A regeneration proof of the central limit theorem for uniformly ergodic Markov chains

by

Ajay Jasra
Department of Mathematics
Imperial College London

and

Chao Yang[1]
Department of Mathematics
University of Toronto

Technical Report No. 0306 March 14, 2006

TECHNICAL REPORT SERIES

University of Toronto

Department of Statistics

A regeneration proof of the central limit theorem for uniformly ergodic Markov chains

By

AJAY JASRA

Department of Mathematics, Imperial College London, SW7 2AZ, London, UK

and

CHAO YANG[1]

Department of Mathematics, University of Toronto, M5S 2E4, Toronto ON, Canada

Abstract

Let (X_n) be a Markov chain on measurable space $(E, \mathcal{E})$ with unique stationary distribution π. Let $h : E \to \mathbb{R}$ be a measurable function with finite stationary mean $\pi(h) := \int_E h(x)\pi(dx)$. Ibragimov & Linnik (1971) proved that if (X_n) is geometrically ergodic, then a central limit theorem (CLT) holds for h whenever $\pi(|h|^{2+\delta}) < \infty$, $\delta > 0$. Cogburn (1972) proved that if a Markov chain is uniformly ergodic, with $\pi(h^2) < \infty$ then a CLT holds for h. The first result was re-proved in Roberts & Rosenthal (2004) using a regeneration approach; thus removing many of the technicalities of the original proof. This raised an open problem: to provide a proof of the second result using a regeneration approach. In this paper we provide a solution to this problem.

Keywords: Markov chains; Central limit theorems

1 Introduction

Let (X_n) be a Markov chain with transition kernel $P : E \times \mathcal{E} \to [0, 1]$ and a unique stationary distribution π. Let $h : E \to \mathbb{R}$ be a real-valued measurable function. We say that h satisfies a Central Limit Theorem (or $\sqrt{n}$−CLT) if there is some $\sigma^2 < \infty$ such that the normalized sum $n^{-\frac{1}{2}} \sum_{i=1}^{n} [h(X_i) - \pi(h)]$ converges weakly to a $N(0, \sigma^2)$ distribution, where $N(0, \sigma^2)$ is a Gaussian distribution with zero mean and variance σ^2 (we allow that $\sigma^2 = 0$), and (e.g. Chan & Geyer (1994), see also Bradley (1985) and Chen (1999))

$$\sigma^2 = \pi(h^2) + 2 \int_E \sum_{n=1}^{\infty} h(x) P^n(h)(x) \pi(dx)$$

with $P^n(h)(x) = \int_E h(y)P^n(x, dy)$ and $P^n(x, dy)$ the n−step transition law for the Markov chain.

To further our discussion we provide the following definitions. Denote the class of probability measures on $(E, \mathcal{E})$ as $\mathcal{P}(E)$. We define the total variation distance between $\mu, \nu \in \mathcal{P}(E)$ as:

$$||\mu - \nu|| := \sup_{A \in \mathcal{E}} |\mu(A) - \nu(A)|.$$

We will be concerned with geometrically and uniformly ergodic Markov chains:

Definition 1.1. *A Markov chain with stationary distribution $\pi \in \mathcal{P}(E)$ is geometrically ergodic if $\forall n \in \mathbb{N}$:*

$$||P^n(x, \cdot) - \pi(\cdot)|| \leqslant M(x)\rho^n$$

[1]Corresponding author, e-mail: chaoyang@math.toronto.edu

where $\rho < 1$ and $M(x) < \infty$ $\pi-$almost everywhere. If $M = \sup_{x\in E} |M(x)|$ is finite then the chain is uniformly ergodic.

Theorem 1.2 (Cogburn, 1972). *If a Markov chain with stationary distribution $\pi \in \mathcal{P}(E)$ is uniformly ergodic, then a $\sqrt{n}-$CLT holds for h whenever $\pi(h^2) < \infty$.*

Ibragimov & Linnik (1971) proved a CLT for h when the chain is geometrically ergodic and $\pi(|h|^{2+\delta}) < \infty$ for some $\delta > 0$, and Roberts & Rosenthal (2004) provided a simpler proof using regeneration arguments. Roberts & Rosenthal (2004) left an open problem: To provide a proof of Theorem 1.2 (originally proved by Cogburn (1972)) using regeneration.

