Advanced computational methods X071521-SDE
Lecture 6

In this lecture, we study long time behaviors of SDEs and numerical schemes of SDEs. We will mainly look at the paper "Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise" by Mattingly, Stuart and Higham.

I plan to write another lecture notes based on Talay's paper , but this will not be gone over in class. This will be left as free reading.

1 Two conditions ensureing geometric ergodicity

We will first of all present two conditions that guarantee the geometric ergodicity with certain weights.

The first one is a Dooblin condition or minorization condition. The second condition is regarding existence of Lyapunov functions.

The existence of Lyapunov function in some sense means that the system will be pulled back to some central compact region if it escapes away. This is some recurrent condition. In discretization, this is usually the one that can be lost.

The minorization condition is a property that says a certain point in a compact set C is uniformly reachable from inside of C. This is a property on compact set and it is often satisfied after discretization.

1.1 A discrete example regarding the minorization property and geometric ergodicity

To enter our main theory, we first of all look at a simple example and try to get the ideas.

Suppose there is a Markov chain taking values in discrete space. Without loss of generality, we label them to be $1, 2, \ldots$. The transition probability is

$$p_{ij} = \mathbb{P}(X^{n+1} = j | X^n = i), \quad \sum_j p_{ij} = 1.$$

Assume that there exist values c_j such that

$$\inf_i p_{ij} \ge c_j \ge 0, \quad c := \sum_j c_j > 0.$$

This implies that there exists j^* such that $c_{j^*} > 0$. In other words, there is a special state such that all other states jump to this state with a significant probability. This condition is called the **Doeblin's condition**, which somehow will be similar to the minorization condition as we shall see.

We define a probability measure

$$\nu_j := c_j / c.$$

Now, we split this from the transition probability as

$$p_{ij} = (1-c)\tilde{p}_{ij} + c\nu_j.$$

It is easy to verify that $\tilde{p}_{ij} \geq 0$ and it is also a transition probability.

The idea is to use ν as the coupling measure. The idea is to construct a coupling betweent two chains. We consider two chains X_1 and X_2 with initial distributions α and β. We now sample i.i.d random variables $U^n \sim \nu$. We also sample i.i.d Bernoulli random variables α^n with $\mathbb{P}(\alpha^n = 1) = c$. We set $\tau = 0$. We now define the following transitions recursively as follows for $n \geq 1$:

- If $\tau = 0$, then we sample Y_1^n with $\mathbb{P}(Y_1^n = j | X_1^n = i) = \tilde{p}_{ij}$ and

  $$X_1^n = (1-\alpha^n)Y_1^n + \alpha^n U^n.$$

  Similarly, we get X_2^n. If $\alpha^n = 1$, we set $\tau = 1$. Move to next n.

- If $\tau = 1$, then $X_1^n = X_2^n$ and the value is determined by the transition probability p_{ij}. Move to next n.

Define the stopping time

$$\mathcal{N} = \inf\{n \geq 1 : \alpha^n = 1\}.$$

We have $\mathbb{P}(X_2^n = i, \mathcal{N} \leq n) = \mathbb{P}(X_2^n = i, \mathcal{N} \leq n)$. Therefore,

$$\begin{aligned}|\mathbb{P}(X_1^n = i) - \mathbb{P}(X_2^n = i)| &= |\mathbb{P}(X_1^n = i, \mathcal{N} > n) - \mathbb{P}(X_2^n = i, \mathcal{N} > n)| \\ &\leq \mathbb{P}(X_1^n = i, \mathcal{N} > n) + \mathbb{P}(X_2^n = i, \mathcal{N} > n).\end{aligned}$$

Taking sum on i, we find

$$TV(\alpha P^n, \beta P^n) \leq 2\mathbb{P}(\mathcal{N} > n) = 2(1-c)^n.$$

If β is the stationary distribution μ, then $\mu P^n = \mu$ and thus

$$TV(\alpha P^n, \mu) \leq 2(1-c)^n.$$

If we do coupling on the initial distribution as well, the bound can be reduced to $2TV(\alpha, \beta)(1-c)^n$.

