# Chapter 4

# Uniform CLT for Markov chains with a general state space

## 4.1 Introduction

Let $(S, \mathcal{G}, P)$ be a probability space and let $\mathcal{F}$ be a set of measurable functions on $S$ with an envelope function $F$ finite everywhere. Let $X_1, X_2, ...$ be a strictly stationary sequence of random variables with distribution $P$, and define the empirical measures $P_n$, based on $\{X_i\}$, as $P_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$. We say the uniform CLT holds over $\mathcal{F}$, if $n^{\frac{1}{2}}(P_n - P)$ converges in law, in the space $l^\infty(\mathcal{F})$ to a Gaussian process. Of course, $l^\infty(\mathcal{F})$ is not separable unless $\mathcal{F}$ is a finite set, but Giné and Zinn [7, p56] includes a suitable definition of weak convergence in non-separable spaces.

Define the covering number of $\mathcal{F}$ with respect to $L^p(S, Q)$ by letting, for $\varepsilon > 0$, $N_p(\varepsilon, \mathcal{F}, Q)$ be the minimum $m$ for which there exists $g_1, ..., g_m$ in $L^p(Q)$ such that, for all $f \in \mathcal{F}$, $\| f - g_i \|_{L^p(Q)} < \varepsilon$, for some $1 \leq i \leq m$. Pollard [12] defined a combinatorial entropy

$$N_2(\varepsilon, \mathcal{F}) = \sup_Q (N_2(\varepsilon, \mathcal{F}, Q))$$

where the sup is taken on all the measures on $S$ with finite support, and proved a uniform CLT for the functions class $\mathcal{F}$ with envelope $F$ in $L^2(P)$

satisfying

$$\int_0^\infty [\log N_2(\varepsilon, \mathcal{F})]^{\frac{1}{2}} d\varepsilon < \infty. \tag{4.1}$$

Dudley [6] proved the above combinatorial condition (4.1) is satisfied in the case where the subgraphs of the functions in $\mathcal{F}$ are a VC class of sets. Levental [9] extended Pollard's result to regenerative processes, whose renewal times satisfy

$$E(N_2 - N_1)^{2+\gamma} < \infty \text{ for some } \gamma > 0,$$

for a uniformly bounded family of functions satisfying the condition (4.1). He also applied the regenerative process result to Harris recurrent Markov chains by embedding Markov chains into a regenerative structure [2, chapter 1].

More precisely, Levental proved a uniform CLT for Markov chains over uniformly bounded classes of functions satisfying (4.1), provided

$$\sup_{x \in C} E_x(\tau_C)^{2+\gamma} < \infty \tag{4.2}$$

for some $\gamma > 0$ where $\tau_C = \min\{k : X_{km} \in C\}$, $C$ is a small set, and $m$ is the order of $C$ [2, section 1 of chapter 1].

We will weaken the condition (4.2) to ergodicity of degree 2 and generalize the family of functions from uniformly bounded to the condition that its envelope function $F$ is in $L^2$ and the CLT holds for $F$.

Using an empirical process CLT for stationary sequences satifying a mixing condition [1], one can obtain similar results. However, these results require the envelope function of the function class in $L^p(P)$ for some $p > 2$. We will give a example such that our conditions hold, but the envelope function $F \notin L^p(P)$ for all $p > 2$. For a uniformly bounded family of functions, the mixing approach needs the absolutely regular mixing coefficient

$$\beta_k = O(k^{-\gamma}) \quad \text{for some } \gamma > 1. \tag{4.3}$$

We also give another example such that $E(N_2 - N_1)^2 < \infty$ i.e. our condition holds, but (4.3) fails.

## 4.2 Uniform CLT for regenerative processes

Let $(S, \mathcal{G})$ be a measurable space and let $\mathcal{F}$ be a family of $\mathcal{G}/\mathcal{B}(\mathbf{R})$ measurable real functions on $S$, where $\mathcal{B}(\mathbf{R})$ is the borel $\sigma$-algebra of $\mathbf{R}$. Since we need to take a supremun over $\mathcal{F}$ and $\mathcal{F}$ is not necessary countable, some restrictions have to be made on $\mathcal{F}$ for measurability considerations. (see [8, p8] and [13, appendix C])

**Definition 4.1.**
(i) We call a set **analytic** if it is a subset of Polish space which is a continuous image of some Polish space. (A Polish space is a complete separable metric space, and see [5, chapter 3] for properties of analytic sets.)
(ii) $\mathcal{F}$ is called a **permissible** set of functions if
(a) $\mathcal{F}$ can be identified (set theory isomorphism) as an analytic subset of a compact metric space and
(b) the function $g : S \times \mathcal{F} \to \mathbf{R}$ defined by $g(s, f) = f(s)$ for $f \in \mathcal{F}$ and $s \in S$ is $\mathcal{G} \times \mathcal{B}(\mathcal{F})/\mathcal{B}(\mathbf{R})$ measurable, where $\mathcal{B}(\mathcal{F})$ is the borel $\sigma$-algebra generated by the metric on $\mathcal{F}$.

