A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains

Witold Bednorz †
Krzysztof Latuszynski ‡
† Institute of Mathematics, Warsaw University. 00-927 Warsaw. Poland.
‡ Department of Mathematical Statistics. Institute of Econometrics. Warsaw School of Economics. 02-554 Warsaw. Poland.

Keywords: Markov Chain, CLT, Uniform Ergodicity, Regeneration.
AMS: 60J05

Abstract

Central limit theorems for functionals of general state space Markov chains are of crucial importance in sensible implementation of Markov chain Monte Carlo algorithms as well as of vital theoretical interest. Different approaches to proving this type of results under diverse assumptions led to a large variety of CTL versions. However due to the recent development of the regeneration theory of Markov chains, many classical CLTs can be reproved using this intuitive probabilistic approach, avoiding technicalities of original proofs. In this paper we provide a regeneration proof of a CLT for functionals of uniformly ergodic Markov chains, thus solve the open problem posed in [8]. Moreover we discuss the difference between one-step and multiple-step small set condition.

1. Introduction

Let \((X_n)_{n\geq 0}\) be a time homogeneous Markov chain on a measurable space \((X,\mathcal{B}(X))\) with initial distribution \(\pi_0\), transition kernel \(P\) and a unique stationary distribution \(\pi\). Let \(g\) be a real valued Borel function on \(X\) and define \(\bar{g}_n = \frac{1}{n} \sum_{i=0}^{n-1} g(X_i)\) and \(E_\pi g = \int_X g(x) \pi(dx)\). We say that a \(\sqrt{n}\)-CLT holds for \((X_n)_{n\geq 0}\) and \(g\), if

\[
\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} N(0, \sigma_g^2), \quad \text{as} \quad n \to \infty,
\]

where \(\sigma_g^2 := \text{var}_\pi g(X_0) + 2 \sum_{n=1}^\infty \text{cov}_\pi \{g(X_0), g(X_n)\} < \infty\). Central limit theorems of this type are crucial for assessing the quality of Markov chain Monte Carlo estimation (see e.g. [5]) and are also of independent theoretical interest. Thus a large body of work on CLTs for functionals of Markov chains exists and a variety of results have been established under different assumptions and with different approaches to proofs (see [4] for a review). We state two classical CLT versions for geometrically ergodic and uniformly ergodic Markov chains.

Let \(\|\mu_1(\cdot) - \mu_2(\cdot)\|_{tv} := 2 \sup_{A \in \mathcal{B}} |\mu_1(A) - \mu_2(A)|\) be the well known total variation distance between probability measures \(\mu_1\) and \(\mu_2\). We say that a Markov chain \((X_n)_{n\geq 0}\) with transition kernel \(P\) and stationary distribution \(\pi\) is geometrically ergodic, if \(\|P^n(x, \cdot) - \pi(\cdot)\|_{tv} \leq M(x)\rho^n\), for some \(\rho < 1\)
and $M(x) < \infty \Rightarrow \pi-$almost everywhere. We say it is uniformly ergodic, if $\|P^n(x, \cdot) - \pi(\cdot)\|_{tv} \leq M \rho^n$, for some $\rho < 1$ and $M < \infty$.

**Theorem 1.1.** If a Markov chain $(X_n)_{n \geq 0}$ with stationary distribution $\pi$ is geometrically ergodic and $\pi(|g|^{2+\delta}) < \infty$ for some $\delta > 0$, then a $\sqrt{n}-CLT$ holds for $(X_n)_{n \geq 0}$ and $g$.

**Theorem 1.2.** If a Markov chain $(X_n)_{n \geq 0}$ with stationary distribution $\pi$ is uniformly ergodic and $\pi(g^2) < \infty$, then a $\sqrt{n}-CLT$ holds for $(X_n)_{n \geq 0}$ and $g$.

Theorem 1.1 due to [3] has been reproved in [8] using the intuitive regeneration approach and avoiding technicalities of the original proof (however see our Section 4). Roberts and Rosenthal posed an open problem, whether Theorem 1.2 due to [2] can also be reproved using direct regeneration arguments.