Many of the recent developments of CLTs for Markov chains are related to the evolution of stochastic simulation algorithms such as Markov chain Monte Carlo (MCMC) (e.g. Robert & Rosenthal (2004)). For example, Roberts & Rosenthal (2004) posed many open problems, including that considered here, for CLTs; see Häggström (2005) for a solution to another open problem. Additionally, Jones (2004) discusses the link between mixing processes and CLTs, with MCMC algorithms a particular consideration. For an up-to-date review of CLTs for Markov chains see: Bradley (1985), Chen (1999) and Jones (2004).

The proof of Theorem 1.2, using regeneration theory, provides an elegant framework for the proof of CLTs for Markov chains. The approach may also be useful for alternative proofs of CLTs for chains with different ergodicity properties; e.g. $V-$uniform ergodicity (see Meyn & Tweedie (1993)).

The structure of this paper is as follows. In Section 2 we provide a new proof (to our knowledge) that E is small if the Markov chain is uniformly ergodic. We also detail the regeneration construction. In Section 3 we give some technical results that are used to prove Theorem 1.2. In Section 4 we use the results of the previous Section to provide a proof of Theorem 1.2 using regenerations.

2 Small Sets and Regeneration Construction

2.1 Small Sets

We recall the notion of a small set:

Definition 2.1. *A set $C \in \mathcal{E}$ is small (or (n_0, ϵ, ν)-small) if there exists an $n_0 \in \mathbb{N}$, $\epsilon > 0$ and a non-trivial $\nu \in \mathcal{P}(E)$ such that the following minorization condition holds $\forall x \in C$:*

$$P^{n_0}(x, \cdot) \geqslant \epsilon\nu(\cdot). \tag{1}$$

It is well-known (e.g. Meyn & Tweedie (1993)) that if P is uniformly ergodic, the state space E is small. Since this result is crucial to what follows, we begin this Section with a new proof of this conclusion.

Lemma 2.1. *If (X_n) on $(E, \mathcal{E})$ with stationary distribution $\pi \in \mathcal{P}(E)$ is uniformly ergodic, then E is small.*

Proof. Fixed $x_0 \in E$, for $\forall x \in E$ and using Definition 1.1, we have (via the triangle inequality):

$$||P^n(x, \cdot) - P^n(x_0, \cdot)|| \leqslant 2M\rho^n. \tag{2}$$

Now we construct joint random variables Z and Y_{x_0}, such that $Z \sim \pi(\cdot)$, $Y_{x_0} \sim P^n(x, \cdot)$. By Proposition 3(g) of Roberts & Rosenthal (2004), we may obtain:

$$\begin{aligned}\mathbb{P}(Z = Y_{x_0}) &= 1 - ||P^n(x_0, \cdot) - \pi(\cdot)|| \\ &\geqslant 1 - M\rho^n.\end{aligned}$$

As a result, for $A \in \mathcal{E}$

$$\begin{aligned}P^n(x_0, A) &= \mathbb{P}(Y_{x_0} \in A) \\ &= \mathbb{P}(Y_{x_0} \in A|Z = Y_{x_0})\mathbb{P}(z = Y_{x_0}) + \mathbb{P}(Y_{x_0} \in A|Z \neq Y_{x_0}) \\ &\geqslant \mathbb{P}(Y_{x_0} \in A|Z = Y_{x_0})\mathbb{P}(z = Y_{x_0}) \\ &\geqslant (1 - M\rho^n)\mathbb{P}(Y_{x_0} \in A|Z = Y_{x_0}). \end{aligned} \tag{3}$$

Therefore by (2) and (3), $\forall x \in E$ and $A \in \mathcal{E}$,

$$P^n(x, A) \geqslant (1 - M\rho^n)\mathbb{P}(Y_{x_0} \in A|Z = Y_{x_0}) - 2M\rho^n$$

.

