THE MARTINGALE CLT
LECTURE NOTES STK4090/9090 SPRING 2025

EMIL A. STOLTENBERG

Abstract. These notes were written for the course STK4090/9090 *Statistical large sample theory* given at the University of Oslo in the spring of 2025. The martingale central limit theorem for discrete time martingales is proved by way of the Lindeberg swapping technique, assuming the conditional Lindeberg condition and convergence in probability of the predictable quadratic variation. To get the basic ideas of the proof, the clt for independent random variables is proved first. In preparing these notes I have, to various extents, drawn on Hall and Heyde (1980), Helland (1982), Pollard (1984) (for Theorem 4), Shiryaev (1996) (for Lemma 7), Lalley (2014) (for Theorem 9), and Häusler and Luschgy (2015) (for Lemma 8).

1. Preliminaries

In this section we collect a few results that are used in the central limit theorems below. Most of these results have been proven in class, so some or most details in the proofs are left to the reader.

Lemma 1. *On* $(\Omega, \mathcal{F}, \Pr)$ *with* $\mathcal{G} \subset \mathcal{F}$*, the random element* $X\colon \Omega \to (\mathcal{X}, \mathcal{A})$ *is* $\mathcal{G}$*-measurable, while* $Y\colon \Omega \to (\mathcal{Y}, \mathcal{B})$ *is independent of* $\mathcal{G}$*. For any measurable function* $g\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ *such that* $\mathrm{E}\,|g(X,Y)| < \infty$*,*

$$\mathrm{E}\,(g(X,Y)\,|\,\mathcal{G}) = \int g(X,y)\,\mathrm{d}P_Y(y), \quad a.s.,$$

where $P_Y = \Pr Y^{-1}$ *is the distribution of* Y*.*

Proof. Sketch of proof from class: Start with $g(x,y) = I_{A\times B}(x,y)$ for $A \in \mathcal{A}$ and $B \in \mathcal{B}$. For any $G \in \mathcal{G}$, $\mathrm{E}\,I_G \mathrm{E}\,(I_{A\times B}(X,Y)\,|\,\mathcal{G}) = \mathrm{E}\,I_G I_{A\times B}(X,Y)$, and $\mathrm{E}\,I_G \int I_{A\times B}(X,y)\,\mathrm{d}P_Y(y) = \mathrm{E}\,I_G I_A(X)\,\mathrm{E}\,I_B(Y) = \mathrm{E}\,I_G I_{A\times B}(X,Y)$ by the assumed independence. This shows that the lemma is true for all sets of the form $A\times B$, with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Next, apply Dynkin's lemma in order to lift this to indicator functions $g(x,y) = I_C(x,y)$ for the potentially more complicated sets $C \in \mathcal{A}\otimes\mathcal{B}$. Then proceed to simple functions, nonnegative functions via monotone convergence, and finally any integrable function. This proof recipe is sometimes referred to as *bootstrapping.* □

Lemma 2. *If* $X_n \to_d X$ *and* $Y_n - X_n \to_p 0$*, then* $Y_n \to_d Y$*.*

Proof. Proved in class: Let F be a closed set, and define $F^{\varepsilon} = \{x\colon d(x,F) \leq \varepsilon\}$, which is closed. Use the Portmanteau theorem to argue that $\limsup_n \Pr(Y_n \leq F) \leq \Pr(X \in F) + \delta$, for any $\delta > 0$. □

Lemma 3. $X_n \to_d X$ *if and only if* $X_n + \sigma Z_n \to_d X + \sigma Z$ *for each* $\sigma > 0$*, where* $Z_n, Z \sim \mathrm{N}(0,1)$ *are independent of* X_n, X*, respectively.*

Date: February 26, 2026.

Proof. Use the same Portmanteau strategy as in the proof of Lemma 2. $\square$

Theorem 4. *Let (X_n) be a sequence of random variables. Then $X_n \to_d X$ if and only if $\mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X)$ for all $g \in C_b^\infty(\mathbb{R})$, i.e., for all bounded and continuous functions $g\colon \mathbb{R} \to \mathbb{R}$ with bounded and continuous derivatives of all orders.*

Proof. Sketch: Since $C_b^\infty(\mathbb{R}) \subset C_b(\mathbb{R}) = \{\text{bounded and continuous functions}\}$, one way is obvious. For the converse, let $Z_n \sim \mathrm{N}(0,1)$ independent of X_n, and $f \in C_b(\mathbb{R})$, then for $\sigma > 0$, $\mathrm{E}\, f(X_n + \sigma Z_n) = \mathrm{E}\, g_\sigma(X_n)$, where

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \int f(u) \exp(-\tfrac{1}{2}(x-u)^2)\, \mathrm{d}u,$$

and $g_\sigma \in C_b^\infty(\mathbb{R})$. In proving that g_σ is bounded and smooth, the dominated convergence theorem is used for two purposes: To prove continuity of the pth derivative $g_\sigma^{(p)}(x)$ of g_σ, and to prove that we can pass the derivative under the integral sign. For both purposes we need $\mathrm{d}u$-integrable functions $K_p(u)$ such that $|x-u|^p \exp(-\frac{1}{2}(x-u)^2) \leq K_p(u)$ for all $x \in S$, where S is an open interval. If $S = (-\varepsilon, \varepsilon)$ for a given $\varepsilon > 0$, for example, then

$$K_p(u) = \begin{cases} |u+\varepsilon|^p \exp(-\frac{1}{2}(u+\varepsilon)^2), & u < -p^{1/2} - \varepsilon, \\ p^{p/2} \exp(-\frac{1}{2}p), & -p^{1/2} - \varepsilon \leq u < p^{1/2} + \varepsilon, \\ |u-\varepsilon|^p \exp(-\frac{1}{2}(u-\varepsilon)^2), & p^{1/2} + \varepsilon \leq u, \end{cases}$$

does the job. Assuming that $\mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X)$ for all $g \in C_b^\infty(\mathbb{R})$, we have that for any $f \in C_b(\mathbb{R})$ $\mathrm{E}\, f(X_n + \sigma Z_n) = \mathrm{E}\, g_\sigma(X_n) \to \mathrm{E}\, g_\sigma(X) = \mathrm{E}\, f(X + \sigma Z)$, for $Z \sim \mathrm{N}(0,1)$ independent of X, and $\sigma > 0$, and $X_n \to_d X$ follows from Lemma 3. $\square$