The aim of this paper is to provide a regeneration proof of Theorem 1.2. The outline of the paper is as follows. In Section 2 we describe the regeneration construction. In Section 3 we prove Theorem 1.2 and we discuss some of the difficulties of the regeneration approach in Section 4.

## 2. Small Sets and the Split Chain

The regeneration construction discovered independently by [7] and [1] is now a well established technique. A systematic development of the theory can be found in e.g. [6] which we exploit in this section.

**Definition 2.1 (Small Set).** A set $C \in \mathcal{B}(X)$ is $\nu_m-$small, if there exist $m > 0, \varepsilon > 0$, and a nontrivial probability measure $\nu_m$ on $\mathcal{B}(X)$, such that for all $x \in C$,

$$P^m(x, \cdot) \geq \varepsilon \nu_m(\cdot).$$

Since ergodic Markov chains are $\pi-$irreducible, Theorem 5.2.2 of [6] implies that for an ergodic chain a small set $C$ with $\pi(C) > 0$ always exists.

A small set $C$ with $\pi(C) > 0$ allows for constructing the split chain for $(X_n)_{n \geq 0}$ which is the central object of the approach (see Section 17.3 of [6] for a detailed description). Let $(X_{nm})_{n \geq 0}$ be the $m-$skeleton of $(X_n)_{n \geq 0}$, i.e. a Markov chain evolving according to the $m-$step transition kernel $P^m$. The small set condition allows to write $P^m$ as a mixture of two distributions:

$$P^m(x, \cdot) = \varepsilon \mathbb{I}_C(x) \nu_m(\cdot) + (1 - \varepsilon \mathbb{I}_C(x)) R(x, \cdot),$$

where $R(x, \cdot) = [1 - \varepsilon \mathbb{I}_C(x)]^{-1} [P(x, \cdot) - \varepsilon \mathbb{I}_C(x) \nu_m(\cdot)]$. Now let $(Y_{nm}, Y_n)_{n \geq 0}$ be the split chain of the $m-$skeleton i.e. let the random variable $Y_n \in \{0, 1\}$ be the level of the split $m-$skeleton at time $nm$. The split chain $(X_{nm}, Y_n)_{n \geq 0}$ is a Markov chain that obeys the following transition rule $\bar{P}$.

$$\bar{P}(Y_n = 1, X_{(n+1)m} \in dy | Y_{nm}, X_{nm} = x) = \varepsilon \mathbb{I}_C(x) \nu_m(dy)$$

$$\bar{P}(Y_n = 0, X_{(n+1)m} \in dy | Y_{nm}, X_{nm} = x) = (1 - \varepsilon \mathbb{I}_C(x)) R(x, dy).$$
and $Y_n$ can be interpreted as a coin toss indicating whether $X_{(n+1)m}$ given
$X_{nm} = x$ should be drawn from $\nu_m(\cdot)$ - with probability $\varepsilon \Pi_C(x)$ - or from
$R(x, \cdot)$ with probability $1 - \varepsilon \Pi_C(x)$.
One obtains the split chain $(X_k, Y_n)_{k \geq 0, n \geq 0}$ of the initial Markov chain
$(X_n)_{n \geq 0}$ by defining appropriate conditional probabilities. To this end let
$X^m_{0} = \{X_0, \ldots, X_{nm-1}\}$ and $Y^m_0 = \{Y_0, \ldots, Y_{n-1}\}$.

\[ \hat{P}(Y_n = 1, X_{nm+1} \in dx_1, \ldots, X_{(n+1)m-1} \in dx_{m-1}, X_{(n+1)m} \in dy) \]

\[ |Y^m_0, X_{0}^{nm}, X_{nm} = x) = \frac{\varepsilon \Pi_C(x) \nu_m(dy)}{P^{m}_{c}(x, dy)} P(x, dx_1) \cdots P(x_{m-1}, dy), \]

\[ \hat{P}(Y_n = 0, X_{nm+1} \in dx_1, \ldots, X_{(n+1)m-1} \in dx_{m-1}, X_{(n+1)m} \in dy) \]

\[ |Y^m_0, X_{0}^{nm}, X_{nm} = x) = \frac{(1 - \varepsilon \Pi_C(x)) R(x, dy)}{P^{m}_{c}(x, dy)} P(x, dx_1) \cdots P(x_{m-1}, dy). \]

Note that the marginal distribution of $(X_k)_{k \geq 0}$ in the split chain is that of the
underlying Markov chain with transition kernel $P$.