Now we construct the n_0 and $\epsilon > 0$ as in Definition 2.1. Firstly, let n_0 be large enough such that:

$$1 - 3M\rho^{n_0} > 0.$$

Secondly, define a probability measure:

$$\nu(A) = \begin{cases} \frac{(1-M\rho^{n_0})\widetilde{\nu}(A) - 2M\rho^{n_0}}{1-3M\rho^{n_0}} & \text{if } \widetilde{\nu}(A) > \frac{2M\rho^{n_0}}{1-M\rho^{n_0}} \\ 0 & \text{otherwise} \end{cases}$$

where $\widetilde{\nu}(A) = \mathbb{P}(Y_{x_0} \in A|Z = Y_{x_0})$. Then let $\epsilon = 1 - 3M\rho^{n_0}$, clearly (1) holds and E is a small. □

2.2 Regeneration Construction

Now we consider the regeneration construction for the proof. Since E is small we use the split chain construction (Nummelin, 1984):

$$P^{n_0}(x, A) = (1 - \epsilon)R(x, A) + \epsilon\nu(A)$$

$\forall x \in E$ and $A \in \mathcal{E}$ where $R(x, A) = (1 - \epsilon)^{-1}[P^{n_0}(x, A) - \epsilon\nu(A)]$. That is, for a single chain X_n, with probability ϵ we choose $X_{n+n_0} \sim \nu$, while with probability $1-\epsilon$ we choose $X_{n+n_0} \sim R(X_n, \cdot)$, if $n_0 > 1$, we fill in the missing values as X_{n+1} using the appropriate Markov kernel and conditionals.

We let $T_1, T_2, \ldots$ be the regeneration times, i.e. the times such that $X_{T_i} \sim \nu$, clearly $T_i = in_0$, so we have the property:

$$\begin{aligned}\pi(dx_{n_0}) &= \int_E \pi(dx_0)P^{n_0}(x_0, dx_{n_0}) \\ &\geqslant \epsilon\nu(dx_{n_0})\end{aligned} \tag{4}$$

and

$$p(dt|x) \quad = \quad \epsilon(1-\epsilon)^{\frac{t}{n_0}-1}c(dt)$$

on measurable space $(S, \mathcal{S})$, $S = \{n_0, 2n_0, \dots\}$, where $\mathcal{S}$ is the associated $\sigma-$algebra of all subsets of S and c is counting measure.

Let $T_0 = 0$ and $r(n) = \sup\{i \geqslant 0 : T_i \leqslant n\}$, using the regeneration time, we can break up the sum $\sum_{i=0}^{n}[h(X_i) - \pi(h)]$ into sums over tours as follows:

$$\sum_{i=0}^{n}[h(X_i) - \pi(h)] = \sum_{j=1}^{r(n)} \sum_{i=T_j}^{T_{j+1}-1} [h(X_i) - \pi(h)] + Q(n)$$

where

$$Q(n) = \sum_{j=0}^{T_1-1}[h(X_j) - \pi(h)] + \sum_{T_{r(n)+1}}^{n} [h(X_j) - \pi(h)].$$

3 Some Technical Results

Given the formulation provided in the previous Section, we now give some results that will assist us in proving Theorem 1.2.

Lemma 3.1. *Under the formulation above, we have that:*

$$\frac{Q(n)}{n^{1/2}} \longrightarrow_p 0. \tag{5}$$

Proof. Let

$$\begin{aligned} Q_1^+(n) &= \sum_{j=0}^{T_1-1}[h(X_j) - \pi(h)]^+ \\ Q_1^-(n) &= \sum_{j=0}^{T_1-1}[h(X_j) - \pi(h)]^- \end{aligned}$$

and

$$\begin{aligned} Q_2^+(n) &= \sum_{T_{r(n)}+1}^{n} [h(X_j) - \pi(h)]^+ \\ Q_2^-(n) &= \sum_{T_{r(n)}+1}^{n} [h(X_j) - \pi(h)]^- \end{aligned}$$

where $[h(X_j) - \pi(h)]^+ = \max\{h(X_j) - \pi(h), 0\}$ and $[h(X_j) - \pi(h)]^- = \max\{-[h(X_j) - \pi(h)], 0\}$.