With the above definitions, the measurability problems in this chapter can be treated as in [8, p8]. For examples of permissible sets, observe that a countable collection of measurable functions is permissible, and also that $\mathcal{F} = \{1_{(a,b)} : a, b \in \mathbf{R}, b > a\}$ is a permissible set.

A regenerative process, informally speaking, is a stochastic process that can be divided into blocks which are identically distributed and independent. To state the results, we need a formal definition and some notations [9].
(i) $S$ is a set and $\mathcal{G}$ is a $\sigma$-algebra of subsets in $S$.
(ii) $\Omega$ stands for the set of all sequences $\{y_i\}_{1 \le i < \infty}$ such that $y_i = (x_i, \phi_i)$ where $x_i \in S$ and $\phi_i \in \{0, 1\}$.
(iii) $\Sigma$ is the minimal $\sigma$-algebra that makes all the coordinate maps $X_n : \Omega \to S$ defined by $X_n(\{y_i\}) = x_n$ and $\Phi_n : \Omega \to \{0, 1\}$ defined by $\Phi_n(\{y_i\}) = \phi_n$, $\Sigma/\mathcal{G}$ and $\Sigma/2^{\{0,1\}}$ measurable respectively.
(iv) $P$ is a probability measure on $\{\Omega, \Sigma\}$.
(v) $N_i = \min\{j \ge 1 : \sum_{1 \le k \le j} \Phi_k = i\}, i = 1, ...$ or $N_i = \infty$ if the set that we minimize over is empty. $\{N_i\}$ are called renewal times. For every $i \ge 1$ $N_i$ is a stopping time relative to the increasing sequence of $\sigma$-algebras $(\sigma\{W_1, ..., W_n\})_{1 \le n}$ where by $W_n$ we denote the coordinate maps $W_n(\{y_i\}) = y_n$. $\mathcal{G}_{N_i}$ is the $\sigma$-algebra associated with the stopping time $N_i$, i.e. : $\mathcal{G}_{N_i} =$

$\sigma\{W_{k\wedge N_i} : k = 1, 2, ...\}$. $\theta_k$ is a shift operator: $\theta_k : \{y_i\}_{i\geq 1} \to \{y_{i+k}\}_{i\geq 1}$ for every $k \geq 1$.

**Definition 4.2**. $\{X_i\}$ will be called a **regenerative process** if $N_i < \infty$ almost surely for every $i \geq 1$ and if for every $f : \Omega \to \mathbf{R}$ which is bounded and $\Sigma/\mathcal{B}(\mathbf{R})$ measurable,

$$E[f(\theta_{N_i}) \mid \mathcal{G}_{N_i}] = E[f(\theta_{N_1})].$$

The following two properties of the process $\{W_i\}$ are equivalent to the above definition:
(i) The post $N_i+1$ process is independent of the occurence up to and including $N_i$, and
(ii) $\mathcal{L}((W_{N_1+1}, ...)) = \mathcal{L}((W_{N_i+1}, ...)) \quad$ for all $i = 1, 2, ....$

We assume that $E(N_2 - N_1) < \infty$ and denote $\mu = E(N_2 - N_1)$ throughout the paper. Define

$$\pi(A) = \frac{1}{\mu} E\Big(\sum_{N_1 < j \leq N_2} 1_A(X_j)\Big) \quad \text{for all } A \subseteq S.$$

Then $\pi$ is a probability measure on $S$ (called a steady state distribution), and by [10, chapter 10] is the usual invariant probability measure for a Markov chain. For all $f \in L^1(S, \pi)$

$$n^{-1} \sum_{k=1}^{n} f(X_k) \to \pi(f) \quad \text{a.s..}$$

See [3, theorem 1, p92] for the proof. (where the statement is formulated for Markov chains but the same proof will work for regenerative processes). Define the centered sum

$$S_n(f) = \sum_{j=1}^{n} (f(X_j) - \pi(f)).$$

We have the following theorem.

**Theorem 4.3**. Suppose $E(N_2 - N_1)^2 < \infty$. Let $\mathcal{F}$ be a permissible family of functions on $S$ satisfying

$$\int_0^\infty [\log N_2(\varepsilon, \mathcal{F})]^{\frac{1}{2}} d\varepsilon < \infty, \tag{4.4}$$

and such that the envelope function $F$ satisfies

$$E[\sum_{N_1<j\leq N_2} F(X_j)]^2 < \infty. \tag{4.5}$$

Then the uniform CLT holds over $\mathcal{F}$. That is, $\{n^{-\frac{1}{2}}S_n(f)\}_{f\in\mathcal{F}}$ converges in law, as random elements of $l^\infty(\mathcal{F})$, to a Gaussian process indexed $\mathcal{F}$ whose sample paths are bound and uniformly continuous with respect to the metric $L^2(S,\pi)$.