Remark 1. *Another common way to prove convergence is to use the Perron-Frobenious theorem and spectral gaps. This is more close to functional analysis for the semigroups (involving log Sobolev inequality, hypercontractivity and so on).*

Remark 2. *If the force field is contracting and the noise is additive, we can do synchronized coupling. For example,*

$$dX = b(X)dt + dW$$

with $(b(X) - b(Y)) \cdot (X - Y) \le -r|X - Y|^2$. *Then, we can consider two processes so that they are given* **the same** *Brownian motion. This will not affect the distribution. Then, we have* $\mathbb{E}|X-Y|^2 \le -r\mathbb{E}|X-Y|^2$. *This gives the gemeotric convergence in Wasserstein distance. We will not cover this in class.*

1.2 The Doeblin and minorization condition

Let $X(t)$ ($t \in \mathbb{R}_+$ or $t \in \mathbb{N}$) be a Markov chain taking values in $\mathbb{R}^d$. Denote the transition kernel

$$P_t(x, A) := \mathbb{P}(X(t) \in A | X(0) = x), \ A \text{ is a Borel set.}$$

To consider the ergodicity, we will look at the Markov chain at discrete times. In particular, assume we observe the Markov chain at $t = nT$ for $T \in I$. Define the transition kernel

$$P(x, A) := P_T(x, A).$$

Let $\{X_n\}$ be the Markov chain generated by the kernel $P(x, A)$. The filtration $\{\mathcal{F}_n\}$ is the one associated with $\{X_n\}$.

Condition 1. *The Markov chain* $\{X(t)\}$ *and its transition kernel* $P_t(x, A)$ *satisfy the following, for some fixed compact Borel set* $C \subset \mathbb{R}^d$*:*

1. *There exists* $y^* \in int(C)$ *such that for any* $\delta > 0$*, there is* $t_1 \in I$ *with*

$$P_{t_1}(x, B(y^*, \delta)) > 0, \ \forall x \in C.$$

2. *For all* $t \in I$*, the transition kernel has a density* $p_t(x, y)$ *on* C *such that*

$$P_t(x, A) = \int_A p_t(x, y)\, dy, \quad \forall x \in C, A \in \mathscr{B}(\mathbb{R}^d) \cap \mathscr{B}(C).$$

Moreover, $p_t(\cdot, \cdot)$ *is continuous on* $C \times C$*.*

Clearly, this condition is an analogy of the Doeblin's condition mentioned above. This condition implies the following *minorization condition*

Lemma 1. *Assume Condition 1 holds. Then, there exist $T \in I$, $\eta > 0$ and a probability measure ν supported on C (i.e. $\nu(C) = 1$) such that*

$$P(x, A) \geq \eta\nu(A), \quad \forall A \in \mathscr{B}(\mathbb{R}^d), \ x \in C.$$

Proof. By Condition 1, there exists $\delta_1 > 0$ and $t_2 > 0$ such that

$$B(y^*, \delta_1) \subset C, \quad P_{t_2}(y^*, B(y^*, \delta_2)) > 0.$$

Together with the continuity of the density, there exists $z^* \in B(y^*, \delta_2) \subset C$ such that

$$p_{t_2}(y, z) \geq \epsilon, \ \forall y \in B(y^*, \delta_3) \subset C, \ z \in B(z^*, \delta_4) \subset C,$$

with some $\delta_3 > 0$ and $\delta_4 > 0$.

Hence, for all $y \in B(y^*, \delta_3)$,

$$P_{t_2}(y, A) \geq \int_{A \cap B(z^*, \delta_4)} p_{t_2}(y, z) \, dz \geq \epsilon |A \cap B(z^*, \delta_4)|.$$

Now, by the first assumption in Condition 1, we can find $t_1 > 0$ such that

$$P_{t_1}(x, B(y^*, \delta_3)) > 0, \ \forall x \in C.$$

By the continuity of the density and dominate convergence theorem, $x \mapsto P(x, B(y^*, \delta_3))$ is continuous. It follows from the compactness of C that

$$\inf_{x \in C} P_{t_1}(x, B(y^*, \delta_3)) \geq \gamma > 0.$$

Taking $T = t_1 + t_2$, for all $x \in C$, we have by Markov property that

$$\begin{aligned} P(x, A) &\geq \int_{B(y^*, \delta_3)} p_{t_1}(x, y) P_{t_2}(y, A) \, dy \\ &\geq \epsilon |A \cap B(z^*, \delta_4)| \int_{B(y^*, \delta_3)} p_{t_1}(x, y) \, dy \geq \epsilon\gamma |A \cap B(z^*, \delta_4)|. \end{aligned}$$

Define

$$\nu(A) = \frac{|A \cap B(z^*, \delta_4)|}{|B(z^*, \delta_4)|}.$$

We find that ν is a probability measure supported on C. Setting

$$\eta = \epsilon\gamma |B(z^*, \delta_4)|$$

then finishes the proof. □

With the minorization condition, we can do the same thing as above. Namely, we construct explicit coupling using ν ad random iterated functions (Kifer, 1988).

