Chapter 6
Ergodicity and Central Limit Theorems for Markov Processes

Thomas G. Kurtz

6.1 Introduction

There are several contexts in the theory of Markov processes in which the term *ergodicity* is used, but in all of these, assertions of the form

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k) = \int h d\pi, \tag{6.1}$$

or in continuous time,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t h(X(s)) ds = \int h d\pi, \tag{6.2}$$

for some probability measure, π, appear. Limits of this form are essentially laws of large numbers, and given such a limit, it is natural to ask about rates of convergence or fluctuations, in particular, to explore the behavior of the rescaled deviations,

$$\sqrt{n}\left(\frac{1}{n} \sum_{k=1}^{n} h(X_k) - \int h d\pi\right) \text{ or } \sqrt{t}\left(\frac{1}{t} \int_0^t h(X(s)) ds - \int h d\pi\right).$$

T.G. Kurtz (✉)
University of Wisconsin, Madison, WI, USA
e-mail: kurtz@math.wisc.edu

Many times during his career, Rabi has studied problems of this form. The goal of these brief comments is to review some of his results and provide some of the background needed to read his papers.

All processes we consider will take values in a complete separable metric space (E, r). They will be temporally homogeneous and Markov in discrete or continuous time. In discrete time, the *transition function* will be denoted by $P(x, \Gamma)$, that is, there is a filtration $\{\mathcal{F}_k\}$ such that the process of interest $X = \{X_k, k = 0, 1, \ldots\}$ satisfies

$$P\{X_{k+1} \in \Gamma | \mathcal{F}_k\} = P(X_k, \Gamma), \quad k = 0, 1, \ldots, \Gamma \in \mathcal{B}(E), \tag{6.3}$$

where $\mathcal{B}(E)$ denotes the Borel subsets of E. The filtration may be larger than the filtration generated by $\{X_k\}$. When X and $\{\mathcal{F}_k\}$ satisfy (6.3), we will say that X is $\{\mathcal{F}_k\}$-*Markov* with transition function P.

In continuous time, the transition function will be denoted by $P(t, x, \Gamma)$ and there will be a filtration $\{\mathcal{F}_t\}$ such that the process $\{X(t), t \geq 0\}$ satisfies

$$P\{X(s + t) \in \Gamma | \mathcal{F}_s\} = P(t, X(s), \Gamma), \quad s, t \geq 0, \Gamma \in \mathcal{B}(E). \tag{6.4}$$

Setting

$$T(t)f(x) = \int_E f(y)P(t, x, dy), \quad f \in B(E),$$

where $B(E)$ is the space of bounded, Borel measurable functions on E, the Markov property implies $\{T(t)\}$ is a *semigroup*, that is

$$T(s)T(t)f = T(s + t)f.$$

The semigroup can (and will be) defined for larger classes of functions as is convenient.

The notion of an operator A being a *generator* for a Markov process can be defined in a variety of ways, but essentially always implies

$$T(t)f = f + \int_0^t T(s)Af ds,$$

which in turn implies

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))ds \tag{6.5}$$

is a martingale for any filtration satisfying (6.4).

The analog in discrete time to the continuous-time semigroup is obtained by defining the linear operator

$$Pf(x) = \int_E f(y)P(x, dy)$$

and observing that

$$E[f(X_{k+n})|\mathcal{F}_k] = P^n f(X_k),$$

and that

$$M_n = f(X_n) - f(X_0) - \sum_{k=0}^{n-1}(Pf(X_k) - f(X_k)) \tag{6.6}$$

is a martingale for every $f \in B(E)$. Consequently,

$$A = P - I$$

in discrete time plays the role of the generator in continuous time. The martingale properties (6.5) and (6.6) are central to the study of Markov processes and are the basis for the central limit theorems that Rabi and others have given.

By the *initial distribution* of a Markov process, we mean the distribution of X_0 in the discrete case and of $X(0)$ in the continuous-time case. The finite dimensional distributions of a Markov process are determined by its initial distribution and its transition function. If we want to emphasize the initial distribution μ of the process, we will write $\{X_k^\mu\}$ or $\{X^\mu(t)\}$.

The following lemma will prove useful in studying discrete time Markov processes.

Lemma 1. *Let $P(x, \Gamma)$ be a transition function on E. There exists a measurable space $(U, \mathcal{U})$, a measurable mapping $\alpha : U \times E \to E$, and a probability distribution ν on $(U, \mathcal{U})$ such that if ξ has distribution ν, then $\alpha(\xi, x)$ has distribution $P(x, \cdot)$.*

Consequently, if X_0 has distribution $\mu \in \mathcal{P}(E)$ and $\xi_1, \xi_2, \ldots$ is a sequence of independent, ν-distributed, U-valued random variables that is independent of X_0, then for $\mathcal{F}_k = \sigma(X_0, \xi_1, \ldots, \xi_k)$, $\{X_k\}$ defined recursively by

$$X_{k+1} = \alpha(\xi_{k+1}, X_k), \quad k = 0, 1, \ldots,$$

is a $\{\mathcal{F}_k\}$-Markov process with initial distribution μ and transition function $P(x, \Gamma)$.

Proof. The construction in [8] gives α for ξ uniformly distributed on $[0, 1] \times [0, 1]$. A slight modification allows ξ to be uniform on $[0, 1]$.