The strategy of the proof is to show that $Q_i^{\pm}(n)/n^{1/2} \rightarrow_p 0$ as $n \rightarrow \infty$. Consider $Q_1^+(n)$,

$$Q_1^+(n) = \sum_{j=0}^{sn_0-1} [h(X_j) - \pi(h)]^+ \quad \text{w.p} \quad \epsilon(1-\epsilon)^{(s-1)} \tag{6}$$

where $s \in \mathbb{N}$. If $Q_1^+(n)/n^{1/2} \to_p 0$, i.e. $\mathbb{P}(\exists\epsilon, Q_1^+(n) > \epsilon n^{1/2}, \text{i.o.}) = 1$ for all n, which means that $\mathbb{P}(Q_1^+(n) = \infty, \text{i.o.}) = 1$, which is impossible from (6). So $Q_i^+(n)/n^{1/2} \to_p 0$ as $n \to \infty$. Similarly $Q_i^-(n)/n^{1/2} \to_p 0$ as $n \to \infty$.

For Q_2 we have $Q_2^+(n) \leqslant \sum_{j=r_n+1}^{l_n} [h(X_j) - \pi(h)]^+ = \widetilde{Q}_2^+(n)$, where $l(n) = \inf\{i \geqslant 0 : T_i \geqslant n\}$. We know that $\widetilde{Q}_2^+(n)$ has the same distribution with $Q_2^+(n)$, so $\widetilde{Q}_i^+(n)/n^{1/2} \to_p 0$ as $n \to \infty$ and therefore, $Q_2^+(n)/n^{1/2} \to_p 0$ as $n \to \infty$. Similarly $Q_2^-(n)/n^{1/2} \to_p 0$ as $n \to \infty$. From the above discussion, we conclude that $Q(n)/n^{1/2} \to_p 0$. □

Lemma 3.2. *Under the formulation above we have:*

$$I = \mathbb{E}\Big[\Big(\sum_{i=T_1}^{T_2-1} [h(X_i) - \pi(h)]\Big)\Big] = 0.$$

Proof. Let

$$I_1 = \mathbb{E}\Big[\Big(\sum_{i=0}^{T_1-1} [h(X_i) - \pi(h)]\Big)^-\Big]$$
$$I_2 = \mathbb{E}\Big[\Big(\sum_{i=0}^{T_2-1} [h(X_i) - \pi(h)]\Big)\Big].$$

Since $X_0 \sim \pi(\cdot)$, we have that

$$\begin{aligned} I_2 &= \int_E \pi(dx_0)\mathbb{E}\Big[\Big(\sum_{i=0}^{T_2-1} [h(X_i) - \pi(h)]\Big)\Big| X_0 = x\Big] \\ &= \sum_{s=2}^{\infty}\sum_{i=1}^{sn_0-1} E_i \cdot [\epsilon(1-\epsilon)^s] \end{aligned} \tag{7}$$

where

$$E_i = \int_E \pi(dx_0)\mathbb{E}\Big[\Big(h(X_i) - \pi(h)\Big)\Big| X_0 = x\Big].$$

It then follows that

$$\begin{aligned} E_i &= \int_E\int_E [h(x_i) - \pi(h)] P^i(x_0, dx_i)\pi(dx_0) \\ &= \int_E [h(x_i) - \pi(h)]\pi(dx_i) \\ &= 0 \end{aligned} \tag{8}$$

where (8) follows from the fact that the Markov kernel P^i is $\pi-$invariant $\pi(dx_i) = \int_E \pi(dx_0)P^i(x_0, dx_i)$. By (7), $I_2 = 0$, and similarly $I_1 = 0$; thus $I = I_2 - I_1 = 0$. □

Lemma 3.3. *Under the formulation above we have:*

$$J = \int_E \nu(dx)\mathbb{E}\Big[\Big(\sum_{i=0}^{T_1-1} [h(X_i) - \pi(h)]\Big)^2\Big| X_0 = x\Big] < \infty. \tag{9}$$