**Remark.** The condition (4.5) implies $F \in L^2(S,\pi)$ as in the remark after Theorem 2.2 in Chapter 2. Hence suppose $\{X_i\}$ is a positive recurrent Markov chain taking values in a countable state space and let $N_i$ be the $i$-th hitting of a fixed state. Then $\{X_i\}$ is a regenerative process with renewal times $N_i$ and we can apply Theorem 4.3 to it.

**Proof of Theorem 4.3**. We follow the proof of Theorem 4.9 in [9]. First we denote

$$S' = \bigcup_{n=1}^{\infty} S^n$$

and define

$$Z_k = (X_{N_k+1}, ..., X_{N_{k+1}})$$

for $k = 1,2,3,..$ on $S'$. Let $P' = \mathcal{L}(Z_1)$, and $P'_n$ be the $n$-th empirical measure of $P'$. For $f : S \to \mathbf{R}$, $f' : S' \to \mathbf{R}$ is defined by

$$(y_1, ..., y_n) \to \sum_{k=1}^{n} f(y_k).$$

Denote $Y_k(f) = f'(Z_k)$. Let $\mathcal{F}' = \{f' : f \in \mathcal{F}\}$. Levental showed that $\mathcal{F}'$ is permissble if $\mathcal{F}$ is permissble in his Lemma (2.11) (d). Define

$$[\delta] = \{(f,g) : f,g \in F, \| f-g \|_{L_2(\pi)} < \delta\}$$

and

$$[[\delta]] = \{(f,g) : f,g \in F, \| f-g \|_{L_2(P')} < \delta\}.$$

The main step of the proof is the following lemma.

**Lemma 4.4.** Suppose $E(N_2 - N_1)^2 < \infty$, and (4.4) and (4.5) hold. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_n P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid \sum_{k=1}^{n} (Y_k - \mu\pi)(f - g) \mid > \varepsilon) < \varepsilon.$$

**Proof**. We follow the proof of Lemma 4.2 in [9]. The main new step of the proof is to show that for every $\lambda > 0$, there exists $\delta > 0$ such that $[\delta] \subseteq [[\lambda]]$.

Fix $\lambda > 0$ and choose $M > 0$ such that

$$E[(\sum_{N_1 < j \le N_2} 2F(X_j))^2 1_{\{N_2 - N_1 > M\}}] < \frac{\lambda^2}{2}.$$

Take $\delta = \lambda/\sqrt{2M\mu}$, let $(f, g) \in [\delta]$, we have to show $(f, g) \in [[\lambda]]$. Write $h = f - g$, then $h' = f' - g'$ and $\| h \|_{L^2(\pi)} < \delta$. Now

$$\begin{aligned} &(\| h' \|_{L^2(P')})^2 = E[\textstyle\sum_{N_1 < j \le N_2} h(X_j)]^2 \\ &= E[(\textstyle\sum_{N_1 < j \le N_2} h(X_j))^2 1_{\{N_2 - N_1 > M\}}] + E[(\textstyle\sum_{N_1 < j \le N_2} h(X_j))^2 1_{\{N_2 - N_1 \le M\}}]. \end{aligned}$$

The left term is less than $\frac{\lambda^2}{2}$ since $\mid h \mid \le 2F$. The right term

$$\begin{aligned} E[(\textstyle\sum_{N_1 < j \le N_2} h(X_j))^2 1_{\{N_2 - N_1 \le M\}}] &\le E[(\textstyle\sum_{N_1 < j \le N_2} h^2(X_j))(N_2 - N_1) 1_{\{N_2 - N_1 \le M\}}] \\ &\le ME[\textstyle\sum_{N_1 < j \le N_2} h^2(X_j)], \end{aligned}$$

and

$$ME[\sum_{N_1 < j \le N_2} h^2(X_j)] = M(\| h \|_{L^2(\pi)})^2 \mu < M\delta^2\mu = \frac{\lambda^2}{2}.$$

Thus $\| h' \|_{L^2(P')} < \lambda$ and hence $(f, g) \in [[\lambda]]$.

We need to show that for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_n P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid (P'_n - P')(f' - g') \mid > \varepsilon) < \varepsilon.$$

It is enough to show proved that for every $\varepsilon > 0$, there exists $\lambda > 0$ such that

$$\limsup_n P(\sup_{[[\lambda]]} n^{-\frac{1}{2}} \mid (P'_n - P')(f' - g') \mid > \varepsilon) < \varepsilon. \tag{4.6}$$

We use Lemma 15 in [13, p150]. Since $\mathcal{F}'$ has an envelope $F'$ in $L^2(P')$ by (4.5), we only have to show that for every $\varepsilon > 0$ there exists $a > 0$ such that

$$\limsup_n P\left(\int_0^a [\log N_2(u, \mathcal{F}', P_n')]^{\frac{1}{2}} du > \varepsilon\right) < \varepsilon. \tag{4.7}$$

