subset of $\mathcal{P}(X)$ such that for each $1 \leq i \leq d$ the family of i th marginals of \mathcal{H} is tight, then \mathcal{H} itself is tight.¹⁸

Lemma 14. *Let X_1, \dots, X_d be metrizable topological spaces, let $X = \prod_{i=1}^d X_i$, and let $\mathcal{H} \subset \mathcal{P}(X)$. Suppose that for each $1 \leq i \leq d$,*

$$\mathcal{H}_i = \{\pi_{i*}\mu : \mu \in \mathcal{H}\}$$

is a tight set in $\mathcal{P}(X_i)$. Then \mathcal{H} is a tight set in $\mathcal{P}(X)$.

Proof. For $\mu \in \mathcal{H}$, write $\mu_i = \pi_{i*}\mu$. Let $\epsilon > 0$ and take $1 \leq i \leq d$. Because \mathcal{H}_i is tight, there is a compact subset K_i of X_i such that for all $\mu_i \in \mathcal{H}_i$,

$$\mu_i(X_i \setminus K_i) < \frac{\epsilon}{d}.$$

Let

$$K = \prod_{i=1}^d K_i = \bigcap_{i=1}^d \pi_i^{-1}(K_i).$$

Then for any $\mu \in \mathcal{H}$,

$$\begin{aligned} \mu(X \setminus K) &= \mu\left(X \setminus \bigcap_{i=1}^d \pi_i^{-1}(K_i)\right) \\ &= \mu\left(\bigcup_{i=1}^d \pi_i^{-1}(K_i)^c\right) \\ &= \mu\left(\bigcup_{i=1}^d \pi_i^{-1}(X_i \setminus K_i)\right) \\ &\leq \sum_{i=1}^d \mu(\pi_i^{-1}(X_i \setminus K_i)) \\ &= \sum_{i=1}^d \mu_i(X_i \setminus K_i) \\ &< \sum_{i=1}^d \frac{\epsilon}{d} \\ &= \epsilon, \end{aligned}$$

which shows that \mathcal{H} is tight. □

¹⁸Luigi Ambrosio, Nicola Gigli, and Giuseppe Savare, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, p. 119, Lemma 5.2.2; V. I. Bogachev, *Measure Theory*, volume II, p. 94, Lemma 7.6.6.

We now prove **Lévy's continuity theorem**, which we shall use to prove the martingale central limit theorem.¹⁹

Theorem 15 (Lévy's continuity theorem). *Suppose that $\mu_n \in \mathcal{P}(\mathbb{R}^d)$, $n \geq 1$.*

1. *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\mu_n \rightarrow \mu$ narrowly, then for any $u \in \mathbb{R}^d$,*

$$\tilde{\mu}_n(u) \rightarrow \tilde{\mu}(u), \quad n \rightarrow \infty.$$

2. *If there is some $\phi : \mathbb{R}^d \rightarrow \mathbb{C}$ to which $\tilde{\mu}_n$ converges pointwise and ϕ is continuous at 0, then there is some $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\phi = \tilde{\mu}$ and such that $\mu_n \rightarrow \mu$ narrowly.*

Proof. Suppose that $\mu_n \rightarrow \mu$ narrowly. For each $u \in \mathbb{R}^d$, the function $x \mapsto e^{ix \cdot u}$ is continuous $\mathbb{R}^d \rightarrow \mathbb{C}$ and is bounded, so

$$\tilde{\mu}_n(u) = \int_{\mathbb{R}^d} e^{ix \cdot u} d\mu_n(x) \rightarrow \int_{\mathbb{R}^d} e^{ix \cdot u} d\mu(x) = \tilde{\mu}(u).$$

Suppose that $\tilde{\mu}_n$ converges pointwise to ϕ and that ϕ is continuous at 0. For $1 \leq i \leq d$, let $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ be the projection map and define $\iota_i : \mathbb{R} \rightarrow \mathbb{R}^d$ by taking the i th entry of $\iota_i(t)$ to be t and the other entries to be 0. Fix $1 \leq i \leq d$ and write $\nu_n = \pi_{i*}\mu_n \in \mathcal{P}(\mathbb{R})$, and for $t \in \mathbb{R}$ we calculate

$$\begin{aligned} \tilde{\nu}_n(t) &= \int_{\mathbb{R}} e^{ist} d\nu_n(s) \\ &= \int_{\mathbb{R}^d} e^{i\pi_i(x)t} d\mu_n(x) \\ &= \int_{\mathbb{R}^d} e^{ix \cdot \iota_i(t)} d\mu_n(x) \\ &= \tilde{\mu}_n(\iota_i(t)), \end{aligned}$$

so $\tilde{\nu}_n = \tilde{\mu}_n \circ \iota_i$. By hypothesis, $\tilde{\nu}_n$ converges pointwise to $\phi \circ \iota_i$. Because ϕ is continuous at $0 \in \mathbb{R}^d$, the function $\phi \circ \iota_i$ is continuous at $0 \in \mathbb{R}$. Then Lemma 13 tells us that the sequence ν_n is tight. That is, for each $1 \leq i \leq d$, the set

$$\{\pi_{i*}\mu_n : n \geq 1\}$$

is tight in $\mathcal{P}(\mathbb{R})$. Thus Lemma 14 tells us that the set

$$\{\mu_n : n \geq 1\}$$

is tight in $\mathcal{P}(\mathbb{R}^d)$.