Proof. By definition:

$$
\begin{aligned}
J &= \int_E \nu(dx_0) \int_S \int_{E^{t_1-1}} \left(\sum_{i=0}^{t_1-1} [h(x_i) - \pi(h)] \right)^2 P(x_0, dx_1) \ldots P(x_{t_1-2}, dx_{t_1-1}) p(dt_1|x_0) \\
&\leqslant \frac{1}{\epsilon} \mathbb{E}_\pi \left[\left(\sum_{i=0}^{T_1-1} [h(x_i) - \pi(h)] \right)^2 \right]
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}_\pi \left[\left(\sum_{i=0}^{T_1-1} [h(x_i) - \pi(h)] \right)^2 \right] &= \int_E \pi(dx_0) \sum_S \int_{E^{t_1-1}} \left(\sum_{i=0}^{t_1-1} h(x_i) - \pi(h) \right)^2 \\
&\quad \times P(x_0, dx_1) \ldots P(x_{t_1-2}, dx_{t_1-1}) \epsilon (1-\epsilon)^{\frac{t_1}{n_0}-1}.
\end{aligned}
$$

and we have used (4).

Applying Cauchy-Schwartz we have:

$$
\mathbb{E}_\pi \left[\left(\sum_{i=0}^{T_1-1} [h(x_i) - \pi(h)] \right)^2 \right] \leqslant \left(\sum_{i=0}^{\infty} \mathbb{E}_\pi \left[\mathbb{I}(i < T_1) H_i^2 \right]^{1/2} \right)^2
$$

with $H_i = h(x_i) - \pi(h)$. For $i = 0$ we can write:

$$
\mathbb{E}_\pi \left[\mathbb{I}(0 < T_1) H_0^2 \right] = \int_E \pi(dx_0) \mathbb{P}(T_1 \geqslant 1 | x_0) H_0^2
$$

For $i \geqslant 1$ it can be seen that we may factorize out the expression

$$
\int_{E^i} P(x_0, dx_1) \ldots P(x_{i-1}, dx_i) H_i^2
$$

from the summation over t_1, thus we have:

$$
\mathbb{E}_\pi \left[\mathbb{I}(i < T_1) H_i^2 \right] = \int_{E^2} \pi(dx_0) P^i(x_0, dx_i) \mathbb{P}(T_1 \geqslant i+1 | x_0) H_i^2
$$

but:

$$
\mathbb{P}(T_1 \geqslant i+1 | x_0) = (1-\epsilon)^{m_i - 1}
$$

where $l_i = \min\{j \in \mathbb{N} : j n_0 \geqslant i+1\}$.

Since $\int_{E^2} H_i^2 \pi(dx_0) P^i(x_0, dx_i)$ is finite, independent of i, L. So we have:

$$
\begin{aligned}
J &\leqslant \frac{L}{\epsilon} \left(\sum_{i=0}^{\infty} (1-\epsilon)^{\frac{l_i - 1}{2}} \right)^2 \\
&\leqslant \frac{L}{\epsilon} \left(\sum_{i=0}^{\infty} (1-\epsilon)^{\frac{i+1}{2n_0} - \frac{1}{2}} \right)^2 \\
&= \frac{L}{\epsilon^3} (1-\epsilon)^{\frac{1}{n_0} - 1} < \infty.
\end{aligned}
$$

□

4 Proof of Theorem 1.2

Using the results in Section 3 we now provide a proof of Theorem 1.2.

Proof of Theorem 1.2. Following Lemma 3.1, we can obtain:

$$\lim_{n\to\infty}\frac{\sum_{i=0}^{n}[h(X_i)-\pi(h)]}{n^{1/2}} = \lim_{n\to\infty}\frac{\sum_{j=1}^{r(n)}\sum_{i=T_j}^{T_{j+1}-1}[h(X_i)-\pi(h)]}{n^{1/2}}. \tag{10}$$

Note that $\{\sum_{i=T_j}^{T_{j+1}-1}[h(X_i)-\pi(h)]\}_{j=1}^{\infty}$ are i.i.d. We denote

$$\begin{aligned}
\tilde{\sigma}^2 &= \mathbb{E}\left[\left(\sum_{i=T_1}^{T_2-1}[h(X_i)-\pi(h)]\right)^2\right] \\
\sigma^2 &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n}[h(X_i)-\pi(h)]\right)^2\right].
\end{aligned}$$