For the applications in the clts below, Theorem 4 is redundant. All we need for the clts of the next two sections is that g is a bounded and continuous function with at least *three* bounded and continuous derivatives, infinitely many are not needed. Let $C_b^k(\mathbb{R})$ be the space of bounded and continuous function $g\colon \mathbb{R} \to \mathbb{R}$ with k bounded and continuous derivatives. We may prove that $X_n \to_d X$ if and only if $\mathrm{E}\, g(X_n) \to \mathrm{E}\, g(X)$ for all $g \in C_b^k(\mathbb{R})$, using the exact same methods as in the proof of Theorem 4, but only caring about the k first derivatives this time.

2. The Lindeberg clt

The proof of the martingale clt presented in Section 3 mimics the proof of the clt for independent square integrable random variables ('the Lindeberg clt'). To easier appreciate the basic ideas of the proof, we prove this latter theorem first. By $(X_{n,i})_{1\leq i\leq k_n, n\geq 1}$ being row-wise independent is meant that $X_{n,1}, \ldots, X_{n,k_n}$ are independent for each n. Here k_n is a sequence of integers increasing with n, and quite often $k_n = n$.

Theorem 5. *Let $(X_{n,i})_{1\leq i\leq k_n, n\geq 1}$ be a triangular array of row-wise independent mean zero square integrable random variables. If*

(a) $\sum_{i=1}^{k_n} \mathrm{E}\, X_{n,i}^2 \to \sigma^2$ *for* $\sigma^2 > 0$*;*

(b) $\sum_{i=1}^{k_n} \mathrm{E}\, X_{n,i}^2 I_{|X_{n,i}|\geq\varepsilon} \to 0$ *for each* $\varepsilon > 0$*,*

then $\sum_{i=1}^{k_n} X_{n,i} \to_d \mathrm{N}(0, \sigma^2)$.

Proof. Write $\sigma_{n,i}^2 = \mathrm{E}\, X_{n,i}^2$, and let $g\colon \mathbb{R} \to \mathbb{R}$ be a bounded and continuous function with at least three bounded and continuous derivatives. We are going to show that

$\mathrm{E}\, g(\sum_{i=1}^{k_n} X_{n,i}) \to \mathrm{E}\, g(\sigma\xi)$, with $\xi \sim \mathrm{N}(0,1)$, which is the same as $\sum_{i=1}^{k_n} X_{n,i} \to_d \sigma\xi$. As shown in class, for any $\varepsilon > 0$ there are $\delta > 0$ and K such that

$$|g(r+x) - g(r) - g'(r)x - \tfrac{1}{2}g''(r)x^2| \le \varepsilon x^2 + Kx^2 I_{|x|\ge\delta}, \tag{1}$$

uniformly in r and x. For each n, let $\xi_{n,1},\dots,\xi_{n,k_n}$ be i.i.d. $\mathrm{N}(0,1)$ and set $Z_{n,i} = \sigma_{n,i}\xi_{n,i}$ for $1 \le i \le k_n$ and each n, so that $\sum_{i=1}^{k_n} Z_{n,i} \sim \xi$ with $\xi \sim \mathrm{N}(0,1)$. Define $R_{n,k} = \sum_{i=1}^{k-1} X_{n,i} + \sum_{i=k+1}^{k_n} Z_{n,i}$. By the Lindeberg swapping trick

$$|\mathrm{E}\, g(\sum_{i=1}^{k_n} X_{n,i}) - \mathrm{E}\, g(\sum_{i=1}^{k_n} Z_{n,i})| \le \sum_{k=1}^{k_n} |\mathrm{E}\,\{g(R_{n,k}+X_{n,k}) - g(R_{n,k}+Z_{n,k})\}|. \tag{2}$$

Since $R_{n,k}$ is independent of $X_{n,k}$ and of $Z_{n,k}$, and $\mathrm{E}\, X_{n,k} = \mathrm{E}\, Z_{n,k} = 0$,

$$\mathrm{E}\, g'(R_{n,k})X_{n,k} = \mathrm{E}\, g'(R_{n,k})Z_{n,k} = 0,$$

and, again using the independence $R_{n,k} \perp X_{n,k}$ and $R_{n,k} \perp Z_{n,k}$ and that $\mathrm{E}\, X_{n,k}^2 = \mathrm{E}\, Z_{n,k}^2 = \sigma_{n,k}^2$, it holds that

$$\mathrm{E}\, g''(R_{n,k})X_{n,k}^2 = \mathrm{E}\, g''(R_{n,k})Z_{n,k}^2 = \sigma_{n,k}^2\, \mathrm{E}\, g''(R_{n,k}),$$

These equalities allow us to add $g(R_{n,k}) + g'(R_{n,k})X_{n,k} + \frac{1}{2}g''(R_{n,k})X_{n,k}^2$ and subtract $g(R_{n,k}) + g'(R_{n,k})Z_{n,k} + \frac{1}{2}g''(R_{n,k})Z_{n,k}^2$ inside the expectation on the right in (2). Thus, using the triangle inequality followed by the bound from (1)

$$\begin{aligned}
&|\mathrm{E}\,\{g(R_{n,k}+X_{n,k}) - g(R_{n,k}+Z_{n,k})\}| \\
&\quad = |\mathrm{E}\,\{g(R_{n,k}+X_{n,k}) - g(R_{n,k}) - g'(R_{n,k})X_{n,k} - \tfrac{1}{2}g''(R_{n,k})X_{n,k}^2 \\
&\qquad - [g(R_{n,k}+Z_{n,k}) - g(R_{n,k}) - g'(R_{n,k})Z_{n,k} - \tfrac{1}{2}g''(R_{n,k})Z_{n,k}^2]\}| \\
&\quad \le 2\varepsilon\sigma_{n,k}^2 + K\,\mathrm{E}\, X_{n,k}^2 I_{|X_{n,k}|\ge\delta} + K\,\mathrm{E}\, Z_{n,k}^2 I_{|Z_{n,k}|\ge\delta}.
\end{aligned} \tag{3}$$