Since

$$\begin{aligned} \|f'\|_{L^2(P_n')} &= \left(\tfrac{1}{n}\sum_{k=1}^n [\sum_{N_k<j\le N_{k+1}} f(X_j)]^2\right)^{\frac{1}{2}} \\ &\le \left(\tfrac{1}{n}\sum_{k=1}^n [\sum_{N_k<j\le N_{k+1}} f^2(X_j)][N_{k+1}-N_k]\right)^{\frac{1}{2}} \\ &= \|f\|_{L^2(Q)} \cdot \left(\tfrac{1}{n}\sum_{k=1}^n [N_{k+1}-N_k]^2\right)^{\frac{1}{2}}, \end{aligned}$$

where $Q = \sum_{k=1}^n [\sum_{N_k<j\le N_{k+1}} \delta_{X_j} \cdot (N_{k+1}-N_k)] / \sum_{k=1}^n [N_{k+1}-N_k]^2$, we have

$$\|f' - g'\|_{L^2(P_n')} \le \varepsilon$$

for $f, g$ satisfying $\|f-g\|_{L^2(Q)} \le \varepsilon / [\frac{1}{n}\sum_{k=1}^n [N_{k+1}-N_k]^2]^{\frac{1}{2}}$. Thus for $n$ large enough, outside of a set of small probability, we have

$$\begin{aligned} N_2(\varepsilon, \mathcal{F}', P_n') &\le N_2(\varepsilon/[\tfrac{1}{n}\textstyle\sum_{k=1}^n [N_{k+1}-N_k]^2]^{\frac{1}{2}}, \mathcal{F}, Q) \\ &\le N_2(\varepsilon/2[E[N_2-N_1]^2]^{\frac{1}{2}}, \mathcal{F}, Q) \\ &\le N_2(\varepsilon/2[E[N_2-N_1]^2]^{\frac{1}{2}}, \mathcal{F}). \end{aligned}$$

The last inequality will lead to (4.7) by (4.4).

**Lemma 4.5**. Let $\mathcal{F}$ be a permissible family of functions on $S$ satisfying

$$\frac{1}{n}\log N_1(\varepsilon, \mathcal{F}, P_n) \to 0 \quad \text{in probability for every } \varepsilon > 0, \tag{4.8}$$

and such that its envelope function $F$ is in $L^1(S, \pi)$. Then

$$\sup_{f\in F} \frac{1}{n}\left|\sum_{k=1}^n [Y_k(f) - \mu\pi(f)]\right| \to 0 \quad \text{a.s..}$$

**Proof.** Levental proved this result for $\mathcal{F}$ uniformly bounded in Lemma (3.2) of [9]. Using the same method as in the first observation of the proof of Lemma (3.6) in Levental we can obtain the desired result.

With some easy modification of Levental's proof, we also have the following Lemma.

**Lemma 4.6.** Let $E(N_2 - N_1)^2 < \infty$, and assume (4.4) and (4.5) hold. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\limsup_n P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid S_n(f) - S_n(g) \mid > \varepsilon) < \varepsilon.$$

**Proof**. This follows from Levental's proof of Lemma (4.6), i.e.

$$\begin{aligned} |S_n(f) - S_n(g)| \;&\le \left|\textstyle\sum_{j=1}^{N_1} [(f-g)(X_j) - \pi(f-g)]\right| + \left|\textstyle\sum_{k=1}^{l(n)-1} (Y_k - \mu\pi)(f-g)\right| \\ + \left|\textstyle\sum_{j=N_{l(n)-1}+1}^{n} [(f-g)(X_j) - \pi(f-g)]\right| &+ \left|\left[\mu(l(n)-1) - N_{l(n)-1} + N_1\right] \cdot \pi(f-g)\right| \end{aligned}$$

where $l(n) = \max\{k : N_k \le n\}$.

The first term on the right hand side of the inequality does not depend on $n$, thus it converges to zero when divided by $\sqrt{n}$. The second term goes to zero following Levental since Lemma 4.5 gives the uniform SLLN for unbounded $\mathcal{F}$. For the third term, applying the SLLN to the sequence $\{Y_k^2(F), k \ge 1\}$, we have

$$n^{-1} \sum_{k=1}^{n} Y_k^2(F) \to EY_1^2(F) < \infty \quad a.s.$$

and hence $n^{-\frac{1}{2}} Y_n(F) \to 0$ a.s.. Since $l(n)/n \to 1/\mu$ a.s., we have $n^{-\frac{1}{2}} Y_{l(n)-1}(F) \to 0$ a.s.. Note that

$$\sup_{[\delta]} \mid \sum_{j=N_{l(n)-1}+1}^{n} (f-g)(X_j) \mid \le Y_{l(n)-1}(2F),$$

thus

$$\sup_{[\delta]} n^{-\frac{1}{2}} \mid \sum_{j=N_{l(n)-1}+1}^{n} (f-g)(X_j) \mid \to 0 \quad \text{a.s.}.$$

Applying the constant function $\pi(2F)$ to the above function $F$, we can obtain

$$\sup_{[\delta]} n^{-\frac{1}{2}} \mid \sum_{j=N_{l(n)-1}+1}^{n} \pi(f-g) \mid \to 0 \quad \text{a.s.}.$$

Thus the third term is also done. The fourth term follows as in Levental's proof.