For a measure $\lambda$ on $(\mathcal{X}, B(\mathcal{X}))$ let $\lambda^*$ denote the measure on $\mathcal{X} \times \{0, 1\}$
(with product $\sigma-$algebra) defined by $\lambda^*(B \times \{1\}) = \varepsilon \lambda(B)$ and $\lambda^*(B \times \{0\}) = (1 - \varepsilon)B$. Now the crucial observation is that on the set $\{Y_n = 1\}$, the
pre--nm process $\{X_k, Y_i : k \leq nm, i \leq n\}$ and the post--$(n+1)m$ process
$\{X_k, Y_i : k \geq (n+1)m, i \geq n+1\}$ are independent and the post--$(n+1)m$
process has the same distribution with $\nu_m^*$ for the initial distribution of $(X_n, Y_0)$. This leads to Theorem 2.2, but we first need
some more notation. Thus let $\sigma(n)$ denote entrance times of the split chain to
the set $C \times \{1\}$, i.e. $\sigma(0) = \min\{k \geq 0 : Y_k = 1\}$, and $\sigma(n) = \min\{k > \sigma(n-1) : Y_k = 1\}$, for $n \geq 1$. Also define $Z_n(g) = \sum_{k=0}^{n-1} g(X_{nm+k})$ and $g_e = g - \pi g$.

**Theorem 2.2 (Theorem 17.3.6 of [6]).** Suppose that $(X_n)_{n \geq 0}$ is ergodic
and let $\nu_m$ be the measure satisfying (2). If the following conditions hold

\[ (i) \quad \hat{E}_{\nu_m} \left[ \left( \sum_{n=0}^{\sigma(0)} Z_n(|g|) \right)^2 \right] < \infty, \quad (ii) \quad \hat{E}_{\nu_m} [\sigma(0)^2] < \infty, \]

then the $\sqrt{n}-$CLT holds for $(X_n)_{n \geq 0}$ and $g$, with

\[ \sigma^2_g = \frac{\varepsilon \pi(C)}{m} \left\{ \hat{E}_{\nu_m} \left[ \left( \sum_{n=0}^{\sigma(0)} Z_n(g_e) \right)^2 \right] + 2\hat{E}_{\nu_m} \left[ \left( \sum_{n=0}^{\sigma(0)} Z_n(g) \right) \left( \sum_{n=\sigma(0)+1}^{\sigma(1)} Z_n(g_e) \right) \right] \right\}. \]

### 3. A Proof

In view of Theorem 2.2 providing a regeneration proof of Theorem 1.2 amounts to establishing conditions (i) and (ii) of (8). To this end we need some
additional facts about small sets for uniformly ergodic Markov chains.
Theorem 3.1. If \((X_n)_{n \geq 0}\) is a Markov chain on \((X, \mathcal{B}(X))\) with stationary distribution \(\pi\) is uniformly ergodic, then \(X\) is \(\nu_m\)-small for some \(\nu_m\).

Hence for uniformly ergodic chains (2) holds for all \(x \in X\). Theorem 3.1 is well known in literature, in particular it results from Theorems 5.2.1 and 5.2.4 in [6] with their \(\psi = \pi\).

We start with proving (ii) of (8) which is now straightforward. Integrating (6) together with the fact that \(X\) is small, yields \(\hat{P}(Y_n = 1|X_0^{nm}, Y_0^{n-1}, X_{nm} = x) = \varepsilon\), thus \(Y_0, Y_1, \ldots\) are independent Bernoulli trials and the distribution of \(\sigma(0)\) is geometric.

Establishing (i) of (8) is the essential part of the proof. Theorem 3.1 implies that for uniformly ergodic Markov chains (3) can be rewritten in operator notation as

\[ P^m = \varepsilon \nu_m + (1 - \varepsilon) R. \]  

The following mixture representation of \(\pi\) will turn out very useful.