Remark 1. If the mapping $x \in E \to P(x, \cdot) \in \mathcal{P}(E)$ is continuous taking the weak topology on $\mathcal{P}(E)$, then α given by the Blackwell and Dubins construction has the property that for each $x_0 \in E$, the mapping $x \in E \to \alpha(x, \xi)$ is almost surely continuous at x_0.

The next section reviews ideas of ergodicity of Markov processes and gives some of the basic results. The final section considers central limit theorems exploiting the martingale properties mentioned above. We assume that all continuous-time Markov processes considered are cadlag.

6.2 Ergodicity for Markov Processes

Ideas of ergodicity for Markov processes all relate to the existence of stationary distributions for the processes. In discrete time, $\pi \in \mathcal{P}(E)$ is a stationary distribution if

$$\int_E P(x, \Gamma)\pi(dx) = \pi(\Gamma), \quad \Gamma \in \mathcal{B}(E),$$

and in continuous time, if

$$\int_E T(t)f(x)\pi(dx) = \int_E f(x)\pi(dx), \quad f \in B(E), t \geq 0,$$

which is equivalent to requiring

$$\int_E Af(x)\pi(dx) = 0$$

for a sufficiently large class of f.

If π is a stationary distribution and we take π to be the initial distribution for the process, then $\{X_k^\pi\}$ (or $\{X^\pi(t)\}$) will be a stationary process. If $\{X_k^\pi\}$ is ergodic as defined generally for stationary processes, that is, the tail σ-field

$$\mathcal{T} = \cap_n \sigma(X_k^\pi, k \geq n)$$

only contains events of probability zero or one, we will say that π is an *ergodic stationary distribution*. If there is only one stationary distribution, it must be ergodic. If π is an ergodic stationary distribution, then taking $X = X^\pi$, (6.1) or (6.2) hold for all $h \in B(E)$, or more generally, for all $h \in L^1(\pi)$.

The questions of existence and uniqueness of stationary distributions are among the fundamental questions in the study of Markov processes. If, as is typically the case,

$$P : C_b(E) \to C_b(E), \tag{6.1}$$

or

$$T(t) : C_b(E) \to C_b(E), \quad t \geq 0, \tag{6.2}$$

where $C_b(E)$ is the space of bounded continuous functions on E, proof of existence of a stationary distribution can be reduced to the proof of relative compactness of a sequence of probability measures.

Theorem 1. *Assume that $\{T(t)\}$ satisfies (6.2) and for a corresponding Markov process $\{X^\mu(t)\}$, define a family of probability measures $\{\nu_t\}$ by*

$$\nu_t f = E[\frac{1}{t} \int_0^t f(X(s))ds] = \frac{1}{t} \int_0^t E[f(X(s))]ds$$

$$= \frac{1}{t} \int_0^t \int_E T(s)fd\mu ds, \quad f \in B(E). \tag{6.3}$$

Then as $t \to \infty$, any weak limit point of $\{\nu_t\}$ is a stationary distribution for $\{T(t)\}$.

Similarly, in the discrete-time case, if P satisfies (6.1), any weak limit point of $\{\nu_n\}$ defined by

$$\nu_n f = \frac{1}{n} \sum_{k=0}^{n-1} E[f(X_k)] \tag{6.4}$$

is a stationary distribution for P.

Proof. Suppose $t_n \to \infty$ and $\{\nu_{t_n}\}$ converges weakly to π. Observe that for each $f \in C_b(E)$ and $t > 0$,

$$\frac{1}{t_n} \int_t^{t+t_n} E[f(X(s))]ds = \nu_{t_n} T(t)f$$

has the same limit as $\nu_{t_n} f$. Consequently, $\pi f = \pi T(t)f$, and π is a stationary distribution.

The proof in the discrete case is essentially the same.

The natural approach to proving the existence of the sequence $\{\nu_{t_n}\}$ is to prove relative compactness for $\{\nu_t\}$. Since relative compactness in $\mathcal{P}(E)$ is equivalent to tightness, we have the following.

Corollary 1. *Let E be compact. If $\{T(t)\}$ satisfies (6.2), then there exists at least one stationary distribution for $\{T(t)\}$. Similarly, if P satisfies (6.1), then there exists at least one stationary distribution for P.*

More generally, relative compactness is usually proved by obtaining a Lyapunov function for the process. In particular, we want to find a function $\psi : E \to [0, \infty)$ such that for each $a \geq 0$, the level set

$$\Gamma_a = \{x \in E : \psi(x) \leq a\}$$

is compact and for some initial distribution μ,

$$K \equiv \sup_{t \geq 0} E[\psi(X^\mu(t))] < \infty.$$

It follows that

$$P\{X^\mu(t) \notin \Gamma_a\} = P[\psi(X^\mu(t)) > a] \leq \frac{K}{a}$$

and that

$$\nu_t(\Gamma_a^c) \leq \frac{\nu_t \psi}{a} \leq \frac{K}{a},$$

so $\{\nu_t\}$ is tight and hence relatively compact.

The notion of a *stochastic Lyapunov functions* was developed in [14] and reflects ideas dating back to [10] and [13]. There is a large literature on constructing such functions. In discrete time, we have the following simple condition.