To finish the proof of Theorem 4.3, we now use [13, theorem 21, p157]. Following Levental's proof, we have asymptotically stochastic equicontinuity from Lemma 4.6, so only need to make sure that $\mathcal{F}$ is totally bounded in $L^2(S,\pi)$. Note that to show that $\mathcal{F}$ is totally bounded in $L^2(S,\pi)$ in a discrete space, we only need $F \in L^2(S,\pi)$. It is easy to show that if $\mathcal{F}$ is totally bounded in $L^1(S,\pi)$ and its envelope function $F$ is in $L^2(S,\pi)$, then $\mathcal{F}$ is totally bounded in $L^2(S,\pi)$. We have $F \in L^2(S,\pi)$ from (4.5) and Levental proved that $\mathcal{F}$ is totally bounded in $L^1(S,\pi)$ by letting $\mathcal{K} = \{| f-g |: f,g \in \mathcal{F}\}$ and $Q_n$ be the n-th empirical measure of an i.i.d. process whose law is $\pi$. Since for every $\varepsilon > 0$

$$N_1(\varepsilon,\mathcal{K},Q_n) \leq \Big(N_2(\frac{\varepsilon}{2},\mathcal{F})\Big)^2 < \infty$$

we see that the conditions of the uniform SLLN for $Q_n$ are satisfied and we have

$$\sup_{h\in\mathcal{K}} |(Q_n-\pi)h| \to 0 \quad \text{a.s..}$$

Hence it follows that $\mathcal{F}$ is totally bounded in $L^1(S,\pi)$.

## 4.3 Uniform CLT for Markov chains

We generalize Theorem 4.3 to weakly regenerative processes, which are needed for application to ergodic Markov chains.

**Definition 4.7.** Using the notations of section 2, we call $\{X_i\}$ a **weakly regenerative process** if $N_i < \infty$ almost surely for every $i \geq 1$ and if for every $f : \Omega \to \mathbf{R}$ which is bounded and $\Sigma/\mathcal{B}(\mathbf{R})$ measurable,

$$E[f(\theta_{N_{i+1}}) \mid \mathcal{G}_{N_i} \vee \sigma(N_{i+1}-N_i)] = E[f(\theta_{N_1})].$$

The following two properties of the process $\{W_i\}$ are equivalent to the above definition:
(i) The post $N_{i+1}+1$ process is independent of the occurence up to and

including $N_i$, and
(ii) $\mathcal{L}((W_{N_1+1},...)) = \mathcal{L}((W_{N_i+1},...))$ for all $i = 1, 2, ....$

We put $S^* = S \cup \{*\}$ where $*$ denotes an ideal point which is not in $S$. We define two processes $\{E_i\}$ and $\{O_i\}$, taking values in $S^*$:
$O_i = X_i$ if $N_{2k+1} < i \leq N_{2k+2}$ for some $k \geq 0$ and $O_i = *$ otherwise
$E_i = X_i$ if $N_{2k} < i \leq N_{2k+1}$ for some $k \geq 0$ and $E_i = *$ otherwise.
Then $\{O_i\}$ and $\{E_i\}$ are regenerative processes taking values in $S^*$, with renewal times $\{N_{2i}\}_{i\geq 1}$ and $\{N_{1i+1}\}_{i\geq 1}$ respectively. Every function $f : S \to \mathbf{R}$ will be considered as defined on $S^*$ with the identification $f(*) = 0$.

Levental's weakly regenerative process approach can now be applied to unbounded $\mathcal{F}$. That is, we have the following theorem.

**Theorem 4.8**. Let $\{X_i\}$ be a weakly regenerative process with renewal times $N_i$ satisfying

$$E(N_2 - N_1)^2 < \infty.$$

Let $\mathcal{F}$ be a permissible family of functions on $S$ with envelope function $F$. If

$$\int_0^\infty [\log N_2(\varepsilon, \mathcal{F})]^{\frac{1}{2}} d\varepsilon < \infty,$$

and

$$E[\sum_{N_1 < j \leq N_2} F(X_j)]^2 < \infty,$$

then the uniform CLT holds over $\mathcal{F}$, i.e. $\{n^{-\frac{1}{2}} S_n(f)\}_{f\in\mathcal{F}}$ converges in law to a Gaussian process indexed $\mathcal{F}$ in the sense of Theorem 4.3.

**Proof.** We need stochastic equicontinuity, i.e.

$$\lim_n \sup P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid S_n(f) - S_n(g) \mid > \varepsilon) < \varepsilon. \tag{4.9}$$

Let $S_n^O(f) = \sum_{j=1}^n (f(O_j) - \frac{1}{2}\pi(f))$ and $S_n^E(f) = \sum_{j=1}^n (f(E_j) - \frac{1}{2}\pi(f))$. Since $f(X_j) = f(O_j) + f(E_j)$

$$\begin{aligned} P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid S_n(f) - S_n(g) \mid > \varepsilon) \leq\ & P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid S_n^O(f) - S_n^O(g) \mid > \varepsilon/2) \\ & + P(\sup_{[\delta]} n^{-\frac{1}{2}} \mid S_n^E(f) - S_n^E(g) \mid > \varepsilon/2). \end{aligned}$$

Thus (4.9) follows from applying Lemma 4.6 to $\{O_j\}$ and $\{E_j\}$.