We define transition kernel

$$\tilde{P}(x, A) = \begin{cases} P(x, A), & x \in C^c, \\ \frac{1}{1-\eta}[P(x, A) - \eta\nu(A)], & x \in C. \end{cases}$$

According to the construction of in Kifer, 1988, $\tilde{P}(x, A)$ can be realized by iteration of random family

$$\tilde{h}(x, w), \ x \in \mathbb{R}^d,$$

where $w \in \Omega$ is a random vector. Then, one can define a new random function as

$$g(x, \omega) = 1_C(x)[\phi\tilde{h}(x, w) + (1 - \phi)\xi] + (1 - 1_C(x))\tilde{h}(x, w).$$

Here, $\omega = (w, \phi, \xi)$ with $\mathbb{P}(\phi = 1) = 1 - \eta$, $\mathbb{P}(\phi = 0) = \eta$, whereas $\xi \sim \nu$. Clearly, the transition kernel given by g is also $P(x, A)$.

Therefore, we can generate

$$X'_{n+1} = g(X'_n, \omega_n), \ \omega_n = (w_n, \phi_n, \xi_n) \ i.i.d.$$

to get a chain that has the same distribution as $\{X_n\}$.

For the coupling construction, we do the same thing. Consider another chain

$$Y'_{n+1} = g(Y'_n, \eta_n), \ \eta_n = (W_n, \phi_n, \xi_n).$$

Here, ϕ_n and ξ_n are the same. W_n is has the same distribution as w_n. Before $\phi_n = 0$, W_n is independent of w_n. After $\phi_n = 0$, we take $W_n = w_n$. Therefore, as long as $\phi_n = 0$, after that $X'_n = Y'_n$.

It is clear now that the measure ν is like an atom. If the chains touch, then they move in the same way. Next, we will check the Lyapunov function condition to guarantee that the two chains return back to C fequently so that there is a high chance that $X'_n = Y'_n$.

Remark 3. *Regarding the minorization condition, we required the existence of density that is continuous. For SDEs, this condition can be guranteed by the hypoelliptic condition of the Fokker-Planck operator $\partial_t - \mathcal{L}^*$. A classical theorem regarding this hypoellipticity is the Hörmander's theorem. Roughly speaking, this requires the vector fields associated with the SDE and their* **iterated** *Poisson brackets will be d-dimensional for each point $x \in \mathbb{R}^d$. Another condition is the irreducibility.*

1.3 The Lyapunov function condition

Condition 2. *With the choice of $T \in I$ above, there is a function $V : \mathbb{R}^d \to [1, \infty)$, $\lim_{|x|\to\infty} V(x) = \infty$, and real numbers $\alpha \in (0, 1)$, $\beta \in [0, \infty)$ such that*

$$\mathbb{E}[V(X_{n+1})|\mathcal{F}_n] \le \alpha V(X_n) + \beta.$$

For SDEs, this condition can be ensured by some condition on the generator of SDEs.

Lemma 2. *Consider the SDE $dX = b(X)dt + \sigma(X)dW$. If there is a function $V : \mathbb{R}^d \to [1, \infty)$ with $\lim_{|x|\to\infty} V(x) = \infty$ such that*

$$\mathcal{L}V(x) \le -aV(x) + d,$$

for some $a > 0$ and $d > 0$, then the above Lyapunov conditon holds for any $T > 0$.