Lemma 2. *Let $\psi : E \to [0, \infty)$. Suppose that there exist $a \geq 0$ and $0 \leq b < 1$ such that*

$$P\psi(x) \leq a + b\psi(x). \tag{6.5}$$

Then for each n

$$P^n\psi(x) \le a\frac{1 - b^{n-1}}{1 - b} + b^n\psi(x),$$

and hence for $\mu \in \mathcal{P}(E)$ satisfying $\int_E \psi d\mu < \infty$,

$$\sup_n E[\psi(X_n^\mu)] \le \frac{a}{1 - b} + \int_E \psi d\mu < \infty.$$

Consequently, if ψ has compact level sets and P satisfies (6.1), then there exists at least one stationary distribution for P.

The analogous result in the continuous-time case is somewhat more delicate. Rewriting (6.5) as

$$(P - I)\psi(x) \le a - (1 - b)\psi(x)$$

and recalling that $P - I$ plays a role analogous to the generator A suggests looking for ψ satisfying

$$A\psi(x) \le a - \epsilon\psi(x),$$

for some positive a and ϵ. To this point, we have only considered A to be defined so that f and Af are in $B(E)$. For many Markov processes, for example, diffusions, the extension of the generator to a large class of unbounded ψ is clear, but even in the diffusion setting with smooth ψ, in general we can only claim that

$$\psi(X(t)) - \psi(X(0)) - \int_0^t A\psi(X(s))ds$$

is a local martingale, not a martingale. Note, however, that if ψ is bounded below and $A\psi$ is bounded above, this local martingale will also be a supermartingale. With that observation in mind, the following lemma provides the desired extension.

The following is essentially a consequence of Fatou's lemma.

Lemma 3. *For $n = 1, 2, \ldots$, let $f_n, Af_n \in B(E)$, and*

$$f_n(X(t)) - f_n(X(0)) - \int_0^t Af_n(X(s))ds$$

be a martingale. Suppose $f_n \ge 0$, $\sup_{n,x} Af_n(x) < \infty$, and for each $x \in E$, $\{f_n(x)\}$ and $\{Af_n(x)\}$ converge. Denote the limits ψ and $A\psi$. Then

$$\psi(X(t)) - \psi(X(0)) - \int_0^t A\psi(X(s))ds \tag{6.6}$$

is a supermartingale.

The supermartingale property is exactly what is needed to give the continuous-time analog of Lemma 2.

Lemma 4. *Let measurable functions* $\psi, A\psi : E \to \mathbb{R}$ *satisfy* $\psi \geq 0$ *and* $\sup_{x \in E} A\psi(x) < \infty$. *For* $\mu \in \mathcal{P}(E)$ *satisfying* $\int_E \psi d\mu < \infty$, *assume that (6.6) with X replaced by* X^μ *is a supermartingale. Suppose*

$$A\psi(x) \leq a - \epsilon\psi(x)$$

Then

$$E[\psi(X^\mu(t))] \leq \frac{a}{\epsilon} \vee \int_E \psi d\mu. \tag{6.7}$$

Consequently, if ψ *has compact level sets and* $\{T(t)\}$ *satisfies (6.2), then there exists at least one stationary distribution for* $\{T(t)\}$.

Proof. Let Z^μ denote the supermartingale. Then

$$
\begin{aligned}
e^{\epsilon t}\psi(X^\mu(t)) &= \psi(X^\mu(0)) + \int_0^t e^{\epsilon s} d\psi(X^\mu(s)) + \int_0^t \epsilon e^{\epsilon s}\psi(X^\mu(s))ds \\
&= \psi(X^\mu(0)) + \int_0^t e^{\epsilon s} dZ^\mu(s) + \int_0^t e^{\epsilon s}(\epsilon\psi(X^\mu(s)) + A\psi(X^\mu(s)))ds \\
&\leq \psi(X^\mu(0)) + \int_0^t e^{\epsilon s} dZ^\mu(s) + \int_0^t e^{\epsilon s}a\,ds.
\end{aligned}
$$

Since $E[\int_0^t e^{\epsilon s} dZ^\mu(s)] \leq 0$,

$$E[\psi(X^\mu(t))] \leq e^{-\epsilon t}\int_E \psi d\mu + \frac{a}{\epsilon}(1 - e^{-\epsilon t}),$$

and the lemma follows.

We can relax the conditions of Lemma 4 and still obtain relative compactness of $\{\nu_t\}$ but without the moment estimate (6.7).

Lemma 5. *Let measurable functions* $\psi, A\psi : E \to \mathbb{R}$ *satisfy* $\psi \geq 0$ *and* $K = \sup_{x \in E} A\psi(x) < \infty$. *For* $\mu \in \mathcal{P}(E)$ *satisfying* $\int_E \psi d\mu < \infty$, *assume that (6.6) with X replaced by* X^μ *is a supermartingale. Suppose that for each* $a > 0$

$$\Gamma_a = \{x : A\psi(x) \geq -a\}$$

is compact. Then $\{\nu_t\}$ *defined by (6.3) is relatively compact, and if* $\{T(t)\}$ *satisfies (6.2), there exists at least one stationary distribution for* $\{T(t)\}$.

If E is locally compact and $\{T(t)\}$ *satisfies (6.2), then existence of a stationary distribution holds as long as* Γ_a *is compact for some* $a > 0$.