Consider a Markov chain $\{X_i\}_{i\geq 0}$ with state space $(S, \mathcal{G})$, transition probability $P(x, A)$ and $n$-step transition probability $P^n(x, A)$ for each $n \geq 1$.

The chain $\{X_i\}_{i\geq 0}$ is called **irreducible** if there exists a $\sigma$-finite measure $\varphi$ on $(S, \mathcal{G})$ which we call an irreducibility measure for $\{X_i\}_{i\geq 0}$, such that for every $A \in \mathcal{G}$

$$\varphi(A) > 0 \quad \Longrightarrow \quad \sum_{n=1}^{\infty} P^n(x, A) > 0 \quad \text{for all } x \in S.$$

There exists a measure $\varphi$ on $(S, \mathcal{G})$ which we call maximal irreducibility measure for $\{X_i\}_{i\geq 0}$, such that $\varphi$ is an irreducibility measure and that all other irreducibility measures are absolutely continuous w.r.t. $\varphi$ [11, proposition 2.4]. Write

$$\mathcal{G}^+ = \{A \in \mathcal{G} : \varphi(A) > 0\},$$

where $\varphi$ is a maximal irreducibility measure.

The chain $\{X_i\}_{i\geq 0}$ is called **Harris recurrent** if

$$P_x(X_n \in A \ \ i.o.) = 1$$

for all $x \in S$ and all $A \in \mathcal{G}^+$.

A $\sigma$-finite measure $\pi$ is called invariant if

$$\pi(A) = \int P(x, A)\pi(dx) \quad \text{for all } A \in \mathcal{G}.$$

$\{X_i\}_{i\geq 0}$ is called **positive** if the invariant measure exists uniquely and is finite.

A irreducible Markov chain satisfies the following **minorzation** condition [11, theorem 2.1]. That is, there exists $C \in \mathcal{G}^+$ such that

$$P^m \geq bI_C \otimes \nu \tag{4.10}$$

for some integer $m \geq 1$, some $b > 0$, and some probability measure $\nu$ on $(S, \mathcal{G})$. In this case we call $C$ a **small set** and the smallest $m$ satisfying (4.10) is called the **order** of $C$.

In the case that $S$ is a countable state space we can take $C = \{x_0\}$ and $\nu(A) = P(x_0, A)$, thus $m = 1$.

Fix a small set $C$ together with a probability measure $\nu$ on $(S, \mathcal{G})$ such that $(C, \nu)$ satisfies (4.10) for some $m \geq 1$ and $\nu(C) > 0$. Let $d$ be the greatest common divisor of the set

$$\{m \geq 1: \text{ there exists } b > 0 \text{ such that (4.10) holds}\}$$

By [11, theorem 2.2], $d$ does not depend on the particular choise of $(C, \nu)$. The chain $\{X_i\}_{i \geq 0}$ is called **aperiodic** if $d = 1$.

The chain $\{X_i\}_{i \geq 0}$ is called **ergodic** if it is positive, Harris recurrent and aperiodic. Let

$$\tau_A = \inf\{n \geq 1 : X_n \in A\}.$$

The chain $\{X_i\}_{i \geq 0}$ is called **ergodic of degree 2** [11, proposition 5.16 and section 6.4] if it is ergodic and

$$E_\pi \tau_A < \infty \quad \text{for all } A \in \mathcal{G}^+, \tag{4.11}$$

where $\pi$ is the invariant measure.

Using the split chain technique [2, chapter 1], ergodic Markov chains can be embedded in a weakly regenerative structure. That is, if $\{X_i\}$ is a ergodic Markov chain, then by Theorem 2.2 of chapter 1 in Chen, $\{X_i\}$ is a weakly regenerative process with renewal times

$$N_i = m\tau(i) + m - 1,$$

where $\tau(i)$ is defined in [2, (2.19) of chapter 1] and $m$ is the order of some fixed small set.

¿From Proposition 5.16 of Nummelin [11], (4.11) is equivalent to

$$\int_A \pi(dx) E_x \tau_A^2 < \infty \quad \text{for all } A \in \mathcal{G}^+. \tag{4.12}$$

Hence (4.31) of chapter 1 in Chen holds by taking $\varphi(s) = s^2$ and $\xi \equiv 1$ in Theorem 4.1 of chapter 1 in Chen. Furthermore, by Theorem 4.6 of chapter 1 in Chen we have $E(N_2 - N_1)^2 < \infty$. Thus ergodicity of degree 2 implies $E(N_2 - N_1)^2 < \infty$.