Prokhorov's theorem²⁰ states that if X is a Polish space, then a subset \mathcal{H} of $\mathcal{P}(X)$ is tight if and only if each sequence of elements of \mathcal{H} has a subsequence

¹⁹cf. Jean Jacod and Philip Protter, *Probability Essentials*, second ed., p. 167, Theorem 19.1.

²⁰V. I. Bogachev, *Measure Theory*, volume II, p. 202, Theorem 8.6.2.

that converges narrowly to some element of $\mathcal{P}(X)$. Thus, there is a subsequence $\mu_{a(n)}$ of μ_n and some $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that $\mu_{a(n)}$ converges narrowly to μ . By the first part of the theorem, we get that $\tilde{\mu}_{a(n)}$ converges pointwise to $\tilde{\mu}$. But by hypothesis $\tilde{\mu}_n$ converges pointwise to ϕ , so $\phi = \tilde{\mu}$.

Finally we prove that $\mu_n \rightarrow \mu$ narrowly. Let $\mu_{b(n)}$ be a subsequence of μ_n . Because $\{\mu_n : n \geq 1\}$ is tight, Prokhorov's theorem tells us that there is a subsequence $\mu_{c(n)}$ of $\mu_{b(n)}$ that converges narrowly to some $\lambda \in \mathcal{P}(\mathbb{R}^d)$. By the first part of the theorem, $\tilde{\mu}_{c(n)}$ converges pointwise to $\tilde{\lambda}$. By hypothesis $\tilde{\mu}_{c(n)}$ converges pointwise to ϕ , so $\tilde{\lambda} = \phi = \tilde{\mu}$. Then $\lambda = \mu$. That is, any subsequence of μ_n itself has a subsequence that converges narrowly to μ , which implies that the sequence μ_n converges narrowly to μ . (For a sequence x_n in a topological space X and $x \in X$, $x_n \rightarrow x$ if and only if each subsequence of x_n has a subsequence that converges to x .) \square

11 Martingale central limit theorem

Let γ_d be the **standard Gaussian measure on \mathbb{R}^d** : γ_d has density

$$\frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{1}{2}|x|^2}$$

with respect to Lebesgue measure on \mathbb{R}^d .

We now prove the **martingale central limit theorem**.²¹

Theorem 16 (Martingale central limit theorem). *Suppose X_j is a sequence in $L^3(\Omega, \mathcal{A}, P)$ satisfying the following, with $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$:*

1. $E(X_j | \mathcal{F}_{j-1}) = 0$.
2. $E(X_j^2 | \mathcal{F}_{j-1}) = 1$.
3. *There is some K for which $E(|X_j|^3 | \mathcal{F}_{j-1}) \leq K$.*

Then $\frac{S_n}{\sqrt{n}}$ converges in distribution to some random variable $Z : \Omega \rightarrow \mathbb{R}$ with $Z_*P = \gamma_1$, where

$$S_n = \sum_{j=1}^n X_j.$$

Proof. For positive integers n and j , define

$$\phi_{n,j}(u) = E(e^{iu \frac{1}{\sqrt{n}} X_j} | \mathcal{F}_{j-1}).$$

²¹Jean Jacod and Philip Protter, *Probability Essentials*, second ed., p. 235, Theorem 27.7.

For each $\omega \in \Omega$, by Taylor's theorem there is some $\xi_{n,j}(\omega)$ between 0 and $X_j(\omega)$ such that

$$e^{iu\frac{1}{\sqrt{n}}X_j(\omega)} = 1 + iu\frac{1}{\sqrt{n}}X_j(\omega) - \frac{u^2}{2n}X_j(\omega)^2 - \frac{iu^3}{6n^{3/2}}\xi_{n,j}(\omega)^3.$$

Because $f \mapsto E(f|\mathcal{F}_{j-1})$ is a positive operator and $|\xi_{n,j}|^3 \leq |X_j|^3$, we have, by the last hypothesis of the theorem,

$$E(|\xi_{n,j}|^3|\mathcal{F}_{j-1}) \leq E(|X_j|^3|\mathcal{F}_{j-1}) \leq K \quad (3)$$

we use this inequality later in the proof. Now, using that $E(X_j|\mathcal{F}_{j-1}) = 0$ and $E(X_j^2|\mathcal{F}_{j-1}) = 1$,

$$\begin{aligned} \phi_{n,j}(u) &= 1 + iu\frac{1}{\sqrt{n}}E(X_j|\mathcal{F}_{j-1}) - \frac{u^2}{2n}E(X_j^2|\mathcal{F}_{j-1}) - \frac{iu^3}{6n^{3/2}}E(\xi_{n,j}^3|\mathcal{F}_{j-1}) \\ &= 1 - \frac{u^2}{2n} - \frac{iu^3}{6n^{3/2}}E(\xi_{n,j}^3|\mathcal{F}_{j-1}). \end{aligned}$$