Remark 2. The assumption that Γ_a is compact only for some $a > 0$ is, in general, not enough to ensure relative compactness of $\{\nu_t\}$. If, however, the process is Harris recurrent (see Section 6.2.2), then existence of a stationary distribution implies convergence of $\{\nu_t\}$.

Proof. The supermartingale property implies

$$- \int_E A\psi d\nu_t = -\frac{1}{t}E[\int_0^t A\psi(X(s))ds] \leq \frac{1}{t}\int_E \psi d\mu - \frac{1}{t}E[\psi(X(t))]] \leq \frac{1}{t}\int_E \psi d\mu,$$

and for $a > 0$,

$$a\nu_t(\Gamma_a^c) \leq K + \frac{1}{t}\int_E \psi d\mu,$$

giving tightness and hence relative compactness for $\{\nu_t\}$.

The second part of the lemma follows by the observation that $\{\nu_t\}$ is relatively compact as a probability measure on the one-point compactification of E and the compactness of Γ_a for some $a > 0$ implies that any limit point ν_∞ satisfies $\nu_\infty(\Gamma_a) > 0$ and hence $\nu_\infty(E) > 0$. Normalizing the restriction of ν_∞ to E to be a probability measure gives a stationary distribution for $\{T(t)\}$. See Theorem 4.9.9 of [9].

The following lemma gives conditions which coupled with some kind of irreducibility imply recurrence, but not necessarily positive recurrence.

Lemma 6. *Let measurable functions* $\psi, A\psi : E \to \mathbb{R}$ *satisfy* $\psi \geq 0$ *and* $K = \sup_{x \in E} A\psi(x) < \infty$. *For* $\mu \in \mathcal{P}(E)$ *satisfying* $\int_E \psi d\mu < \infty$, *assume that (6.6) with X replaced by* X^μ *is a supermartingale. Suppose that for each* $a > 0$,

$$\Gamma_a = \{x : \psi(x) \leq a\}$$

is compact and that there exists a_0 *such that*

$$\sup_{x \in \Gamma_{a_0}^c} A\psi \leq 0.$$

Let $\tau_0 = \inf\{t \geq 0 : X^\mu(t) \in \Gamma_{a_0}$ *and* $\gamma_a = \inf\{t \geq 0 : X^\mu(t) \notin \Gamma_a\}$. *Then*

$$\lim_{a \to \infty} P\{\gamma_a \leq \tau_0\} = 0. \tag{6.8}$$

Proof. It is at least not immediately obvious that $\gamma_a < \infty$ implies $\psi(X^\mu(\gamma_a)) \geq a$, so some randomization may be necessary for a complete proof, but assuming this inequality holds, the supermartingale property implies

$$aP\{\gamma_a \leq \tau_0\} \leq E[\psi(X^\mu(\gamma_a \wedge \tau_0))] \leq \int_E \psi d\mu,$$

and (6.8) follows.

Example 1. In [3], Rabi gives a class of ψ of the form $\psi(x) = F(|x - z|)$ for nondegenerate diffusion processes which satisfy the conditions of Lemma 6. (Actually, in Rabi's notation, we need to set $\psi(x) = -F(|x - z|)$.) The non-degeneracy assumption then ensures Harris recurrence (see below). He also formulates similar conditions that imply transience and gives a construction of an F such that $\psi(x) = -F(|x - z|)$ satisfies the conditions of the second part of Lemma 5.

A central idea in the study of uniqueness of stationary distributions is the notion of *Harris recurrence*.

6.2.1 Harris Recurrence

Harris irreducibility requires the existence of a measure φ on $\mathcal{B}(E)$ such that $\varphi(B) > 0$ implies that the Markov process visits B with positive probability, regardless of the initial distribution. If the process visits such B infinitely often with probability one, or in the continuous time case, the process visits B for arbitrarily large times, that is, $\tau_n = \inf\{t > n : X(t) \in B\}$ is finite almost surely for each n, the process is *Harris recurrent*. As long as φ is σ-finite, without loss of generality, we can and will assume φ is a probability measure. In discrete time, the classical conditions for Harris recurrence can be formulated under the assumption that there exists a function $\varepsilon : E \to [0, 1]$ such that the transition function satisfies

$$P(x, B) \geq \varepsilon(x)\varphi(B) \tag{6.9}$$

and that for each initial condition μ, the Markov process satisfies

$$P\{\sum_{k=1}^{\infty} \varepsilon(X_k^\mu) = \infty\} = 1. \tag{6.10}$$

The following lemma illustrates the significance of these conditions.

Lemma 7. *Let $\mu \in \mathcal{P}(E)$, and suppose that (6.9) and (6.10) hold. Then there exists a probability space with a process X^μ, a filtration $\{\mathcal{F}_k^\mu\}$, and a $\{\mathcal{F}_k^\mu\}$-stopping time τ^μ such that X^μ is $\{\mathcal{F}_k^\mu\}$-Markov with initial distribution μ and transition function $P(x, \Gamma)$ and the distribution of $X_{\tau^\mu}^\mu$ is φ.*

Proof. We enlarge the state space to be $E \times \{-1, 1\}$ and define the new transition function by

$$Q(x, \theta, \Gamma \times \{\theta\}) = P(x, \Gamma) - \varepsilon(x)\varphi(\Gamma)$$

and

$$Q(x, \theta, \Gamma \times \{-\theta\}) = \varepsilon(x)\varphi(\Gamma).$$

If (X^μ, Θ) is a Markov process with this transition function such that X_0^μ has distribution μ, then X^μ is a Markov process with transition function $P(x, \Gamma)$ and initial distribution μ, and the desired stopping time is $\tau^\mu = \min\{k : \theta_k \neq \theta_{k-1}\}$. Note that

$$P\{\tau^\mu > n\} = E[\prod_{k=0}^{n-1}(1 - \varepsilon(X_k^\mu))]$$

and (6.10) implies $P\{\tau^\mu < \infty\} = 1$.