Summing up and using Conditions (a) and (b), we then get

$$|\mathrm{E}\, g(\sum_{i=1}^{n} X_{n,i}) - \mathrm{E}\, g(\sum_{i=1}^{n} Z_{n,i})| \le 2\varepsilon\sigma^2 + \sum_{k=1}^{k_n} \mathrm{E}\, Z_{n,k}^2 I_{|Z_{n,k}|\ge\delta} + o(1),$$

as $n \to \infty$, where the $o(1)$ term comprises the convergence in (a) and the Lindeberg condition in (b). Since $Z_{n,i} \sim \sigma_{n,i}\xi_{n,i}$ and $\xi_{n,1},\dots,\xi_{n,k_n}$ are i.i.d. $\mathrm{N}(0,1)$, we get that for $\xi \sim \mathrm{N}(0,1)$

$$\sum_{k=1}^{k_n} \mathrm{E}\, Z_{n,k}^2 I_{|Z_{n,k}|\ge\delta} = \sum_{k=1}^{k_n} \sigma_{n,k}^2 \mathrm{E}\, \xi^2 I_{|\xi|\ge\delta/\sigma_{n,k}} \le \sigma^2\, \mathrm{E}\, \xi^2 I_{|\xi|\ge\delta/\gamma_n} + o(1),$$

where $\gamma_n^2 = \max_{k\le n} \sigma_{n,k}^2$ and the $o(1)$ term comes from Condition (a). But by Condition (b),

$$\gamma_n^2 = \max_{i\le n} \mathrm{E}\, X_{n,i}^2 \le \eta^2 + \sum_{i=1}^{n} \mathrm{E}\, X_{n,i}^2 I_{|X_{n,i}|\ge\eta} \to \eta^2, \quad \text{for any } \eta > 0,$$

and therefore $\gamma_n \to 0$, which implies that $\mathrm{E}\, \xi^2 I_{|\xi|\ge\delta/\gamma_n} \to 0$ since ξ^2 is integrable. We conclude that

$$\limsup_n |\mathrm{E}\, g(\sum_{i=1}^{k_n} X_{n,i}) - \mathrm{E}\, g(\sum_{i=1}^{k_n} Z_{n,i})| \le 2\varepsilon\sigma^2,$$

but since $\varepsilon > 0$ was arbitrary, the theorem is proved. □

3. The martingale CLT

For each n, $(X_{n,i})_{1\le i\le k_n}$ is a sequence of random variables, and $(\mathcal{F}_{n,i})_{0\le i\le k_n}$ is a filtration. The triangular array $(X_{n,i},\mathcal{F}_{n,i})_{1\le i\le k_n,n\ge 1}$ is then a square integrable martingale difference array if for all $1\le i\le k_n$ and $n\ge 1$.

(i) $X_{n,i}$ is $\mathcal{F}_{n,i}$-measurable;
(ii) $\mathrm{E}\,X_{n,i}^2<\infty$;
(iii) $\mathrm{E}\,(X_{n,i}\,|\,\mathcal{F}_{n,i-1})=0$, almost surely.

Given such an array we can construct a sequence of square integrable martingales

$$M_{n,k}=\sum_{i=1}^{k}X_{n,i},\quad 1\le k\le k_n, \tag{4}$$

where these are martingales with respect to $(\mathcal{F}_{n,k})_{0\le k\le k_n}$ for each n. Define also the predictable quadratic variation

$$\langle M_n,M_n\rangle_k=\sum_{i=1}^{k}\mathrm{E}\,(X_{n,i}^2\,|\,\mathcal{F}_{n,i-1}),\quad \text{for } 1\le k\le k_n,$$

and $\langle M_n,M_n\rangle_0=0$, and notice that $\langle M_n,M_n\rangle_k$ is $\mathcal{F}_{n,k-1}$-measurable. Any process with this peek-ahead property is said be *predictable.*

In Theorem 9 we prove that $\sum_{i=1}^{k_n}X_{n,i}\to \mathrm{N}(0,\sigma^2)$ provided

(i) $\langle M_n,M_n\rangle_{k_n}\to_p \sigma^2$ for a constant $\sigma^2>0$;
(ii) $\sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,i-1})\to_p 0$, for each $\varepsilon>0$,

The second of these conditions is referred to as the conditional Lindeberg condition, or simply the Lindeberg condition. Many of the equalities in the following only hold almost surely, without that being explicitly stated each time.

Lemma 6. *If (ii) holds, then* $\max_{j\le k_n}\mathrm{E}\,(X_{n,j}^2\,|\,\mathcal{F}_{n,j-1})\to_p 0$.

Proof. For $\varepsilon>0$, $\max_{j\le k_n}\mathrm{E}\,(X_{n,j}^2\,|\,\mathcal{F}_{n,j-1})\le\varepsilon^2+\sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,i-1})$, and the lemma follows since $\varepsilon>0$ is arbitrary. □

The next lemma is Lenglart's inequality in discrete time. The continuous time version can be found in Jacod and Shiryaev (2003, p. 35).