Following (10), we have:

$$\begin{aligned}
\sigma^2 &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\left(\sum_{i=1}^{n}[h(X_i)-\pi(h)]\right)^2\right] \\
&= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{r(n)}\sum_{i=T_j}^{T_{j+1}-1}[h(X_i)-\pi(h)]^2\right] \quad \text{by } (10) \\
&= \lim_{n\to\infty}\frac{r_n}{n}\mathbb{E}\left[\left(\sum_{i=T_1}^{T_2-1}[h(X_i)-\pi(h)]\right)^2\right] \\
&= \tilde{\sigma}^2\lim_{n\to\infty}\frac{r_n}{n}.
\end{aligned}$$

By the elementary renewal theorem (e.g. Feller (1968)), $\lim_{n\to\infty}\frac{r_n}{n} = E(T_2-T_1)$. Since $P[T_2-T_1 = n_0 s] = \varepsilon(1-\varepsilon)^{(s-1)}$, $E(T_2-T_1) = \sum_{s=1}^{\infty}[n_0 s\varepsilon(1-\varepsilon)^{(s-1)}] = \frac{n_0}{\varepsilon} < \infty$. Therefore

$$\sigma^2 = \frac{n_0}{\varepsilon}\tilde{\sigma}^2 \tag{11}$$

By using Lemmas 3.2, 3.3, the CLT for i.i.d random variables and the result (11), we obtain:

$$\begin{aligned}
\lim_{n\to\infty}\frac{\sum_{j=1}^{r(n)}\sum_{i=T_j}^{T_{j+1}-1}[h(X_i)-\pi(h)]}{n^{1/2}} &= \lim_{n\to\infty}\frac{\sum_{j=1}^{r(n)}\sum_{i=T_j}^{T_{j+1}-1}[h(X_i)-\pi(h)]}{r_n^{1/2}}\cdot\frac{r_n^{1/2}}{n^{1/2}} \\
&\longrightarrow_d \frac{n_0}{\varepsilon}N(0,\tilde{\sigma}^2) \\
&= N(0,\sigma^2)
\end{aligned}$$

as $n\to\infty$. $\square$

Acknowledgement

Both authors would like to thank Jeffrey Rosenthal for his assistance in writing this paper. The first author was supported by an Engineering and Physical Sciences Research Council Studentship and would like to thank Dave Stephens and Chris Holmes for their advice relating to this paper.

REFERENCES

BRADLEY, R. C. 1985. On the central limit question under absolute regularity. Ann. Prob., 13, 1314–1325.

CHAN, K. S. & GEYER, C. J. 1994. Discussion of Markov chains for exploring posterior distributions. Ann. Statist., 22, 1747–1758.

CHEN, X. 1999. Limit theorems for functionals of ergodic Markov chains with general state space. Mem. Amer. Math. Soc., 139.

COGBURN, R. 1972. The central limit theorem for Markov processes. In Le Cam, L. E., Neyman, J. & Scott, E. L. (Eds.) Proc. Sixth Ann. Berkley Symp. Math. Statist. and Prob., 2, 485–512.

FELLER, W. 1968. An Introduction to Probability Theory and its Applications. 3rd ed, Wiley, Chichester.

HÄGGSTRÖM, O. 2005. On the central limit theorem for geometrically ergodic Markov chains. Prob. Theory and Rel. Fields, 132, 74–82.

IBRAGIMOV, I. A. & LINNIK, Y. V. 1971. Independent and stationary sequences of random variables. Wolter-Noordhoff, Groiningen.

JONES, G. L. 2004. On the Markov chain central limit theorem. Prob. Surveys, 1, 299–320.

MEYN, S. P. & TWEEDIE, R. L. 1993. Markov chains and stochastic stability. Springer, New York. http://probability.ca/MT.

NUMMELIN, E. 1984. General irreducible Markov chains and non-negative operators. Cambridge University Press, Cambridge.

ROBERTS, G. O. & ROSENTHAL, J. S. 2004. General state space Markov chains and MCMC algorithms. Prob. Surveys, 1, 20–71.