For any $\gamma \in (\alpha, 1)$ and $s \in [1, \infty)$, we can define

$$c(s) = \frac{s\beta}{\gamma - \alpha},$$

and set

$$C(s) = \{x : V(x) \le c(s)\}.$$

With Condition 2, we find easily that

$$\mathbb{E}(V(X_{n+1})|\mathcal{F}_n) \le \gamma V(X_n) + s\beta 1_{C(s)}(X_n).$$

Fixing $s = 2$, and define $c := c(s)$, $C := C(2)$. Then, we have the following result

Lemma 3. *Let $\mathcal{N}$ be a stopping time and fix $n \ge 0$. Under Condition 2, we have*

$$\mathbb{E}[V(X_n)1_{N\ge n}] \le \kappa\gamma^n \left[V(x_0) + \mathbb{E}(\sum_{j=1}^{n\wedge\mathcal{N}} \gamma^{-j} 1_C(X_{j-1})) \right]$$

Proof. Using the inequaltiy, $\mathbb{E}(V(X_{n+1})|\mathcal{F}_n) \le \gamma V(X_n) + s\beta 1_{C(s)}(X_n)$, it is natural to consider the random varible

$$M_n = \gamma^{-n} V(X_n).$$

With the stopping time, we use the technique from Martingale

$$\mathbb{E}M_{n\wedge\mathcal{N}} = \mathbb{E}M_0 + \sum_{j=1}^{n} \mathbb{E}(1_{j\leq\mathcal{N}}(M_j - M_{j-1}))$$

Since $\{j \leq \mathcal{N}\} = \{\mathcal{N} \leq j-1\}^c \in \mathcal{F}_{j-1}$, we find

$$\mathbb{E}(1_{j\leq\mathcal{N}}(M_j - M_{j-1})) = \mathbb{E}(1_{j\leq\mathcal{N}}\mathbb{E}(M_j - M_{j-1}|\mathcal{F}_{j-1}))$$

Using $\mathbb{E}(M_j - M_{j-1}|\mathcal{F}_{j-1}) \leq \gamma^{-j}(2\beta 1_C(X_{j-1}))$ (recall $s = 2$), we have

$$\mathbb{E}M_{n\wedge\mathcal{N}} \leq V(x_0) + \sum_{j=1}^{n} \gamma^{-j} 2\beta\mathbb{E}(1_{j\leq\mathcal{N}} 1_C(X_{j-1}))$$

Finally, noting that $\mathbb{E}\gamma^{-n}V(X_n)1_{N\geq n} \leq \mathbb{E}M_{n\wedge\mathcal{N}}$ and letting $\kappa = 2\beta$, the claim follows. □

Corollary 1. *Assume Condition 2 holds. Let γ and C be defined as above (Note that C depends on γ). Define $\tau_C = \inf\{n > 0 : X_n \in C\}$. Then,*

$$\mathbb{P}(\tau_C > n) \leq \kappa_1\gamma^n(V(x_0) + 1),$$

and

$$\mathbb{E}(\frac{1}{\gamma})^{\tau_C} \leq \kappa_2(V(x_0) + 1).$$

In fact, $\mathbb{E}(\sum_{j=1}^{n\wedge\mathcal{N}} \gamma^{-j} 1_C(X_{j-1}) = \gamma^{-1}1_C(X_0)$ (only the first term is nonzero). Moreover, the left handside is estimated as $\mathbb{E}[V(X_n)1_{\tau_C\geq n}] \geq \mathbb{E}[V(X_n)1_{\tau_C>n}] \geq c\mathbb{P}(\tau_C > n)$. The second claim follows by definition

$$\mathbb{E}(\frac{1}{\gamma})^{\tau_C} = \sum_{n=1}^{\infty}(\gamma^{-1})^n\mathbb{P}(\tau_C = n).$$

Clearly, $\mathbb{P}(\tau_C = n) \leq \mathbb{P}(\tau_C > n-1) \leq \kappa_3\gamma_1^{-n+1}$ where $\alpha < \gamma_1 < \gamma$. Then, done.

Lastly, we use the control above to conclude the existence of invariant measure.

Lemma 4. *Assume Condition 2 holds. Then, X_n has an invariant measure.*

Taking $\mathcal{N} = n$, one find that $\sup_{n\ge 0}\mathbb{E}V(X_n) \le \frac{\kappa(V(x_0)+1)}{1-\gamma}$. It follows that

$$\lim_{R\to\infty}\sup_n \mathbb{P}(|X_n| \ge R) = 0.$$

Define the sequence of mesures

$$\mu_n(A) = \frac{1}{n}\sum_{k=0}^{n}\mathbb{P}(X_k \in A).$$

The above property implies that μ_n is tight. (There is a compact set so that $\sup_n \mu_n(A^c) < \epsilon$ for any given ϵ). Then, there is a subsequence that converges weakly to a probability measure μ. The limit measure is an invariant measure.