Much of the work on Harris recurrence is done under weaker conditions of the form

$$\sum_{n=1}^{\infty} a_n P^n(x, \Gamma) \geq \varepsilon(x)\varphi(\Gamma),$$

where $a_n(x) \geq 0$, $\sum_{n=1}^{\infty} a_n(x) = 1$, or in continuous time,

$$\int_0^{\infty} P(t, x, \Gamma)a_x(dt) \geq \varepsilon(x)\varphi(\Gamma),$$

where a_x is a probability distribution on $(0, \infty)$, and typically, $\varepsilon(x)$ has the form $\varepsilon \mathbf{1}_C(x)$ for some constant $\varepsilon > 0$ and $C \in \mathcal{B}(E)$. The analog of Lemma 7 holds under these conditions, at least if (6.10) is replaced by

$$P\{\sum_{k=1}^{\infty} \varepsilon(X_k^\mu)^2 = \infty\} = 1 \text{ or } P\{\int_0^{\infty} \varepsilon(X^\mu(s))^2 ds = \infty\} = 1.$$

The existence of these stopping times implies the desired uniqueness of the stationary distribution and convergence in total variation of ν_n and ν_t.

Lemma 8. *Let $\varphi \in \mathcal{P}(E)$. Suppose that for each $\mu \in \mathcal{P}(E)$, on some probability space, there exists a process X^μ a filtration $\{\mathcal{F}_k^\mu\}$, and a $\{\mathcal{F}_k^\mu\}$-stopping time τ^μ such that X^μ is $\{\mathcal{F}_k^\mu\}$-Markov with initial distribution μ and transition function P and $X_{\tau^\mu}^\mu$ has distribution φ. Then there is at most one stationary distribution for P.*

If there is a stationary distribution π, then for each initial distribution $\mu \in \mathcal{P}(E)$, $\{\nu_n\}$ defined by (6.4) converges in total variation to π.

The analogous result holds in continuous time.

Proof. Suppose π_1 and π_2 are stationary distributions for P. Let X^{π_1} and X^{π_2} satisfy the hypotheses of the lemma. By the ergodic theorem, for each $h \in B(E)$, we can define

$$H_{\pi_i}^h = \lim_{n \to \infty} \frac{1}{n}\sum_{k=0}^{n-1} h(X_k^{\pi_i}) = \lim_{n \to \infty} \frac{1}{n}\sum_{k=\tau^{\pi_i}}^{\tau^{\pi_i}+n-1} h(X_k^{\pi_i}) = \lim_{n \to \infty} \frac{1}{n}\sum_{k=0}^{n-1} h(X_{\tau^{\pi_i}+k}^{\pi_i}).$$

By the strong Markov property, $H_{\pi_1}^h$ and $H_{\pi_2}^h$ must have the same distribution. Since $\pi_i h = E[H_{\pi_i}^h]$, $\pi_1 = \pi_2$.

Under the hypotheses of the lemma, for $h \in B(E)$ and $\mu \in \mathcal{P}(E)$,

$$|E[\frac{1}{n}\sum_{k=0}^{n-1} h(X_k^\mu)] - E[\frac{1}{n}\sum_{k=0}^{n-1} h(X_k^\varphi)]|] \le \|h\|(P\{\tau^\mu > k\} + \frac{2k}{n}),$$

and hence

$$|E[\frac{1}{n}\sum_{k=0}^{n-1} h(X_k^\mu)] - \pi h| \le \|h\|(P\{\tau^\mu > k\} + P\{\tau^\pi > k\} + \frac{4k}{n}), \qquad (6.11)$$

and taking the sup over $h \in B(E)$ with $\|h\| \le 1$ gives the convergence in total variation.

If τ^φ satisfies $0 < E[\tau^\varphi] < \infty$, then $\tau_1 = 0$ and $\tau_2 = \tau^\varphi$ provide an example of τ_1 and τ_2 in the following lemma.

Lemma 9. *Let X be $\{\mathcal{F}_k\}$-Markov with transition function P, and let τ_1 and τ_2 be stopping times satisfying $\tau_1 \le \tau_2$ and $0 < E[\tau_2 - \tau_1] < \infty$ such that X_{τ_1} and X_{τ_2} have the same distribution. Then π defined by*

$$\pi h = \frac{E[\sum_{k=\tau_1+1}^{\tau_2} h(X_k)]}{E[\tau_2 - \tau_1]}$$

is a stationary distribution for P.
In continuous time,

$$\pi h = \frac{E[\int_{\tau_1}^{\tau_2} h(X(s))ds]}{E[\tau_2 - \tau_1]}.$$

Remark 3. In the case $0 < E[\tau^\varphi] < \infty$, this observation is essentially the renewal argument of [1] and [16].

