If, for fixed $i$, $B_i^+$ consists of the union of all $B_j$ such that $P^Z (B_j | A \cap A_i) - P^Z(B_j) > 0$ and $B_i^-$ is the union of the remaining $B_j$, then the double sum can be rewritten

$$
\sum_i \left( |P^Z (B_i^+ | A \cap A_i) - P^Z (B_i^+) | + |P^Z (B_i^- | A \cap A_i) - P^Z (B_i^-) | \right) P^Y (A \cap A_i).
$$

The sum between round brackets is bounded above by $2\phi(h)$, by the definition of $\phi$. Thus the display is bounded above by $2\phi(h) P^Y (A)$.

4.14 Theorem. If $X_t$ is a strictly stationary time series with mean zero such that $E|X_t|^p < \infty$ and $\sum_h \phi(h)^{1/p} \hat{\phi}(h)^{1/q} < \infty$ for some $p, q > 0$ with $p^{-1} + q^{-1} = 1$, then the series $v = \sum \gamma_X(h)$ converges absolutely and $\sqrt{n} X_n \rightsquigarrow N(0, v)$.

Proof. For a given $M > 0$ let $X_t^M = X_t 1\{|X_t| \leq M\}$ and let $Y_t^M = X_t - X_t^M$. Because $X_t^M$ is a measurable transformation of $X_t$, it is immediate from the definition of the mixing coefficients that $Y_t^M$ is mixing with smaller mixing coefficients than $X_t$. Therefore, by (4.2) and Lemma 4.13,

$$
\text{var} \sqrt{n}(X_n - \overline{X_n^M}) \leq 2 \sum_h \phi(h)^{1/p} \hat{\phi}(h)^{1/q} \|Y_0^M\|_p \|Y_n^M\|_q.
$$

As $M \to \infty$, the right side converges to zero, and hence the left side converges to zero, uniformly in $n$. This means that we can reduce the problem to the case of uniformly bounded time series $X_t$, as in the proof of Theorem 4.7.

Because the $\alpha$-mixing coefficients are bounded above by the $\phi$-mixing coefficients, we have that $\sum_h \alpha(h) < \infty$. Therefore, the second part of the proof of Theorem 4.7 applies without changes.

4.5 Martingale Differences

The martingale central limit theorem applies to the special time series for which the partial sums $\sum_{t=1}^n X_t$ are a martingale (as a process in $n$), or equivalently the increments $X_t$ are "martingale differences". In Chapter 13 score processes (the derivative of the log likelihood) will be an important example of application.

The martingale central limit theorem can be seen as another type of generalization of the ordinary central limit theorem. The partial sums of an i.i.d. sequence grow by increments $X_t$ that are independent from the "past". The classical central limit theorem shows that this induces asymptotic normality, provided that the increments are centered and not too big (finite variance suffices). The mixing central limit theorems relax the independence to near independence of variables at large time lags, a condition that involves the whole distribution. In contrast, the martingale central limit theorem imposes conditions on the conditional first and second moments of the increments given the past, without directly involving other aspects of the distribution, and in this sense is closer to
the ordinary central limit theorem. The first moments given the past are assumed zero; the second moments given the past must not be too big.

The “past” can be given by an arbitrary filtration. A filtration \( \mathcal{F}_t \) is a nondecreasing collection of \( \sigma \)-fields \( \cdots \subset \mathcal{F}_1 \subset \mathcal{F}_0 \subset \mathcal{F}_t \subset \cdots \). The \( \sigma \)-field \( \mathcal{F}_t \) can be thought of as the “events that are known” at time \( t \). Often it will be the \( \sigma \)-field generated by the variables \( X_t, X_{t-1}, X_{t-2}, \ldots \); the corresponding filtration is called the natural filtration of the time series \( X_t \), or the filtration generated by this series. A martingale difference series relative to a given filtration is a time series \( X_t \) such that

(i) \( X_t \) is \( \mathcal{F}_t \)-measurable;
(ii) \( E(X_t | \mathcal{F}_{t-1}) = 0 \).

The second requirement implicitly includes the assumption that \( E[X_t] < \infty \), so that the conditional expectation is well defined; the identity is understood to be in the almost-sure sense.

### 4.15 Exercise

Show that a martingale difference series with finite variances is a white noise series.

### 4.16 Theorem

If \( X_t \) is a martingale difference series relative to the filtration \( \mathcal{F}_t \) such that \( n^{-1} \sum_{t=1}^{n} E(X_t^2 | \mathcal{F}_{t-1}) \overset{P}{\to} v \) for a positive constant \( v \), and such that \( n^{-1} \sum_{t=1}^{n} E(X_t^2 1_{|X_t| \geq \varepsilon \sqrt{n}} | \mathcal{F}_{t-1}) \overset{P}{\to} 0 \) for every \( \varepsilon > 0 \), then \( \sqrt{n} X_n \sim N(0, v) \).