2 The ergodicity theorem and the proof

Theorem 1. *Consider the Markov chain $\{X(t)\}$ with transition kernel $P_t(x, A)$. Assume that there is $T \in I$ such that the minorization condition in Lemma 1 and Condition 2 hold with*

$$C = \left\{x : V(x) \le \frac{2\beta}{\gamma - \alpha}\right\},$$

where $\gamma \in (\alpha^{1/2}, 1)$. Then, the embedded chain $\{X_n\}$ has a uniqe invariant measure π. Further, there exist $r \in (0,1)$ and $\kappa > 0$ such that for any f with $|f| \le V$, we have

$$\left|\mathbb{E}_x f(X_n) - \int f\,d\pi\right| \le \kappa V(x) r^n.$$

The proof will rely on the coupling argument as above. Recall that we have constructed two chains $\{X_n'\}$ and $\{Y_n'\}$ with random variables $\omega_n(w_n, \phi_n, \xi_n)$ and $\eta_n = (W_n, \phi_n, \xi_n)$. After $\phi_n = 0$, the two chains become the same. Moreover, we enlarge the filtration to $\mathcal{F}_n$ that are generated by X_n' and Y_n'. The lemma proved above are still valid. Without loss of generality, we can assume f is nonnegative because any function can be decomposed into the difference of two nonnegative functions $f = f^+ - f^-$.

Consider the coupling time (which is a stopping time)

$$\zeta := \inf\{n \ge 0 : (X_n', Y_n') \in C \times C,\ \phi_n = 0\}.$$

Using this definition, we have:

$$\mathbb{E}(f(X'_n)-f(Y'_n)) = \mathbb{E}((f(X'_n)-f(Y'_n))1_{\zeta>n}) \leq \max\{\mathbb{E}(V(X'_n)1_{\zeta>n}), \mathbb{E}(V(Y'_n)1_{\zeta>n})\}.$$

We first of all establish the following key result:

Lemma 5. *With the assumptions and notations in the theorem, for any $\gamma \in (\alpha^{1/2}, 1)$, we have that*

$$\max\{\mathbb{E}(V(X'_n)1_{\zeta>n}), \mathbb{E}(V(Y'_n)1_{\zeta>n})\} \leq \kappa[\mathbb{E}(V(X_0)+V(Y_0))+1]r^n,$$

with some $r \in (0,1)$ and $\kappa > 0$.

Proof. The proof here relies on the lemmas applied to the Markov chain

$$Z_n := (X'_n, Y'_n),$$

with Lyapunov function

$$U(z) = V(x) + V(y).$$

Clearly, we have

$$\mathbb{E}(U(Z_{n+1})|\mathcal{F}_n) \leq \alpha U(Z_n) + 2\beta.$$

Similarly, we have

$$\mathbb{E}(U(Z_{n+1})|\mathcal{F}_n) \leq \gamma U(Z_n) + 2\beta 1_{C_1}(Z_n),$$

where

$$C_1 = \{z = (x,y) : U(z) \leq \frac{2\beta}{\gamma - \alpha}\} \subset C \times C.$$

Now, it is exactly this reason why we choose $s = 2$ above.

Similarly, define the coupling time

$$\theta = \inf\{n \geq 0 : Z_n \in C_1, \phi_n = 0\}.$$

Then,we have $\zeta \leq \theta$ and

$$\max\{\mathbb{E}(V(X'_n)1_{\zeta>n}), \mathbb{E}(V(Y'_n)1_{\zeta>n})\} \leq \mathbb{E}(V(X'_n)1_{\zeta>n}+V(Y'_n)1_{\zeta>n}) \leq \mathbb{E}(U(Z_n)1_{\theta>n})$$

Hence, it sufficies to consider $\mathbb{E}(U(Z_n)1_{\theta>n})$.

For $\theta > n$, there are typically two cases: The chain returns to C_1 but at the time when it returns $\phi = 1$; the second case is that the chain Z_n never comes back to C_1. Intuitively, both cases are unlikely. We need to make the intuition rigorous.

We define τ_k to be the kth time that the chain returns to C_1, which is clearly a stopping time. We also define $\tau_0 = 0$. For real number s, τ_s is understood as τ_k with k being the smallest interger no less than s. Fix

$$a \in (0,1)$$

to be determined.