Lemma 3.2. If \((X_n)_{n \geq 0}\) is an ergodic Markov chain with transition kernel \(P\) and (9) holds, then

\[ \pi = \varepsilon \mu := \varepsilon \sum_{n=0}^{\infty} \nu_m(1 - \varepsilon)^n R^n. \]  

Proof. Since \(\varepsilon \sum_{n=0}^{\infty} \nu_m(1 - \varepsilon)^n R^n)(X) = \varepsilon \sum_{n=0}^{\infty}(1 - \varepsilon)^n(\nu_m R^n)(X) = 1\), the measure in question is a probability measure. It is also invariant for \(P^m\). By (9) we obtain

\[ \left( \sum_{n=0}^{\infty} \nu_m(1 - \varepsilon)^n R^n \right) P^m \varepsilon \nu_m + \sum_{n=1}^{\infty} \nu_m(1 - \varepsilon)^n R^n = \sum_{n=0}^{\infty} \nu_m(1 - \varepsilon)^n R^n. \]

Hence by ergodicity \(\varepsilon \mu = \varepsilon \mu P^m \to \pi\), as \(n \to \infty\). Thus \(\varepsilon \mu = \pi\).

Corollary 3.3. The decomposition in Lemma 3.2 implies that

(i) \( \hat{E}_{\nu_m} \left( \sum_{n=0}^{\sigma(0)} \mathbb{1}_{\{X_{nm} \in A\}} \right) = \hat{E}_{\nu_m} \left( \sum_{n=0}^{\infty} \mathbb{1}_{\{X_{nm} \in A\}} \mathbb{1}_{\{Y_0 = 0, \ldots, Y_{n-1} = 0\}} \right) = \varepsilon^{-1} \pi(A) \)

(ii) \( \hat{E}_{\nu_m} \left( \sum_{n=0}^{\infty} f(X_{nm}, X_{nm+1}, \ldots; Y_n, Y_{n+1}, \ldots) \mathbb{1}_{\{Y_0 = 0, \ldots, Y_{n-1} = 0\}} \right) = \varepsilon^{-1} \hat{E}_{\pi^*} f(X_0, X_1, \ldots; Y_0, Y_1, \ldots) \).

Proof. (i) is a direct consequence of (10). To see (ii) note that \(Y_n\) is a coin toss independent of \(\{Y_0, \ldots, Y_{n-1}\}\) and \(X_{nm}\), this allows for \(\pi^*\) instead of \(\pi\) on the RHS of (ii). Moreover the evolution of \(\{X_{nm+1}, X_{nm+2}, \ldots; Y_{n+1}, Y_{n+2}, \ldots\}\) depends only (and explicitly by (6) and (7)) on \(X_{nm}\) and \(Y_n\). Now use (i).
Our object of interest

\[ I = \mathcal{E}_n \left[ \left( \sum_{n=0}^{\sigma(0)} Z_n(|g|) \right)^2 \right] = \mathcal{E}_n \left[ \left( \sum_{n=0}^{\infty} Z_n(|g|)I\{\sigma(0) \geq n\} \right)^2 \right] \]

\[ = \mathcal{E}_n \left[ \sum_{n=0}^{\infty} Z_n(|g|)^2 I\{Y_0=0,Y_1=0,\ldots,Y_{n-1}=0\} \right] + 2\mathcal{E}_n \left[ \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} Z_n(|g|)I\{\sigma(0) \geq n\} Z_k(|g|)I\{\sigma(0) \geq k\} \right] = A + B \quad (11) \]

Now we can use Corollary 3.3 and then the inequality \(2ab \leq a^2 + b^2\) to bound the term \(A\) in (11).

\[ A = \frac{1}{\varepsilon} \mathcal{E}_n Z_0(|g|)^2 = \frac{1}{\varepsilon} \mathcal{E}_n \left( \sum_{k=0}^{m-1} |g(X_k)| \right)^2 \leq \frac{m}{\varepsilon} \mathcal{E}_n \left( \sum_{k=0}^{m-1} g^2(X_k) \right) \leq \frac{m^2}{\varepsilon} \pi g^2 < \infty. \]

