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

Jordan Bell

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3 Filtrations

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

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

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

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

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

4 Martingales

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

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

In particular,

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

i.e.

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

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

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

In particular,

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

i.e.

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

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

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

In particular,

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

i.e.

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

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

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

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

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

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 $\mathscr{F}_n$ and let $B \in \mathscr{B}_{\mathbb{R}}$. Then*

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

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

[2] Zdzisław 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 adapated to the filtration $\mathscr{F}_k$, $A_k^c \in \mathscr{F}_k$ and $A_n \in \mathscr{F}_n$, and because $\mathscr{F}_k$ is a filtration, the right-hand side of the above belongs to $\mathscr{F}_n$. □

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

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

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

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

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

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

and

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

and therefore

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

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

[3] Zdzisław 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 $\mathscr{F}_n$ be a filtration of a σ-algebra $\mathscr{A}$ and let τ be a stopping time with respect to $\mathscr{F}_n$.*

1. *If ξ_n is a submartingale with respect to $\mathscr{F}_n$ then so is $\xi_{\tau\wedge n}$.*
2. *If ξ_n is a supermartingale with respect to $\mathscr{F}_n$ then so is $\xi_{\tau\wedge n}$.*
3. *If ξ_n is a martingale with respect to $\mathscr{F}_n$ 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 $B \in \mathscr{B}_{\mathbb{R}}$, (i) if $0, 1 \notin B$ then

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

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

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

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

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

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

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

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

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 $\mathscr{F}_n$ then $\xi_{\tau\wedge n}$ is adapted to $\mathscr{F}_n$, 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 $\alpha_1, \ldots, \alpha_{n+1}$ is $\mathscr{F}_n$-measurable and is bounded,

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

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

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

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

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

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

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

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

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 to $\mathscr{F}_n$, and hence

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

so

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

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 $\sum_{k=n+1}^{\infty} E(\xi_k 1_{\{\tau=k\}}) \to 0$. With the hypothesis $E(\xi_n 1_{\{\tau>n\}}) \to 0$, as $n \to \infty$ we have

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

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

[5] Zdzisław 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:

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

Let $\xi_n = \sum_{k=1}^n \eta_k$ and let $\mathscr{F}_n = \sigma(\eta_1, \ldots, \eta_n)$. 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 $\mathscr{F}_n$-measurable and belongs to $L^\infty(P)$ and as η_{n+1} is independent of the σ-algebra $\mathscr{F}_n$,

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

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

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 $\mathscr{F}_n$. With some work,[6] one shows that (i) $P(\tau > 2Kn) \to 0$ as $n \to \infty$, (ii) $E(|\xi_\tau^2 - \tau|) < \infty$, and (iii) $E((\xi_n^2 - n)1_{\{\tau > n\}}) \to 0$ as $n \to \infty$. 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]

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

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

Theorem 5 (Doob's maximal inequality)**.** *Suppose that $\mathscr{F}_n$ is a filtration of a σ-algebra $\mathscr{A}$, that ξ_n is a submartingale with respect to $\mathscr{F}_n$, 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 $\mathscr{F}_n$-measurable, and define $\tau : \Omega \to \{1, \dots, n\}$ 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 $1 \leq k \leq n$,

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

and for $k > n$,

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

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

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 $\mathscr{F}_k$ and because ξ_k is a submartingale with respect to this filtration,

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

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

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 $\mathscr{F}_n$ is a filtration of a σ-algebra $\mathscr{A}$ and that ξ_n is a submartingale with respect to $\mathscr{F}_n$ 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-

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

Schwarz inequality,

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

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

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

and so

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

proving the claim. □

7 Upcrossings

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

$$\tau_0 = 0,$$

and by induction for $m \ge 1$,

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

and

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

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

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

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 $\mathscr{F}_n$ 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

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

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 that[9]

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

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

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

□

[9] I 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 uprossings 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 $\mathscr{F}_n$ and that*

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

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

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

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,

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

This implies that

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

Let

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

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

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

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

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

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 $\xi : (\Omega, \mathscr{A}, P) \to \mathbb{R}$ be a random variable. It is a fact[11] 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 $L^1(\Omega, \mathscr{A}, P)$ 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