**Proof.** For simplicity of notation, let \( E_t \) denote conditional expectation given \( \mathcal{F}_{t-1} \). Because the events \( A_i = \{ n^{-1} \sum_{j=1}^{t} E_{j-1} X_j^2 \leq 2v \} \) are \( \mathcal{F}_{t-1} \)-measurable, the variables \( X_{n,t} = n^{-1/2} X_t 1_{A_i} \) are martingale differences relative to the filtration \( \mathcal{F}_t \). They satisfy

\[
\sum_{t=1}^{n} E_{t-1} X_{n,t}^2 \leq 2v,
\]

(4.6)

\[
\sum_{t=1}^{n} E_{t-1} X_{n,t}^2 \overset{P}{\to} v,
\]

\[
\sum_{t=1}^{n} E_{t-1} X_{n,t}^2 1_{|X_{n,t}| \geq \varepsilon \sqrt{n}} \overset{P}{\to} 0, \quad \text{every } \varepsilon > 0.
\]

To see this note first that \( E_{t-1} X_{n,t}^2 = 1_{A_i} n^{-1} E_{t-1} X_t^2 \), and that the events \( A_i \) are decreasing: \( A_1 \supset A_2 \supset \cdots \supset A_n \). The first relation in the display, then follows from the definition of the events \( A_i \). The second relation follows, because the probability of the event \( \cap_{t=1}^{n} A_t = A_n \) tends to 1 by assumption, and the left side of this relation is equal to \( n^{-1} \sum_{t=1}^{n} E_{t-1} X_t^2 \) on \( A_n \), which tends to \( v \) by assumption. The third relation is immediate from the conditional Lindeberg assumption on the \( X_t \) and the fact that \( |X_{n,t}| \leq X_t / \sqrt{n} \), for every \( t \).

On the event \( A_n \) we also have that \( X_{n,t} = X_t / \sqrt{n} \) for every \( t = 1, \ldots, n \) and hence the theorem is proved once it has been established that \( \sum_{t=1}^{n} X_{n,t} \sim N(0, v) \).

We have that \( |e^{ix} - 1 - ix| \leq x^2 / 2 \) for every \( x \in \mathbb{R} \). Furthermore, the function \( R: \mathbb{R} \to \mathbb{C} \) defined by \( e^{ix} - 1 - ix + x^2 / 2 = x^2 R(x) \) satisfies \( |R(x)| \leq 1 \) and \( R(x) \to 0 \) as
\( x \to 0. \) If \( \delta, \varepsilon \geq 0 \) are chosen such that \( |R(x)| < \varepsilon \) if \( |x| \leq \delta \), then \( |e^{ix} - 1 - ix + x^2/2| \leq x^2 1_{|x| > \delta} + \varepsilon x^2. \) For fixed \( u \in \mathbb{R} \) define
\[
R_{n,t} = E_{t-1}(e^{iuX_{n,t}} - 1 - iuX_{n,t}).
\]
It follows that \( |R_{n,t}| \leq \frac{1}{2} u^2 E_{t-1}X_{n,t}^2 \) and
\[
\sum_{t=1}^{n} |R_{n,t}| \leq \frac{1}{2} u^2 \sum_{t=1}^{n} E_{t-1}X_{n,t}^2 \leq u^2 v,
\]
\[
\max_{1 \leq t \leq n} |R_{n,t}| \leq \frac{1}{2} u^2 \left( \sum_{t=1}^{n} E_{t-1}X_{n,t}^2 1_{|X_{n,t}| > \delta} + \delta^2 \right),
\]
\[
\sum_{t=1}^{n} |R_{n,t} + \frac{1}{2} u^2 E_{t-1}X_{n,t}^2| \leq u^2 \sum_{t=1}^{n} \left( E_{t-1}X_{n,t}^2 1_{|X_{n,t}| > \delta} + \varepsilon E_{t-1}X_{n,t}^2 \right).
\]
The second and third inequalities together with (4.6) imply that the sequence \( \max_{1 \leq t \leq n} |R_{n,t}| \) tends to zero in probability and that the sequence \( \sum_{t=1}^{n} R_{n,t} \) tends in probability to \(-\frac{1}{2} u^2 v. \)

The function \( S: \mathbb{R} \to \mathbb{R} \) defined by \( \log(1 - x) = -x + xS(x) \) satisfies \( S(x) \to 0 \) as \( x \to 0. \) It follows that \( \max_{1 \leq t \leq n} |S(R_{n,t})| \to 0, \) and
\[
\prod_{t=1}^{n} (1 - R_{n,t}) = e - \sum_{t=1}^{n} R_{n,t} + \sum_{t=1}^{n} R_{n,t}S(R_{n,t}) \to e^{-u^2v/2}.
\]
We also have that \( \prod_{t=1}^{k} |1 - R_{n,t}| \leq \exp \sum_{t=1}^{n} |R_{n,t}| \leq e^{u^2v}, \) for every \( k \leq n. \) Therefore, by the dominated convergence theorem, the convergence in the preceding display is also in absolute mean.