Lemma 7. *Let $(X_{n,i},\mathcal{F}_{n,i})_{1\le i\le k_n,n\ge1}$ be a square integrable martingale difference array, and let $M_{n,j}$ be as defined in (4). For any $\varepsilon,\delta>0$,*

$$\Pr(\max_{j\le k_n}|M_{n,j}|\ge\varepsilon)\le\frac{\delta}{\varepsilon^2}+\Pr(\langle M_n,M_n\rangle_{k_n}\ge\delta).$$

Proof. Let τ_n be a stopping time with respect to $(\mathcal{F}_{n,j})_{0\le 1\le k_n}$, i.e., τ_n has range $\{0,1,\dots,k_n\}$. Then $\mathrm{E}\,M_{n,\tau_n}^2=\langle M_n,M_n\rangle_{\tau_n}$, essentially by Doob's optional stopping theorem (see, e.g. Shiryaev (1996, pp. 484–486) for a proof). Note that the equality $\mathrm{E}\,M_{n,k}^2=\langle M_n,M_n\rangle_k$ is straightforward to derive, so the difficulty lies in $k\mapsto M_{n,k}^2$ being evaluated in a stopping time. For some $\varepsilon>0$, set

$$r_n=\min\{k\in\{0,1,\dots,\tau_n\}\colon M_{n,k}^2\ge\varepsilon\},$$

and take $r_n=\tau_n$ if the set is empty. Then r_n is a stopping time w.r.t. $(\mathcal{F}_{n,j})_{0\le j\le k_n}$ (since $\{r_n=j\}=\{M_{n,j}^2\ge\varepsilon\}\in\mathcal{F}_{n,j}$ for each j), and

$$\begin{aligned}\mathrm{E}\,\langle M_n,M_n\rangle_{\tau_n}&\ge\mathrm{E}\,\langle M_n,M_n\rangle_{r_n}=\mathrm{E}\,M_{n,r_n}^2\\&\ge\mathrm{E}\,M_{n,r_n}^2 I_{\max_{j\le\tau_n}|M_{n,j}|\ge\varepsilon}\ge\varepsilon^2\Pr(\max_{j\le\tau_n}|M_{n,j}|\ge\varepsilon).\end{aligned}$$

This gives the inequality

$$\Pr(\max_{j\le\tau_n}|M_{n,j}|\ge\varepsilon)\le\frac{\mathrm{E}\,\langle M_n,M_n\rangle_{\tau_n}}{\varepsilon^2},$$

valid for all stopping times τ_n. For any $\varepsilon>0$ and $\delta>0$

$$\Pr(\max_{j\le k_n}M_{n,j}^2\ge\varepsilon^2)\le\Pr(\max_{j\le k_n}M_{n,j}^2\ge\varepsilon^2,\langle M_n,M_n\rangle_{k_n}<\delta)+\Pr(\langle M_n,M_n\rangle_{k_n}\ge\delta).$$

The second probability on the right is fine, so we concentrate on the first one. Define

$$s_n=\min\{k\in\{0,1,\dots,k_n-1\}\colon\langle M_n,M_n\rangle_{k+1}\ge\delta\},$$

and take $s_n=k_n$ if there is no such k. Then s_n is a stopping time because $\langle M_n,M_n\rangle_k$ is predictable. Since $k\mapsto\langle M_n,M_n\rangle_k$ is increasing and $s_n\le k_n$,

$$\max_{j\le k_n}M_{n,j}^2I_{\langle M_n,M_n\rangle_{k_n}<\delta}\le\max_{j\le k_n}M_{n,j}^2I_{\langle M_n,M_n\rangle_{s_n}<\delta}\le\max_{j\le k_n\wedge s_n}M_{n,j}^2,$$

which combined with the inequality just derived gives

$$\Pr(\max_{j\le k_n}M_{n,j}^2\ge\varepsilon^2,\langle M_n,M_n\rangle_{k_n}<\delta)=\Pr(\max_{j\le k_n}M_{n,j}^2I_{\langle M_n,M_n\rangle_{k_n}<\delta}\ge\varepsilon^2)$$
$$\le\Pr(\max_{j\le k_n\wedge s_n}M_{n,j}^2\ge\varepsilon^2)\le\frac{1}{\varepsilon^2}\mathrm{E}\,\langle M_n,M_n\rangle_{k\wedge s_n},$$

and since $\langle M_n,M_n\rangle_{k\wedge s_n}\le\delta$ by the definiton of s_n, the inequality is proved. □

The following lemma provides a reduction that simplifies the proof of the clt.

Lemma 8. *Let $(X_{n,i},\mathcal{F}_{n,i})_{1\le i\le k_n,n\ge1}$ be a square integrable martingale difference array satisfying (i) and (ii). Then there exists a square integrable martingale difference array $(Y_{n,i},\mathcal{G}_{n,i})_{1\le i\le k_n+n,n\ge1}$ satisfying (ii), such that*

$$\langle W_n,W_n\rangle_{k_n+n}=\sigma^2,\ \textit{a.s., for all } n,\quad\textit{and}\quad\sum_{i=1}^{k_n}X_{n,i}-\sum_{i=1}^{k_n+n}Y_{n,i}\to_p0,\ \textit{as } n\to\infty,$$

where $W_{n,k}=\sum_{i=1}^kY_{n,i}$ for $k=1,\dots,k_n+n$.

Proof. Define $\sigma_{n,i}^2=\mathrm{E}\,(X_{n,i}^2\,|\,\mathcal{F}_{n,i-1})$ for $1\le i\le k_n$, $n\ge1$. For each n, let $(a_{n,k})_{k_n+1\le k\le k_n+n}$ be i.i.d. random variables with distribution $\Pr(a_{n,i}=1)=\Pr(a_{n,i}=-1)=1/2$, and define $\mathcal{A}_{n,k}=\sigma(a_{n,k_n+1},\dots,a_{n,k})$ for $k_n+1\le k\le k_n+n$, and for each n take the $(a_{n,n+1},\dots,a_{n,2n})$ to be independent of $\mathcal{F}_{n,k_n}$. Define

$$\tau_n=\max\{k\in\{0,1,\dots,k_n\}\colon\langle M_n,M_n\rangle_k\le\sigma^2\},$$

and notice that $\{\tau_n\ge k\}=\{\langle M_n,M_n\rangle_k\le\sigma^2\}$ since $k\mapsto\langle M_n,M_n\rangle_k$ is increasing for each n, and that $\{\tau_n\ge k\}\in\mathcal{F}_{n,k-1}$ since $\langle M_n,M_n\rangle_k$ is predictable. Define also

$$v_n^2=\sigma^2-\langle M_n,M_n\rangle_{\tau_n},\quad n\ge1,$$

and

$$Y_{n,k}=\begin{cases}X_{n,k}I_{k\le\tau_n}, & 1\le k\le k_n,\\ v_na_{n,k}/\sqrt{n}, & k_n+1\le k\le k_n+n,\end{cases}$$

as well as

$$\mathcal{G}_{n,k}=\begin{cases}\mathcal{F}_{n,k}, & 0\le k\le k_n,\\ \mathcal{F}_{n,k_n}\vee\mathcal{A}_{n,k}, & k_n+1\le k\le k_n+n.\end{cases}$$