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

[11] Zdzisław 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 $\xi \in L^1(\Omega, \mathscr{A}, P)$ and that $\mathscr{F}_n$ is a filtration of $\mathscr{A}$. Then $E(\xi|\mathscr{F}_n)$ is a martingale with respect to $\mathscr{F}_n$ 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 $\mathscr{F}_n$, and that the sequence ξ_n is uniformly integrable. Then there is some $\xi \in L^1(\Omega, \mathscr{A}, P)$ such that $\xi_n \to \xi$ 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 $\xi \in L^1(\Omega, \mathscr{A}, P)$ such that for almost all $\omega \in \Omega$,

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

Because ξ_n is uniformly integrable and converges almost surely to ξ, the Vitali convergence theorem[14] tells us that $\xi_n \to \xi$ in L^1. □

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*

$$\mathscr{F}_n = \sigma(\xi_1, \ldots, \xi_n).$$

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

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

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

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

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

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

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

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

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

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

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 $E(|\xi_m - \xi|) \to 0$ as $m \to \infty$. 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 $E(f|\mathscr{F}_n)$ is the unique element of $L^1(\Omega, \mathscr{F}_n, P)$ such that for each $A \in \mathscr{F}_n$,

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

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

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

proving the claim. □

10 Lévy's continuity theorem

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

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

is continuous $\mathscr{P}(X) \to \mathbb{C}$.

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

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

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

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

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

Lemma 12. *Let $\mu \in \mathscr{P}(\mathbb{R})$. 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]

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

Lemma 13. *Suppose that $\mu_n \in \mathscr{P}(\mathbb{R})$ and that $\tilde{\mu}_n$ converges pointwise to a function $\phi : \mathbb{R} \to \mathbb{C}$ 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

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

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

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

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

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

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. □

[17] Alexander 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 $\pi_i : X \to X_i$ be the projection map. We establish that if $\mathscr{H}$ is a subset of $\mathscr{P}(X)$ such that for each $1 \leq i \leq d$ the family of ith marginals of $\mathscr{H}$ is tight, then $\mathscr{H}$ 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 $\mathscr{H} \subset \mathscr{P}(X)$. Suppose that for each $1 \leq i \leq d$,*

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

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

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

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

Let

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

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

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

which shows that $\mathscr{H}$ is tight. □

[18] Luigi 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 $\mu_n \in \mathscr{P}(\mathbb{R}^d)$, $n \geq 1$.*

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

$$\tilde{\mu}_n(u) \to \tilde{\mu}(u), \qquad n \to \infty.$$

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

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

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

Suppose that $\tilde{\mu}_n$ converges pointwise to ϕ and that ϕ is continuous at 0. For $1 \leq i \leq d$, let $\pi_i : \mathbb{R}^d \to \mathbb{R}$ be the projection map and define $\iota_i : \mathbb{R} \to \mathbb{R}^d$ by taking the ith entry of $\iota_i(t)$ to be t and the other entries to be 0. Fix $1 \leq i \leq d$ and write $\nu_n = \pi_{i*}\mu_n \in \mathscr{P}(\mathbb{R})$, 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 $\mathscr{P}(\mathbb{R})$. Thus Lemma 14 tells us that the set

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

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

Prokhorov's theorem[20] states that if X is a Polish space, then a subset $\mathscr{H}$ of $\mathscr{P}(X)$ is tight if and only if each sequence of elements of $\mathscr{H}$ has a subsequence

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

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

that converges narrowly to some element of $\mathscr{P}(X)$. Thus, there is a subsequence $\mu_{a(n)}$ of μ_n and some $\mu \in \mathscr{P}(\mathbb{R}^d)$ 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 $\mu_n \to \mu$ 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 $\lambda \in \mathscr{P}(\mathbb{R}^d)$. 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 $x \in X$, $x_n \to x$ if and only if each subsequence of x_n has a subsequence that converges to x.) □

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 $L^3(\Omega, \mathscr{A}, P)$ satisfying the following, with $\mathscr{F}_k = \sigma(X_1, \dots, X_k)$:*

1. $E(X_j|\mathscr{F}_{j-1}) = 0$.

2. $E(X_j^2|\mathscr{F}_{j-1}) = 1$.

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

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

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

Proof. For positive integers n and j, define

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

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

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

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

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

For $p \geq 1$,

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

which we write as

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

Now using (3) we get

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

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

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

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

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

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

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

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

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

as $n \to \infty$.

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 $\tilde{\mu}_n \to \phi$ 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.

□