For every \( t, \)
\[
E_{n-1}e^{iuX_{n,n}}(1 - R_{n,n}) = (1 - R_{n,n})E_{n-1}(e^{iuX_{n,n}} - 1 - iuX_{n,n} + 1)
\]
\[
= (1 - R_{n,n}) (R_{n,n} + 1) = 1 - R_{n,n}^2.
\]

Therefore, by conditioning on \( F_{n-1}, \)
\[
E \prod_{t=1}^{n} e^{iuX_{n,t}} (1 - R_{n,t}) = E \prod_{t=1}^{n-1} e^{iuX_{n,t}} (1 - R_{n,t}) - E \prod_{t=1}^{n-1} e^{iuX_{n,t}} (1 - R_{n,t}) R_{n,n}^2.
\]

By repeating this argument, we find that
\[
\left| E \prod_{t=1}^{n} e^{iuX_{n,t}} (1 - R_{n,t}) - 1 \right| = \left| - \sum_{k=1}^{n} E \prod_{t=1}^{k-1} e^{iuX_{n,t}} (1 - R_{n,t}) R_{n,k}^2 \right| \leq e^{u^2v} E \sum_{t=1}^{n} R_{n,t}^2.
\]
This tends to zero, because \( \sum_{t=1}^{n} |R_{n,t}| \) is bounded above by a constant and \( \max_{1 \leq t \leq n} |R_{n,t}| \) tends to zero in probability.
We combine the results of the last two paragraphs to conclude that

\[ \left| \prod_{t=1}^{n} e^{i u X_{n,t}} e^{u^2 v/2} - 1 \right| = \left| \prod_{t=1}^{n} e^{i u X_{n,t}} e^{u^2 v/2} - \prod_{t=1}^{n} e^{i u X_{n,t}} (1 - R_{n,t}) \right| + o(1) \to 0. \]

The theorem follows from the continuity theorem for characteristic functions. ■

Apart from the structural condition that the sums \( \sum_{t=1}^{n} X_t \) form a martingale, the martingale central limit theorem requires that the sequence of variables \( Y_t = \mathbb{E}(X^2_t | F_{t-1}) \) satisfies a law of large numbers and that the variables \( Y_{t,\epsilon,n} = \mathbb{E}(X^2_t 1_{|X_t| > \epsilon \sqrt{n}} | F_{t-1}) \) satisfy a (conditional) Lindeberg-type condition. These conditions are immediate for “ergodic” sequences, which by definition are (strictly stationary) sequences for which any “running transformation” of the type \( Y_t = g(X_t, X_{t-1}, \ldots) \) satisfies the Law of Large Numbers. The concept of ergodicity is discussed in Section 7.2.

4.17 Corollary. If \( X_t \) is a strictly stationary, ergodic martingale difference series relative to its natural filtration with mean zero and \( v = \mathbb{E}X^2_t < \infty \), then \( \sqrt{n} X_n \Rightarrow \mathcal{N}(0, v) \).

Proof. By strict stationarity there exists a fixed measurable function \( g: \mathbb{R}^\infty \to \mathbb{R}^\infty \) such that \( \mathbb{E}(X^2_t | X_{t-1}, X_{t-2}, \ldots) = g(X_{t-1}, X_{t-2}, \ldots) \) almost surely, for every \( t \). The ergodicity of the series \( X_t \) is inherited by the series \( Y_t = g(X_{t-1}, X_{t-2}, \ldots) \) and hence \( \overline{Y}_n \to \mathbb{E}Y_1 = \mathbb{E}X^2_t \) almost surely. By a similar argument the averages \( \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(X^2_t 1_{|X_t| > M} | F_{t-1}) \) converge almost surely to their expectation, for every fixed \( M \). This expectation can be made arbitrarily small by choosing \( M \) large. The sequence \( \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}(X^2_t 1_{|X_t| > \epsilon \sqrt{n}} | F_{t-1}) \) is bounded by this sequence eventually, for any \( M \), and hence converges almost surely to zero. ■

* 4.6 Projections

Let \( X_t \) be a centered time series and \( \mathcal{F}_0 = \sigma(X_0, X_{-1}, \ldots) \). For a suitably mixing time series the covariance \( \mathbb{E}(X_n \mathbb{E}(X_j | \mathcal{F}_0)) \) between \( X_n \) and the best prediction of \( X_j \) at time 0 should be small as \( n \to \infty \). The following theorem gives a precise and remarkably simple sufficient condition for the central limit theorem in terms of these quantities.

4.18 Theorem. Let \( X_t \) be a strictly stationary, mean zero, ergodic time series with \( \sum_{h} |\gamma_X(h)| < \infty \) and, as \( n \to \infty \),

\[ \sum_{j=0}^{\infty} |\mathbb{E}(X_n \mathbb{E}(X_j | \mathcal{F}_0))| \to 0. \]