Then $\mathcal{G}_{n,k-1} \subset \mathcal{G}_{n,k}$ for all $1 \le k \le k_n + n$, so $(\mathcal{G}_{n,k})_{0\le k\le k_n+n}$ is a filtration for each n, and $Y_{n,k}$ is clearly adapted to this filtration, i.e., $Y_{n,k}$ is $\mathcal{G}_{n,k}$-measurable for each n and k. Importantly, for $1 \le k \le k_n$

$$\mathrm{E}\,(Y_{n,k}\,|\,\mathcal{G}_{n,k}) = \mathrm{E}\,(X_{n,k}I_{k\le\tau_n}\,|\,\mathcal{F}_{n,k}) = I_{k\le\tau_n}\mathrm{E}\,(X_{n,k}\,|\,\mathcal{F}_{n,k}) = 0,$$

and since v_n is $\mathcal{F}_{n,k_n-1}$-measurable, we see that for $k_n + 1 \le k \le k_n + n$

$$\mathrm{E}\,(Y_{n,k}\,|\,\mathcal{G}_{n,k-1}) = \frac{v_n}{\sqrt{n}}\,\mathrm{E}\,(a_{n,k}\,|\,\mathcal{F}_{n,k-1} \vee \mathcal{A}_{n,k-1}) = \frac{v_n}{\sqrt{n}}\,\mathrm{E}\,a_{n,k} = 0,$$

using that $a_{n,k}$ is independent of $\mathcal{F}_{n,k-1} \vee \mathcal{A}_{n,k-1}$. By square integrability of $X_{n,k}$ and $a_{n,k}$, we also have $\mathrm{E}\,Y_{n,k}^2 < \infty$ for all $1 \le k \le k_n + n$ and $n \ge 1$. This shows that $(Y_{n,i}, \mathcal{G}_{n,i})_{1\le i\le k_n+n,n\ge1}$ is a square integrable martingale difference array. Moreover,

$$\begin{aligned}\langle W_n, W_n\rangle_{k_n+n} &= \sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{i\le\tau_n}\,|\,\mathcal{F}_{n,i-1}) + \frac{v_n^2}{n}\sum_{i=k_n+1}^{k_n+n}\mathrm{E}\,(a_{n,k}^2\,|\,\mathcal{F}_{n,k-1}\vee\mathcal{A}_{n,k-1})\\ &= \langle M_n, M_n\rangle_{\tau_n} + v_n^2 = \langle M_n, M_n\rangle_{\tau_n} + (\sigma^2 - \langle M_n, M_n\rangle_{\tau_n}) = \sigma^2,\end{aligned}$$

since $\mathrm{E}\,(a_{n,k}^2\,|\,\mathcal{F}_{n,k-1} \vee \mathcal{A}_{n,k-1}) = \mathrm{E}\,a_{n,k} = 1$. Thus $\langle W_n, W_n\rangle_{k_n+n} = \sigma^2$ a.s. for all n. That it also satisfies the conditional Lindeberg condition follows from its construction and Conditions (i) and (ii): Given $\varepsilon > 0$

$$\begin{aligned}\sum_{i=1}^{k_n+n}\mathrm{E}\,(Y_{n,i}^2 I_{|Y_{n,i}|\ge\varepsilon}\,|\,\mathcal{G}_{n,i-1}) &= \sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{k\le\tau_n} I_{|X_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,i-1})\\ &\quad + \frac{1}{n}\sum_{i=k_n+1}^{k_n+n}\mathrm{E}\,(v_n^2 a_{n,i}^2 I_{|v_n a_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,k_n}\vee\mathcal{A}_{n,i-1})\\ &= \frac{1}{n}\sum_{i=k_n+1}^{k_n+n}\mathrm{E}\,(v_n^2 a_{n,i}^2 I_{|v_n a_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,k_n}\vee\mathcal{A}_{n,i-1}) + o_p(1)\\ &= v_n^2 I_{|v_n|\ge\varepsilon} + o_p(1),\end{aligned}$$

where the $o_p(1)$ is due to by Condition (ii) by which

$$\sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{k\le\tau_n} I_{|X_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,i-1}) \le \sum_{i=1}^{k_n}\mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,i-1}) = o_p(1),$$

and

$$\frac{1}{n}\sum_{i=k_n+1}^{k_n+n}\mathrm{E}\,(v_n^2 a_{n,i}^2 I_{|v_n a_{n,i}|\ge\varepsilon}\,|\,\mathcal{F}_{n,k_n}\vee\mathcal{A}_{n,i-1}) = v_n^2 I_{|v_n|\ge\varepsilon} = v_n^2 I_{|v_n|\ge\varepsilon},$$

since $a_{n,i} \in \{-1, 1\}$ and $\mathcal{F}_{n,k_n}$-measurability of v_n (the equality can also be obtained by Lemma 1). We now show that $v_n^2 = o_p(1)$. To see that it is, introduce the events

$$A_n = \{\langle M_n, M_n\rangle_{k_n} \le \sigma^2\}, \quad n = 1, 2, \ldots.$$

Notice that $\tau_n = k_n$ on A_n, while on A_n^c

$$\langle M_n, M_n\rangle_{\tau_n} + \sigma_{n,\tau_n+1}^2 = \langle M_n, M_n\rangle_{\tau_n+1} > \sigma^2.$$

This entails that on A_n^c

$$\max_{1\le j\le k_n}\sigma_{n,j}^2 \ge \sigma_{n,\tau_n+1}^2 > \sigma^2 - \langle M_n, M_n\rangle_{\tau_n} = v_n^2 \ge 0.$$