We can similarly proceed with term \(B\).

\[ B = 2\mathcal{E}_n \left[ \sum_{n=0}^{\infty} Z_n(|g|)I\{\sigma(0) \geq n\} \sum_{k=1}^{\infty} Z_0(|g|)I\{\sigma(0) \geq n+k\} \right] \]

\[ = \frac{2}{\varepsilon} \mathcal{E}_n \left[ \sum_{k=1}^{\infty} Z_0(|g|) \sum_{k=1}^{\infty} Z_k(|g|)I\{\sigma(0) \geq k\} \right] = \frac{2}{\varepsilon} \sum_{k=1}^{\infty} \mathcal{E}_n \left[ I\{\sigma(0) \geq k\} Z_0(|g|)Z_k(|g|) \right] \quad (12) \]

Let \(C_k := \mathcal{E}_n \left[ I\{\sigma(0) \geq k\} Z_0(|g|)Z_k(|g|) \right].\) By Cauchy-Schwarz,

\[ C_k \leq \sqrt{\mathcal{E}_n \left[ I\{\sigma(0) \geq k\} Z_0(|g|)^2 \right]} \sqrt{\mathcal{E}_n \left[ Z_k(|g|)^2 \right]} \]

\[ = \sqrt{\mathcal{E}_n \left[ I\{Y_0=0,Y_1=0,\ldots,Y_{k-1}=0\} Z_0(|g|)^2 \right]} \sqrt{\mathcal{E}_n \left[ Z_0(|g|)^2 \right]}. \]

Now observe that \(\{Y_1,\ldots,Y_{k-1}\}\) and \(\{X_0,\ldots,X_{m-1}\}\) are independent. Moreover we drop \(I\{Y_0=0\}\) to obtain

\[ C_k \leq (1-\varepsilon)^{k+1} \mathcal{E}_n Z_0(|g|)^2 \leq (1-\varepsilon)^{k+1} \frac{m^2}{\varepsilon} \pi g^2 \quad (13) \]

Combining (12) and (13) yields that \(B < \infty.\) This completes the proof.

**4. The difference between \(m = 1\) and \(m \neq 1\)**

Assume the small set condition (2) holds and consider the split chain defined by (6) and (7). The following tours

\[ \{X_{\sigma(n)+1}m, X_{\sigma(n)+1}m+1, \ldots, X_{\sigma(n)+1}m+n-1, n = 0, 1, \ldots \} \]
that start whenever \( X_k \sim \nu_m \) are of crucial importance to the regeneration theory and are eagerly analyzed by researchers. In virtually every paper on the subject there is a claim these objects are independent identically distributed random variables. This claim is usually considered obvious and no proof is provided. However this is not true if \( m > 1 \).

In fact formulas (6) and (7) should be convincing enough, as \( X_{mn+1}, \ldots, X_{(n+1)m} \) given \( Y_n = 1 \) and \( X_{nm} = x \) are linked in a way described by \( P(x, dx_1) \cdots P(x_{m-1}, dy) \). In particular consider a Markov chain on \( X = \{a, b, c, d, e\} \) with transition probabilities \( P(a, b) = P(a, c) = P(b, b) = P(b, d) = P(c, c) = P(c, e) = 1/2 \), and \( P(d, a) = P(e, a) = 1 \). Let \( \nu_4(d) = \nu_4(e) = 1/2 \) and \( \varepsilon = 1/8 \). Clearly \( P^4(x, \cdot) \geq \varepsilon \nu_4(\cdot) \) for every \( x \in X \), hence we established (2) with \( C = X \). Note that for this simplistic example each tour can start with \( d \) or \( e \). However if it starts with \( d \) or \( e \) the previous tour must have ended with \( b \) or \( c \) respectively. This makes them dependent!

Similar examples with general state space \( X \) and \( C \neq X \) can be easily provided. Hence Theorem 2.2 is critical to providing regeneration proofs of CLTs and standard arguments that involve iid random variables are not valid.

5. Bibliography