Proof. Since

$$M_n = h(X_n) - h(X_0) - \sum_{k=0}^{n-1}(Ph(X_k) - h(X_k))$$

is a martingale,

$$0 = E[h(X_{\tau_2}) - h(X_{\tau_1})] = E[\sum_{k=\tau_1}^{\tau_2-1}(Ph(X_k) - h(X_k)),$$

and hence,

$$\pi Ph = \pi h,$$

so π is a stationary distribution for P.

Example 2. In [6], Rabi and Mukul Majumdar consider processes in $E = (0, 1)$ of the form

$$X_{n+1} = \xi_{n+1} X_n (1 - X_n),$$

where the $\{\xi_k\}$ are iid with values in $(0, 4)$. Clearly, this process satisfies (6.1). Under the assumption that the distribution of ξ has an absolutely continuous part with a density that is strictly positive on some interval, they give conditions for Harris recurrence.

Example 3. The inequality in (6.11) and the analogous inequality in continuous time,

$$|E[\frac{1}{t} \int_0^t h(X^\mu(s))ds] - \pi h| \le \|h\|(P\{\tau^\mu > r\} + P\{\tau^\pi > r\} + \frac{4r}{t}), \qquad (6.12)$$

actually give rates of convergence. Under aperiodicity assumptions, one can replace the average by $E[h(X^\mu(t))]$ and eliminate the $O(t^{-1})$ term. In [7], Rabi and Aramian Wasielak give conditions under which this can be done for a class of diffusion processes.

6.2.2 Conditions without Harris Recurrence

Harris recurrence is very useful when it holds, or perhaps more to the point, when it can be shown that it holds. In general, it does not hold, even in relatively simple settings. Perhaps the best known example is the "Markov process" in $[0, 1)$ given by the recursion

$$X_{n+1} = X_n + z \bmod 1,$$

for some irrational z.

For an example with more interesting probabilistic structure, let $E = \{-1, 1\}^\infty$, and consider a generator of the form

$$A f(x) = \sum_{k=1}^{\infty} \lambda_k (f(\eta_k x) - f(x)), \qquad (6.13)$$

where $\lambda_k > 0$, $\sum_k \lambda_k < \infty$, and $\eta_k x$ is obtained by replacing x_k by $-x_k$. If $x, y \in E$ differ on infinitely many components, then $P(t, x, \cdot)$ and $P(t, y, \cdot)$ are mutually singular for all t, but for all $x \in E$, $P(t, x, \cdot)$ converges weakly to the distribution under which the components are independent symmetric Bernoulli.

In general, infinite dimensional processes provide a source of examples that are not Harris recurrent even if ergodic. We will not address any more examples of this type, but see [11] for recent work in this direction.

There is a need for techniques for studying ergodicity for processes that are not Harris recurrent. One approach that appears frequently in Rabi's work involves the notion of *splitting* and is discussed in the paper by Ed Waymire in this volume. A second approach considered by Rabi and Gopal Basak in [2] is by verifying *asymptotical flatness*, that is,

by showing that X^x and X^y can be coupled in such a way that for each compact $K \subset E$ and $\varepsilon > 0$,

$$\lim_{t\to\infty} \sup_{x,y\in K} P\{|X^x(t) - X^y(t)| > \varepsilon\} = 0.$$

For example, if one rewrites the generator in (6.13) as

$$Af(x) = \sum_{k=1}^{\infty} 2\lambda_k\left(\frac{1}{2}f(\eta_k^1 x) + \frac{1}{2}f(\eta_k^{-1} x) - f(x)\right), \tag{6.14}$$

where $\eta_k^1 x$ is obtained from x by replacing x_k by 1 and $\eta_k^{-1} x$ is obtained by replacing x_k by -1, then the coupling can be obtained using independent Poisson processes $\{N_k\}$, N_k with intensity $2\lambda_k$, and at the lth jump of N_k replacing x_k by ξ_{kl}, where the $\{\xi_{kl}\}$ are independent symmetric Bernoulli.

Example 4. In [2], Rabi and Gopal Basak consider diffusions of the form

$$X^x(t) = x + \int_0^t BX^x(s)ds + \int_0^t \sigma(X^x(s))dW(s).$$

One has a natural coupling simply by using the same Brownian motion W for both X^x and X^y. Lyapunov-type arguments are again employed but with analytic estimates rather than simply compactness arguments. In particular, the arguments employ ψ (v in the notation of the paper) of the form

$$\psi(x) = (x \cdot Cx)^{1-\varepsilon},$$

for appropriately chosen positive definite C and $\varepsilon \in [0, 1)$. Different choices of C are applied to $\psi(X^x(t))$ to ensure the existence of a stationary distribution and to $\psi(X^x(t) - X^y(t))$ to give the asymptotic flatness.

6.3 Central Limit Theorems

There are many version of the martingale central limit theorem. See, for example, [15, 17, 12]. The following version is from Theorem 7.1.4 of [9].

Theorem 2. *Let $\{M_n\}$ be a sequence of cadlag, $\mathbb{R}^d$-valued martingales, with $M_n(0) = 0$, and let $A_n = [M_n]$ be the matrix of covariations, that is,*

$$A_n^{ij}(t) = [M_n^i, M_n^j]_t.$$

Suppose that for each $t \geq 0$,

$$\lim_{n\to\infty} E[\sup_{s\leq t}|M_n(s) - M_n(s-)|] = 0$$

and

$$\lim_{n \to \infty} A_n(t) = A(t),$$

where A is deterministic and continuous. Then $\{M_n\}$ converges in distribution to a Gaussian process M such that M has independent increments, $E[M(t)] = 0$, and $[M^i, M^j]_t = E[M^i(t)M^j(t)] = A^{ij}(t)$.