Thus for any $\eta > 0$,

$$\begin{aligned}\Pr(v_n^2 \geq \eta) &= \Pr(\{v_n^2 \geq \eta\}, A_n) + \Pr(\{v_n^2 \geq \eta\}, A_n^c) \\ &\leq \Pr(\{|\sigma^2 - \langle M_n, M_n\rangle_{k_n}| \geq \eta\}, A_n) + \Pr(\{\max_{1\leq j\leq k_n} \sigma_{n,j}^2 \geq \eta\}, A_n^c) \\ &\leq \Pr(|\sigma^2 - \langle M_n, M_n\rangle_{k_n}|) + \Pr(\max_{1\leq j\leq k_n} \sigma_{n,j}^2 \geq \eta),\end{aligned}$$

where the first term on the right tends to zero by Condition (i) and the second term tends to zero by Lemma 6. We conclude that

$$\sum_{i=1}^{k_n+n} \mathrm{E}\,(Y_{n,i}^2 I_{|Y_{n,i}|\geq\varepsilon} \,|\, \mathcal{G}_{n,i-1}) \to_p 0,$$

which is to say that $(Y_{n,i}, \mathcal{G}_{n,i})_{1\leq i\leq k_n+n, n\geq 1}$ satisfies the conditional Lindeberg condition. Finally,

$$\begin{aligned}\sum_{i=1}^{k_n+n} Y_{n,i} - \sum_{i=1}^{k_n} X_{n,i} &= \sum_{i=1}^{k_n} X_{n,i} I_{i\leq\tau_n} - \sum_{i=1}^{k_n} X_{n,i} + \frac{v_n}{\sqrt{k_n}} \sum_{i=k_n+1}^{k_n+n} a_{n,i} \\ &= \sum_{i=1}^{k_n} X_{n,i} I_{i>\tau_n} + \frac{v_n}{\sqrt{n}} \sum_{i=k_n+1}^{k_n+n} a_{n,i}.\end{aligned}$$

The second term on the right tends to zero in probability because $v_n \to_p 0$ and $n^{-1/2} \sum_{i=k_n+1}^{k_n+n} a_{n,i} \to_d \mathrm{N}(0,1)$ by the clt for i.i.d. random variables. For the first term on the right, notice that $(X_{n,i} I_{i>\tau_n})_{1\leq i\leq k_n}$ are square integrable martingale differences with respect to $(\mathcal{F}_{n,i})_{0\leq i\leq k_n}$ for each $n \geq 1$. By Lemma 7, to show that $\sum_{i=1}^{k_n} X_{n,i} I_{i>\tau_n} \to_p 0$, it therefore suffices to show that $\sum_{i=1}^{k_n} \mathrm{E}\,(X_{n,i}^2 I_{i>\tau_n} \,|\, \mathcal{F}_{n,i-1})$ tends to zero in probability. But

$$\begin{aligned}\sum_{i=1}^{k_n} \mathrm{E}\,(X_{n,i}^2 I_{i>\tau_n} \,|\, \mathcal{F}_{n,i-1}) &= \sum_{i=1}^{k_n} \sigma_{n,i}^2 I_{i>\tau_n} = \langle M_n, M_n\rangle_{k_n} - \langle M_n, M_n\rangle_{\tau_n} \\ &= \langle M_n, M_n\rangle_{k_n} - \sigma^2 + v_n^2,\end{aligned}$$

where $\langle M_n, M_n\rangle_{k_n} - \sigma^2 \to_p 0$ by (i), and $v_n^2 \to_p 0$ as shown above. This entails that $\sum_{i=1}^{k_n+n} Y_{n,i} - \sum_{i=1}^{k_n} X_{n,i} \to_p 0$, and the lemma is proven. □

Theorem 9. (MARTINGALE CLT). *Let $(X_{n,i}, \mathcal{F}_{n,i})_{1\leq i\leq k_n, n\geq 1}$ be a square integrable martingale difference array, and let $M_{n,j}$ be as defined in (4). If*

(i) *$\langle M_n, M_n\rangle_{k_n} \to_p \sigma^2$ for a constant $\sigma^2 > 0$;*

(ii) *$\sum_{i=1}^{k_n} \mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\geq\varepsilon} \,|\, \mathcal{F}_{n,i-1}) \to_p 0$, for each $\varepsilon > 0$,*

then $M_{n,n} \to_d \mathrm{N}(0, \sigma^2)$.

Proof. By Lemma 2 and Lemma 8 it is enough to prove that $M_{n,n} \to_d \mathrm{N}(0, \sigma^2)$ under the simplifying assumption that

$$\langle M_n, M_n\rangle_{k_n} = \sigma^2, \quad \text{a.s., for all } n, \tag{5}$$

instead of Condition (i). So in the remainder of the proof we assume that (5) holds. Define $\sigma_{n,i}^2 = \mathrm{E}\,(X_{n,i}^2 \,|\, \mathcal{F}_{n,i-1})$, so that $\langle M_n, M_n\rangle_k = \sum_{i=1}^{k} \sigma_{n,i}^2$, and notice that under (5) we have $\langle M_n, M_n\rangle_k \leq \sigma^2$ a.s. for all $1 \leq k \leq n$ and each n. Let $g\colon \mathbb{R} \to \mathbb{R}$ a bounded and continuous function with at least three bounded and continuous derivatives. Given $\varepsilon > 0$, let $\delta > 0$ and K be such that the bound in (1) holds. If we show that

$$\mathrm{E}\, g(\sum_{i=1}^{k_n} X_{n,i}) \to \mathrm{E}\, g(\sigma\xi), \quad \xi \sim \mathrm{N}(0,1), \tag{6}$$

then by Theorem 4 the job is done. For each n, let $\xi_{n,1}, \ldots, \xi_{n,k_n}$ be i.i.d. $\mathrm{N}(0,1)$, independent of $\mathcal{F}_{n,k_n}$, and define

$$Z_{n,i} = \sigma_{n,i}\xi_{n,i}, \quad 1 \le i \le k_n, n \ge 1.$$

Then, due to (5), $\sum_{i=1}^n Z_{n,i} \sim \mathrm{N}(0,\sigma^2)$ (which you may check with using characteristic functions), so proving (6) is the same as proving

$$\mathrm{E}\, g(\sum_{i=1}^{k_n} X_{n,i}) - \mathrm{E}\, g(\sum_{i=1}^{k_n} Z_{n,i}) \to 0$$