By Theorem 2.3 of chapter 2 in Chen, $F \in L^2(\pi)$ and that $n^{-\frac{1}{2}} S_n(F)$ converges in law to a normal distribution, imply

$$E[\sum_{N_1 < j \leq N_2} (F - \pi(F))(X_j)]^2 < \infty.$$

Then, since $E(N_2 - N_1)^2 < \infty$, we have

$$E[\sum_{N_1 < j \leq N_2} F(X_j)]^2 < \infty.$$

Hence we have the following theorem.

**Theorem 4.9.** Let $\{X_i\}$ be a Markov chain which is ergodic of degree 2. Let $\mathcal{F}$ be a permissible family of functions on $S$ with envelope function $F \in L^2(\pi)$ such that

$$\int_0^\infty [\log N_2(\varepsilon, \mathcal{F})]^{\frac{1}{2}} d\varepsilon < \infty, \tag{4.13}$$

and assume $n^{-\frac{1}{2}} S_n(F)$ converges in law to a normal distribution. Then the uniform CLT holds over $\mathcal{F}$, i.e. $\{n^{-\frac{1}{2}} S_n(f)\}_{f \in \mathcal{F}}$ converges in law to a Gaussian process indexed $\mathcal{F}$ in the sense of Theorem 4.3.

## 4.4 Compare to mixing results

First we introduce a combinatorial condition on a class of sets. Let $\mathcal{C}$ be a class of subsets of $S$. For $x_1, \cdots, x_n \in S$, let

$$\Delta^{\mathcal{C}}(x_1, \cdots, x_n) := \text{ card } \{C \cap \{x_1, \cdots, x_n\} : C \in \mathcal{C}\} \leq 2^n$$

and

$$m^{\mathcal{C}}(n) := \sup\{\Delta^{\mathcal{C}}(x_1, \cdots, x_n) : x_1, \cdots, x_n \in S\}.$$

If $m^{\mathcal{C}}(n) < 2^n$ for some $n \geq 1$, then $\mathcal{C}$ is called a **Vapnik-Cervonenkis** (or VC) class. Many classes of interest in applications, such as the class of all rectangles in $\mathbf{R}^d$, are VC classes.

A class $\mathcal{F}$ of real functions on $S$ is called a **VC graph class** if the class

$$\mathcal{R} := \{\{(s,x) : 0 \leq x \leq f(s) \text{ or } f(s) \leq x \leq 0\} : f \in \mathcal{F}\}$$

of regions in $S\times$ $\mathbf{R}$ which lie between $S \times \{0\}$ and the graph of some $f \in \mathcal{F}$ is a VC class of sets.

Dudley showed that the VC graph classes satisfy (4.13) in [6]. It is known that an ergodic Markov chain satisfies some mixing conditions [4]. Using empirical process CLT's for stationary sequences satisfying some mixing conditions one can also obtain a uniform CLT over the VC graph class. We will present examples such that our conditions for Markov chains are less restrictive than those required for a mixing process application to these problems.

Arcones and Yu [1] obtained an empirical process uniform CLT over VC graph classes $\mathcal{F}$ under the condition that the envelope function of $\mathcal{F}$ is in $L^p$ for some $2 < p < \infty$ and the $\beta$-mixing coefficients satisfy

$$k^{p/(p-2)}(\log k)^{2(p-1)/(p-2)}\beta_k \to 0 \quad \text{as } k \to \infty.$$

For uniformly bounded classes they need that $\beta_k = O(k^{-\gamma})$ for some $\gamma > 1$, where the mixing coefficients $\beta_k$ are defined by

$$\beta_k = \tfrac{1}{2}\sup\{\sum_{i=1}^{I}\sum_{j=1}^{J}|P\left(A_i \cap B_j\right) - P(A_i)P(B_j)|\,; \{A_i\}_{i=1}^{I}\text{is a partition of the sample space in } \sigma_1^l, \{B_j\}_{j=1}^{J} \text{ is a partition of the sample space in } \sigma_{l+k}^{\infty}, l \geq 1\}.$$

The following example is a Markov chain such that the envelope function $F$ satisfies our condition for the uniform CLT, but $F$ is only in $L^2$.

**Example 4.10.** Let $\{X_i\}$ be a stationary Markov chain with transition probability

$$P_{k^2,k^2+1} = \frac{k^a}{4^{2k+1}\left(k+1\right)^a}, P_{k^2,1} = 1 - P_{k^2,k^2+1}$$

and $P_{n,n+1} = 1$ if $n \neq k^2$ for some positive integer $k$. Let $N_i$ be the $i$-th hitting time of state 1. Let $F$ be a function such that

$$F(j) = 2^{k^2} \quad \text{for } (k-1)^2 < j \leq k^2.$$

If $a > 3$, then we have $E(N_2 - N_1)^2 < \infty$ and

$$E(\sum_{N_1 < j \leq N_2} F(X_j))^2 \approx \sum_{k=1}^{\infty} \frac{(2k-1)^2}{k^a} < \infty.$$

Hence the remark following Theorem 4.3 implies the uniform CLT for this Markov chain provided $\mathcal{F}$ is a permissible VC graph classes with envelope function $F$ i.e. VC graph class satisfy condition (4.1) by Dudley's work. However,

$$E[F(X_1)]^{2+\gamma} \approx \sum_{k=1}^{\infty} 2^{\gamma k^2}\frac{(2k-1)}{k^a} = \infty$$

for any $\gamma > 0$, and hence the envelope function $F \notin L^p(P)$ for all $p > 2$, and the results from [1] fails.