If

$$A(t) = \int_0^t \sigma(s)\sigma(s)^T ds,$$

for some $d \times m$-matrix-valued function σ, then we can write

$$M(t) = \int_0^t \sigma(s)dW(s),$$

where W is an $\mathbb{R}^m$-valued standard Brownian motion.

Example 5. Let π be an ergodic stationary distribution for a Markov semigroup $\{T(t)\}$. Then $\{T(t)\}$ extends to $L^2(\pi)$ and is strongly continuous on $L^2(\pi)$. Let A be the Hille-Yosida generator for the semigroup on $L^2(\pi)$. Then for each $f \in \mathcal{D}(A)$, the domain of A,

$$M^f(t) = f(X^\pi(t)) - f(X^\pi(0)) - \int_0^t Af(X^\pi(s))ds$$

is a square integrable martingale.

Then, for $h \in L^2(\pi)$, ergodicity implies

$$\lim_{n \to \infty} \frac{1}{n} \int_0^{nt} h(X^\pi(s))ds = \pi h t,$$

and Theorem 2.1 of [4] gives the functional central limit theorem for the scaled deviations,

$$Z_n^h(t) = \frac{1}{\sqrt{n}} \int_0^{nt} (h(X^\pi(s)) - \pi h)ds.$$

The key assumption is that there exists $f \in \mathcal{D}(A)$ such that $Af = h - \pi h$. Then

$$Z_n^h(t) = \frac{1}{\sqrt{n}}(f(X^\pi(nt)) - f(X^\pi(0))) - \frac{1}{\sqrt{n}}M^f(nt).$$

Consequently, we have the functional central limit theorem for $\{Z_n^h\}$ provided we can prove the functional central limit theorem for $\{\frac{1}{\sqrt{n}}M^f(n\cdot)\}$. Observe that the quadratic variation of $\frac{1}{\sqrt{n}}M^f(n\cdot)$ is the same as the quadratic variation for $U_n(t) = \frac{1}{\sqrt{n}}f(X^\pi(n\cdot))$ and that by Itô's formula,

$$[U_n]_t = U_n(t)^2 - \frac{1}{n} \int_0^t 2f(X^\pi(ns-))df(X^\pi(ns)) \qquad (6.1)$$

$$= U_n(t)^2 - \frac{1}{n} \int_0^t 2f(X^\pi(ns-))dM^f(ns) - \int_0^t 2f(X^\pi(ns))Af(X^\pi(ns))ds$$

$$\rightarrow -t \int_E 2f(x)Af(x)\pi(dx).$$

By Theorem 2, the convergence of Z_n^h follows. Of course, under the assumptions of Lemma 8, the same result will hold for X^μ for all $\mu \in \mathcal{P}(E)$.

If f is smooth and X^π is an $\mathbb{R}^d$-valued diffusion,

$$X^\pi(t) = X^\pi(0) + \int_0^t \sigma(X(s))dW(s) + \int_0^t b(X(s))ds,$$

then

$$f(X^\pi(t)) = f(X^\pi(0)) + \int_0^t \nabla f(X^\pi(s))^T \sigma(X^\pi(s))dW(s) + R(t),$$

where R is continuous with finite variation, so we can also write

$$[U_n]_t = \frac{1}{n} \int_0^{nt} \nabla f(X^\pi(s))^T \sigma(X^\pi(s))\sigma(X^\pi(s))\nabla f(X^\pi(s))ds \qquad (6.2)$$

$$\rightarrow \int_{\mathbb{R}^d} \nabla f(x)^T \sigma(x)\sigma(x)^T \nabla f(x)\pi(dx).$$

Example 6. In [5], Rabi considers diffusions in $\mathbb{R}^d$ of the form

$$X(t) = X(0) + \int_0^t u_0 b(X(s))ds + \int_0^t \sigma(X(s))dW(s),$$

under the assumption that σ is the square root of a positive definite matrix and σ and b are periodic in the sense that

$$\sigma(x + z) = \sigma(x) \text{ and } b(x + z) = b(x) \quad z \in \mathbb{Z}^d.$$

At least under additional regularity assumptions on σ and b, $Y(t) = X(t) \bmod \mathbf{1}, \mathbf{1} \in \mathbb{Z}^d$, the vector with each component 1, is a Markov process in $[0, 1)^d$ which has a unique, ergodic stationary distribution π. Then

$$\lim_{n \to \infty} \frac{1}{n} X(nt) = \lim_{t \to \infty} \frac{1}{n} \int_0^{nt} u_0 b(X(s))ds = \lim_{n \to \infty} \frac{1}{n} \int_0^{nt} u_0 b(Y(s))ds = u_0 \bar{b} t,$$

where $\bar{b} = \pi b$. Rabi gives the corresponding central limit theorem showing the convergence of

$$V_n(t) = \frac{1}{\sqrt{n}}(X(nt) - nu_0\bar{b}t).$$

For simplicity, assume $X(0) = 0$. Setting

$$M_1^n(t) = \frac{1}{\sqrt{n}}\left(X(nt) - \int_0^{nt} u_0 b(Y(s))ds\right) = \frac{1}{\sqrt{n}}\int_0^{nt}\sigma(Y(s))dW(s),$$

the convergence of M_1^n follows from Theorem 2 and the ergodicity of Y.