Now we proceed as in the proof of Theorem 5. Define

$$R_{n,k} = \sum_{i=1}^{k-1} X_{n,i} + \sum_{i=k+1}^{k_n} Z_{n,i}$$

Using the Lindeberg swapping trick and the tower property of conditional expectation followed by the triangle inequality, we get

$$|\mathrm{E}\,\{g(\sum_{i=1}^{k_n} X_{n,i}) - g(\sum_{i=1}^{k_n} Z_{n,i})\}| \le \mathrm{E} \sum_{k=1}^{k_n} |\mathrm{E}\,\{g(R_{n,k}+X_{n,k}) - g(R_{n,k}+Z_{n,k}) \,|\, \mathcal{F}_{n,k-1}\}|.$$

As in the proof of Theorem 5, the key to what follows is that

$$\mathrm{E}\,(g'(R_{n,k})X_{n,k} \,|\, \mathcal{F}_{n,k-1}) = \mathrm{E}\,(g'(R_{n,k})Z_{n,k} \,|\, \mathcal{F}_{n,k-1}), \quad \text{a.s.}, \tag{7}$$

and

$$\mathrm{E}\,(g''(R_{n,k})X_{n,k}^2 \,|\, \mathcal{F}_{n,k-1}) = \mathrm{E}\,(g''(R_{n,k})Z_{n,k}^2 \,|\, \mathcal{F}_{n,k-1}), \quad \text{a.s.}. \tag{8}$$

Because if these equalities hold, we can add and subtract just as in (3), only inside the conditional expectation $\mathrm{E}\,\{g(R_{n,k}+X_{n,k}) - g(R_{n,k}+Z_{n,k}) \,|\, \mathcal{F}_{n,k-1}\}$ this time, and obtain

$$|\mathrm{E}\,\{g(\sum_{i=1}^{k_n} X_{n,i}) - g(\sum_{i=1}^{k_n} Z_{n,i})\}| \le 2\varepsilon\sigma^2 + K \sum_{i=1}^{k_n} \mathrm{E}\, X_{n,i}^2 I_{|X_{n,i}|\ge\delta} + K \sum_{i=1}^{k_n} \mathrm{E}\, Z_{n,i}^2 I_{|Z_{n,i}|\ge\delta}.$$

For the time being, let us assume that (7) and (8) hold, so that the above display is true. By (ii) and

$$\sum_{i=1}^{k_n} \mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\ge\delta} \,|\, \mathcal{F}_{n,k-1}) \le \langle M_n, M_n\rangle_{k_n} \le \sigma^2, \quad \text{a.s., for all } n$$

by (5), the dominated convergence theorem yields

$$\sum_{i=1}^{k_n} \mathrm{E}\, X_{n,i}^2 I_{|X_{n,i}|\ge\delta} = \mathrm{E} \sum_{i=1}^{k_n} \mathrm{E}\,(X_{n,i}^2 I_{|X_{n,i}|\ge\delta} \,|\, \mathcal{F}_{n,k-1}) \to 0.$$

Set $\gamma_n^2 = \max_{i\le n}\sigma_{n,i}^2$, and let $\phi(z) = \exp(-\frac{1}{2}z^2)/\sqrt{2\pi}$ be the standard normal density. Then

$$\begin{aligned}
\sum_{i=1}^{k_n} \mathrm{E}\, Z_{n,i}^2 I_{|Z_{n,i}|\ge\delta} &= \mathrm{E}\sum_{i=1}^{k_n} \mathrm{E}\,(Z_{n,i}^2 I_{|Z_{n,i}|\ge\delta}\,|\,\mathcal{F}_{n,k-1}) \\
&= \mathrm{E}\sum_{i=1}^{k_n} \sigma_{n,i}^2 \mathrm{E}\,(\xi_{n,i}^2 I_{\sigma_{n,i}|\xi_{n,i}|\ge\delta}\,|\,\mathcal{F}_{n,k-1}) \\
&= \mathrm{E}\sum_{i=1}^{k_n} \sigma_{n,i}^2 \int x^2 I(\sigma_{n,i}|x|\ge\delta)\phi(x)\,\mathrm{d}x \\
&\le \mathrm{E}\sum_{i=1}^{k_n} \sigma_{n,i}^2 \int x^2 I(\gamma_n|x|\ge\delta)\phi(x)\,\mathrm{d}x \\
&= \sigma^2\,\mathrm{E}\int x^2 I(\gamma_n|x|\ge\delta)\phi(x)\,\mathrm{d}x \\
&= \sigma^2\int x^2\,\Pr(\gamma_n|x|\ge\delta)\phi(x)\,\mathrm{d}x \to 0,
\end{aligned}$$

where the third equality comes from Lemma 1, and the last equality from Fubini–Tonelli. To see that $\int x^2\Pr(\gamma_n|x|\ge\delta)\phi(x)\,\mathrm{d}x \to 0$, notice that since $\gamma_n \to_p 0$ by Lemma 6, we have $x^2\Pr(\gamma_n|x|\ge\delta)\to 0$ for each x; and since $x^2\Pr(\gamma_n|x|\ge\delta)\le x^2$ for all n, and $\int x^2\phi(x)\,\mathrm{d}x = 1$, the convergence follows from the dominated convergence theorem. We have now shown that $\limsup_n |\mathrm{E}\,g(\sum_{i=1}^{k_n} X_{n,i}) - \mathrm{E}\,g(\sigma\xi)| \le 2\varepsilon\sigma^2$ for $\xi\sim\mathrm{N}(0,1)$, provided (7) and (8) are true. Since $\varepsilon>0$ was arbitrary, this means that the theorem is proven, provided (7) and (8) are true.