For uniformly bounded classes of functions our result only needs $E\,(N_2 - N_1)^2 < \infty$. The following example show that we may have $E\,(N_2 - N_1)^2 < \infty$, but $\beta_k \neq O(k^{-\gamma})$ for any $\gamma > 1$.

**Example 4.11**. Let $\{X_i\}$ be a stationary Markov chain with transition probability

$$P_{n,n+1} = (\frac{nLn}{(n+1)L(n+1)})^2, P_{n,1} = 1 - P_{n,n+1} \text{ for all } n \geq 1,$$

where $Ln = \ln(\max\{e, n\})$. Let $N_i$ be the $i$-th hitting time of state 1. Then

$$\begin{aligned} P\,(N_2 - N_1 = n) &= (\tfrac{1L1}{2L2})^2(\tfrac{2L2}{3L3})^2 \cdots (\tfrac{(n-1)L(n-1)}{nLn})^2(1 - (\tfrac{nLn}{(n+1)L(n+1)})^2) \\ &= (nLn)^{-2}(1 - (\tfrac{nLn}{(n+1)L(n+1)})^2) \approx n^{-3}(Ln)^{-2}, \end{aligned}$$

since $(1 - (\frac{nLn}{(n+1)L(n+1)})^2) = O(n^{-1})$. Thus

$$E\,(N_2 - N_1)^2 = \sum_{n=1}^{\infty} n^2 P\,(N_2 - N_1 = n) \approx \sum_{n=1}^{\infty} n^{-1}(Ln)^{-2} < \infty.$$

Recall that mixing coefficients $\beta_k$ is defined by

$$\beta_k = \tfrac{1}{2}\sup\{\sum_{i=1}^{I}\sum_{j=1}^{J} |P\,(A_i \cap B_j) - P(A_i)P(B_j)|\,; \{A_i\}_{i=1}^{I} \text{is a partition of the sample space in } \sigma_1^l, \{B_j\}_{j=1}^{J} \text{ is a partition of the sample space in } \sigma_{l+k}^{\infty}, l \geq 1\}.$$

We take

$$l = 1, \quad A_1 = \{X_1 = 1\}, \quad A_2 = A_1^c,$$
$$B_j = \{X_{k+1} = k + 1 + j\} \quad \text{for } j = 1, 2, \cdots, J \quad \text{and} \quad B_{J+1} = (\bigcup_{j=1}^{J} B_j)^c.$$

Note that $P(A_1 \cap B_j) = 0$ for $j = 1, 2, \cdots, J$. Thus

$$\beta_k \geq \frac{1}{2}\sup_J \sum_{j=1}^{J} P(A_1)P(B_j) \approx \sum_{j=1}^{\infty}(\frac{1}{(k+1+j)L(k+1+j)})^2 \approx \frac{1}{k(\log k)^2}.$$

# Bibliography


[1] Arcones, M. A. and Yu, B. Central limit theorems for empirical and U-processes of stationary mixing sequences. Journal of Theoretical Probability. Vol. 7 No.1, 1994, 47-71.

[2] Chen, X. Limit theorems for functionals of ergodic Markov chains with general state space. Ph.D. Thesis, 1997, Case Western Reserve University.

[3] Chung, K. L. Markov chain with stationary probability, Springer, 1967.

[4] Davydov, Yu. A. Mixing conditions for Markov chains. Theory of probability and its Application 28, 1973, 313-318.

[5] Dellacherie, C. and Meyer, P-A. Probabilities and potential. Part A. Amsterdam: North Hollard. 1978.

[6] Dudley, R. M. Universal Dunsker classes and metric entropy. Ann. Probab. 15, 1987, 1306-1326.

[7] Giné, E., Zinn, J. Lectures on the central limit theorem for empirical processes. Lect. Notes Math. Val. 1221, 1986, 50-113.

[8] Levental, S. Uniform limit theorems. Ph.D. Thesis, 1986, UW-Madison.

[9] Levental, S. Uniform limit theorems for Harris recurrent Markov chains. Probab. Th. Rel. Fields, 80, 1988, 101-118.

[10] Meyn, S.P. and Tweedie, R.L. Markov chains and stochastic stability. Springer-Verlag, 1993.

[11] Nummelin, E. General irreducible Markov chains and non-negetive operators. London, New York, 1984.

[12] Pollard, D. A central limit theorem for empirical processes. J. Austral. Math. Soc. (Series A) 33, 1982, 235-248.

[13] Pollard, D. Convergence of stochastic processes, Springer 1984.