For $p \geq 1$,

$$\begin{aligned} E(e^{iu\frac{1}{\sqrt{n}}S_p}) &= E(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}e^{iu\frac{1}{\sqrt{n}}X_p}) \\ &= E(E(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}e^{iu\frac{1}{\sqrt{n}}X_p}|\mathcal{F}_{p-1})) \\ &= E(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}E(e^{iu\frac{1}{\sqrt{n}}X_p}|\mathcal{F}_{p-1})) \\ &= E(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\phi_{n,p}(u)) \\ &= E\left(e^{iu\frac{1}{\sqrt{n}}S_{p-1}}\left(1 - \frac{u^2}{2n} - \frac{iu^3}{6n^{3/2}}E(\xi_{n,p}^3|\mathcal{F}_{p-1})\right)\right), \end{aligned}$$

which we write as

$$E\left(e^{i\frac{u}{\sqrt{n}}S_p} - \left(1 - \frac{u^2}{2n}\right)e^{i\frac{u}{\sqrt{n}}S_{p-1}}\right) = -E\left(e^{i\frac{u}{\sqrt{n}}S_{p-1}}\frac{iu^3}{6n^{3/2}}E(\xi_{n,p}^3|\mathcal{F}_{p-1})\right).$$

Now using (3) we get

$$\begin{aligned} &\left|E\left(e^{i\frac{u}{\sqrt{n}}S_p} - \left(1 - \frac{u^2}{2n}\right)e^{i\frac{u}{\sqrt{n}}S_{p-1}}\right)\right| \\ &\leq E\left(\left|e^{i\frac{u}{\sqrt{n}}S_{p-1}}\frac{iu^3}{6n^{3/2}}E(\xi_{n,p}^3|\mathcal{F}_{p-1})\right|\right) \\ &= E\left(\frac{|u|^3}{6n^{3/2}}|E(\xi_{n,p}^3|\mathcal{F}_{p-1})|\right) \\ &\leq \frac{|u|^3}{6n^{3/2}} \cdot K. \end{aligned}$$

Let $u \in \mathbb{R}$ and let $n = n(u)$ be large enough so that $0 \leq 1 - \frac{u^2}{2n} \leq 1$. For $1 \leq p \leq n$, multiplying the above inequality by $\left(1 - \frac{u^2}{2n}\right)^{n-p}$ yields

$$\left|\left(1 - \frac{u^2}{2n}\right)^{n-p} E(e^{iu\frac{1}{\sqrt{n}}S_p}) - \left(1 - \frac{u^2}{2n}\right)^{n-p+1} E(e^{iu\frac{1}{\sqrt{n}}S_{p-1}})\right| \leq K \frac{|u|^3}{6n^{3/2}}. \quad (4)$$

Now, because $\sum_{p=1}^n (a_p - a_{p-1}) = a_n - a_0$,

$$\begin{aligned} & \sum_{p=1}^n \left(\left(1 - \frac{u^2}{2n}\right)^{n-p} E(e^{iu \frac{1}{\sqrt{n}} S_p}) - \left(1 - \frac{u^2}{2n}\right)^{n-(p-1)} E(e^{iu \frac{1}{\sqrt{n}} S_{p-1}}) \right) \\ &= E(e^{iu \frac{1}{\sqrt{n}} S_n}) - \left(1 - \frac{u^2}{2n}\right)^n E(e^{iu \frac{1}{\sqrt{n}} S_0}) \\ &= E(e^{iu \frac{1}{\sqrt{n}} S_n}) - \left(1 - \frac{u^2}{2n}\right)^n. \end{aligned}$$

Using this with (4) gives

$$\left| E(e^{iu \frac{1}{\sqrt{n}} S_n}) - \left(1 - \frac{u^2}{2n}\right)^n \right| \leq n \cdot K \frac{|u|^3}{6n^{3/2}} = K \frac{|u|^3}{6n^{1/2}}.$$

But if $|a_n - b_n| \leq c_n$, $c_n \rightarrow 0$, and $b_n \rightarrow b$, then $a_n \rightarrow b$. As

$$\lim_{n \rightarrow \infty} \left(1 - \frac{u^2}{2n}\right)^n = e^{-\frac{u^2}{2}}$$

and $K \frac{|u|^3}{6n^{1/2}} \rightarrow 0$, we therefore get that

$$E(e^{iu \frac{1}{\sqrt{n}} S_n}) \rightarrow e^{-\frac{u^2}{2}}$$

as $n \rightarrow \infty$.

Let $\mu_n = \left(\frac{S_n}{\sqrt{n}}\right)_* P$ and let $\phi(u) = e^{-\frac{u^2}{2}}$. We have just established that $\tilde{\mu}_n \rightarrow \phi$ pointwise. The function ϕ is continuous at 0, so Lévy's continuity theorem tells us that there is a Borel probability measure μ on \mathbb{R} such that $\phi = \tilde{\mu}$ and such that μ_n converges narrowly to μ . But $\phi(u) = e^{-\frac{u^2}{2}}$ is the characteristic function of γ_1 , so we have that μ_n converges narrowly to γ_1 . \square