Finally, we show that (5) implies (7)&(8). Let $g^{(1)} = g'$ and $g^{(2)} = g''$. Since $\mathcal{F}_{n,k-1}\subset\mathcal{F}_{n,k_n}$ for all $1\le k\le k_n$, we get from the tower property of conditional expectation and $\mathcal{F}_{n,k_n}$-measurability of $X_{n,k}$ that for $p=1,2$

$$\mathrm{E}\,(g^{(p)}(R_{n,k})X_{n,k}^p\,|\,\mathcal{F}_{n,k-1}) = \mathrm{E}\,(\mathrm{E}\,(g^{(p)}(R_{n,k})\,|\,\mathcal{F}_{n,k_n})X_{n,k}^p\,|\,\mathcal{F}_{n,k-1}). \tag{9}$$

To compute the inner conditional expectation on the right, we need to find the conditional distribution of $R_{n,k} = \sum_{i=1}^{k-1} X_{n,i} + \sum_{i=k+1}^{k_n} Z_{n,i}$ given $\mathcal{F}_{n,k_n}$, and this is where the assumption in (5) comes in. Define

$$v_{n,k}^2 = \sigma^2 - \langle M_n, M_n\rangle_k, \quad 1\le k\le n, n\ge 1,$$

and note that $v_{n,k}^2\ge 0$ and $\sum_{j=k+1}^{k_n}\sigma_{n,j}^2 = v_{n,k}^2$ almost surely by (5), and that $v_{n,k}^2$ is $\mathcal{F}_{n,k-1}$-measurable for each $1\le k\le k_n$. By the tower property and Lemma 1,

$$\begin{aligned}
\mathrm{E}\,\exp(it\sum_{j=k+1}^{k_n} Z_{n,j}) &= \mathrm{E}\,\mathrm{E}\,\{\exp(it\sum_{j=k+1}^{k_n}\sigma_{n,j}\xi_{n,j})\,|\,\mathcal{F}_{n,k_n}\} \\
&= \mathrm{E}\int \exp(it\sum_{j=k+1}^{k_n}\sigma_{n,j}x_j)\phi(x_{k+1})\cdots\phi(x_{k_n})\,\mathrm{d}x_{k+1}\cdots\mathrm{d}x_n \\
&= \mathrm{E}\prod_{j=k+1}^{k_n}\int \exp(it\sigma_{n,j}x)\phi(x)\,\mathrm{d}x = \mathrm{E}\,\exp(-\tfrac{1}{2}t^2\sum_{j=k+1}^{k_n}\sigma_{n,j}^2) \\
&= \mathrm{E}\,\exp(-\tfrac{1}{2}t^2 v_{n,k}^2) = \mathrm{E}\,\exp(itv_{n,k}\xi),
\end{aligned}$$

where $\xi \sim \mathrm{N}(0,1)$ is independent of $\mathcal{F}_{n,k_n}$. But this shows that under (5),

$$R_{n,k} \,|\, \mathcal{F}_{n,k_n} \sim \sum_{j=1}^{k-1} X_{n,j} + v_{n,k}\xi,$$

for $\xi \sim \mathrm{N}(0,1)$ independent of $\mathcal{F}_{n,k-1}$. In other words, conditionally on $\mathcal{F}_{n,k_n}$, the random variable $R_{n,k}$ is equal in distribution to a random variable consisting solely of $\mathcal{F}_{n,k_n}$-measurable components and a component independent of $\mathcal{F}_{n,k_n}$. But this means that we can apply Lemma 1 to the inner conditional expectation on the right in (9),

$$\mathrm{E}\,(g^{(p)}(R_{n,k}) \,|\, \mathcal{F}_{n,k_n}) = \int g^{(p)}(\sum_{j=1}^{k-1} X_{n,j} + v_{n,k}x)\phi(x)\,\mathrm{d}x, \quad \text{a.s.},$$

where we see that the right hand side in $\mathcal{F}_{n,k-1}$-measurable (since $g^{(p)}$ is measurable, but the implication we use here is not completely trivial), and therefore

$$\mathrm{E}\,(g^{(p)}(R_{n,k}) \,|\, \mathcal{F}_{n,k-1}) = \int g^{(p)}(\sum_{j=1}^{k-1} X_{n,j} + v_{n,k}x)\phi(x)\,\mathrm{d}x, \quad \text{a.s.},$$

by the tower property of conditional expectation. Combining this with (9), we get

$$\mathrm{E}\,(g'(R_{n,k})X_{n,k} \,|\, \mathcal{F}_{n,k-1}) = \mathrm{E}\,(g'(R_{n,k}) \,|\, \mathcal{F}_{n,k-1})\,\mathrm{E}\,(X_{n,k} \,|\, \mathcal{F}_{n,k-1}) = 0,$$

and

$$\mathrm{E}\,(g''(R_{n,k})X_{n,k}^2 \,|\, \mathcal{F}_{n,k-1}) = \mathrm{E}\,(g''(R_{n,k}) \,|\, \mathcal{F}_{n,k-1})\,\sigma_{n,k}^2.$$

Since $Z_{n,k} = \sigma_{n,k}\xi_{n,k}$ where $\sigma_{n,k}$ is $\mathcal{F}_{n,k-1}$-measurable and $\xi_{n,k}$ is independent of $R_{n,k}$ and of $\mathcal{F}_{n,k-1}$, we get

$$\mathrm{E}\,(g^{(p)}(R_{n,k})Z_{n,k}^p \,|\, \mathcal{F}_{n,k-1}) = \mathrm{E}\,(g^{(p)}(R_{n,k}) \,|\, \mathcal{F}_{n,k-1})\,\sigma_{n,k}^p\mathrm{E}\,(\xi_{n,k}^p),$$

and since $\mathrm{E}\,(\xi_{n,k}) = 0$ and $\mathrm{E}\,(\xi_{n,k}^2) = 1$, the equalities in (7) and (8) hold, and the theorem is proved. □

References

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic press, New York.

Häusler, E. and Luschgy, H. (2015). *Stable Convergence and Stable Limit Theorems.* Springer, Heidelberg.

Helland, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scandinavian Journal of Statistics*, 9:79–94.

Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes. Second Edition.* Springer, Berlin.

Lalley, S. (2014). The martingale central limit theorem. *Lecture notes, Department of Statistics, University of Chicago.*

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Shiryaev, A. N. (1996). *Probability. Second Edition.* Springer, New York.

Department of Mathematics, University of Oslo

Email address: emilas@math.uio.no