Note that $V_n = M_1^n + Z^n$, where

$$Z^n(t) = \frac{1}{\sqrt{n}}u_0\int_0^{nt}(b(Y(s)) - \bar{b})ds.$$

Then Z^n is of the form treated in [4], Example 5. Let $\hat{A}$ denote the generator for Y, which will satisfy $\hat{A}f(x \bmod \mathbf{1}) = Af(x)$, if f extends periodically to an element in the domain of A. Rabi shows the existence of a twice continuously differentiable g satisfying

$$\hat{A}g(y) = b(y) - \bar{b},$$

and setting

$$M_2^n(t) = \frac{u_0}{\sqrt{n}}\left(g(Y(nt)) - g(0) - \int_0^{nt}(b(Y(s)) - \bar{b})ds\right),$$

we have

$$V_n(t) = M_1^n(t) - M_2^n(t) + \frac{u_0}{\sqrt{n}}(g(Y(nt)) - g(0)).$$

Since

$$M_2^n(t) = \frac{u_0}{\sqrt{n}}\int_0^{nt}\nabla g(X(s))^T\sigma(X(s))dW(s) + R_n(t),$$

where R_n is continuous with finite variation,

$$[M_1^n - M_2^n]_t = \frac{1}{n}\int_0^{nt}(I - u_0\nabla g(X(s))^T)\sigma(X(s))\sigma(X(s))^T(I - u_0\nabla g(X(s)))ds.$$

Setting $a = \sigma\sigma^T$,

$$D = \int_{[0,1)^d}(I - u_0\nabla g(y)^T)a(y)(I - u_0\nabla g(y))\pi(dy),$$

and V_n converges in distribution to a mean zero Brownian motion with covariance matrix D. The form of D derived here differs from the form in [5], but compare (6.1) and (6.2).

References

[1] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978. ISSN 0002-9947. doi: 10.2307/1998882. URL http://dx.doi.org/10.2307/1998882.

[2] Gopal K. Basak and Rabi N. Bhattacharya. Stability in distribution for a class of singular diffusions. *Ann. Probab.*, 20(1):312–321, 1992. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(199201)20:1<312:SIDFAC>2.0.CO;2-L&origin=MSN.

[3] R. N. Bhattacharya. Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *Ann. Probab.*, 6(4):541–553, 1978. ISSN 0091-1798.

[4] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. verw. Gebiete*, 60(2):185–201, 1982. ISSN 0044-3719. doi: 10.1007/BF00531822. URL http://dx.doi.org/10.1007/BF00531822.

[5] Rabi Bhattacharya. A central limit theorem for diffusions with periodic coefficients. *Ann. Probab.*, 13(2):385–396, 1985. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(198505)13:2<385:ACLTFD>2.0.CO;2-S&origin=MSN.

[6] Rabi Bhattacharya and Mukul Majumdar. Stability in distribution of randomly perturbed quadratic maps as Markov processes. *Ann. Appl. Probab.*, 14(4):1802–1809, 2004. ISSN 1050-5164. doi: 10.1214/105051604000000918. URL http://dx.doi.org/10.1214/105051604000000918.

[7] Rabi Bhattacharya and Aramian Wasielak. On the speed of convergence of multidimensional diffusions to equilibrium. *Stoch. & Dyn.*, 12(1):1150003, 19, 2012. ISSN 0219-4937. doi: 10.1142/S0219493712003638. URL http://dx.doi.org/10.1142/S0219493712003638.

[8] David Blackwell and Lester E. Dubins. An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.*, 89(4):691–692, 1983. ISSN 0002-9939. doi: 10.2307/2044607. URL http://dx.doi.org/10.2307/2044607.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-08186-8.

[10] F. G. Foster. On the stochastic matrices associated with certain queuing processes. *Ann. Math. Statistics*, 24:355–360, 1953. ISSN 0003-4851.

[11] Martin Hairer and Jonathan C. Mattingly. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.*, 16:no. 23, 658–738, 2011. ISSN 1083-6489. doi: 10.1214/EJP.v16-875. URL http://dx.doi.org/10.1214/EJP.v16-875.

[12] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. ISBN 0-12-319350-8. Probability and Mathematical Statistics.

[13] R. Z. Hasminskii. Ergodic properties of recurrent diffusion processes and stabilization of the solution of the Cauchy problem for parabolic equations. *Theory Probab. Appl.*, 5:179–196, 1960. ISSN 0040-361x.

[14] Harold J. Kushner. *Stochastic Stability and Control*. Mathematics in Science and Engineering, Vol. 33. Academic Press, New York-London, 1967.

[15] D. L. McLeish. Dependent central limit theorems and invariance principles. *Ann. Probab.*, 2:620–628, 1974.

[16] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. verw. Gebiete*, 43(4):309–318, 1978. ISSN 0178-8051.

[17] Rolando Rebolledo. Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete*, 51(3):269–286, 1980. ISSN 0044-3719. doi: 10.1007/BF00587353. URL http://dx.doi.org/10.1007/BF00587353.