Then,

$$\mathbb{E}(U(Z_n)1_{\theta>n}) = \mathbb{E}(U(Z_n)1_{\theta>n}1_{\tau_{an}} \le n) + \mathbb{E}(U(Z_n)1_{\theta>n}1_{\tau_{an}>n}) = I_1 + I_2.$$

We consider I_1:

$$\begin{aligned}\mathbb{E}(U(Z_n)1_{\theta>n}1_{\tau_{an}} \le n) &\le \mathbb{E}(U(Z_n)1_{\tau_{an}<\theta}1_{\tau_{an}} \le n) \\ &= \mathbb{E}(1_{\tau_{an}\le n}\mathbb{E}(U(Z_n)1_{\theta>\tau_{an}}|\mathcal{F}_{\tau_{an}}))\end{aligned}$$

Note that $\{\tau_{an} \le n\} \in \mathcal{F}_{\tau_{an}}$.

Next, we apply Lemma 3 to the conditional expectation using strong Markov property. We then have

$$\mathbb{E}(U(Z_n)1_{\theta>\tau_{an}}|\mathcal{F}_{\tau_{an}}) \le \kappa\gamma^{n-\tau_{an}}U(X_{\tau_{an}}) + \kappa \sum_{j=1}^{\theta\wedge n-\tau_{an}} \gamma^{n-\tau_{an}-j}$$

Therefore, we have

$$1_{\tau_{an}\le n}\mathbb{E}(U(Z_n)1_{\theta>\tau_{an}}|\mathcal{F}_{\tau_{an}}) \le \kappa\|U\|_{L^\infty(C_1)}\gamma^{n-\tau_{an}} + \kappa_1 1_{\tau_{an}<\theta}$$

Clearly, if there is no θ, the second term will be $O(1)$ (τ_{an} could be small).

For the second term:

$$\begin{aligned}I_2 = \mathbb{E}(U(Z_n)1_{\theta>n}1_{\tau_{an}} > n) &\le \sum_{k=0}^{[an]-1} \mathbb{E}(U(Z_n)1_{\tau_k\le n}1_{\tau_{k+1}>n}) \\ &= \sum_{k=0}^{[an]-1} \mathbb{E}(1_{\tau_k\le n}\mathbb{E}(U(Z_n)1_{\tau_{k+1}>n}|\mathcal{F}_{\tau_k}))\end{aligned}$$

We apply Lemma 3 again using strong Markov property:

$$\mathbb{E}(U(Z_n)1_{\tau_{k+1}>n}|\mathcal{F}_{\tau_k}) \le \gamma^{n-\tau_k}\kappa(U(Z_{\tau_k}) + \gamma^{-1}) \le \kappa\gamma^{n-\tau_k}$$

The second term is like that because $Z_m \notin C_1$ for any $m \in (\tau_k, n]$.

Overall, we therefore have

$$I_1 + I_2 \le \kappa\gamma^n \sum_{k=0}^{[an]-1} \mathbb{E}\gamma^{-\tau_k} + \kappa\mathbb{E}1_{\tau_{an}<\theta} \le \kappa\gamma^n \sum_{k=0}^{[an]-1} \mathbb{E}\gamma^{-\tau_k} + \kappa(1-\eta)^{an}$$

For the first term, we use strong Markov property.

$$\mathbb{E}\gamma^{-\tau_k} = \mathbb{E}(\gamma^{-\tau_{k-1}}\mathbb{E}(\gamma^{-(\tau_k-\tau_{k-1})}|\mathcal{F}_{\tau_{k-1}})) \le L\mathbb{E}(\gamma^{-\tau_{k-1}})$$

for some $L > 1$ independent of k by Corollary 1. Therefore, we finally have

$$I_1 + I_2 \le \kappa(\gamma^n L^{an} + (1-\eta)^{an})$$

Choose a close to 1, we have $\gamma L^a < 1$ and $(1-\eta)^a < 1$, and the claim then follows. □

Now, we finish the proof:

Since the chain has an invariant measure, we can then let Y_0' to have the distribution as π. Then, it will preserves to have this distribution.

$$|\mathbb{E}_x f(X_n') - \int f\, d\pi| \le 2\kappa[V(x) + \int V\, d\pi + 1]r^n.$$

The uniqueness of invariant measure and the claim all follow.

Above, we only proved the ergodicity for $t_n = nT$ with $T \in I$. What about general $t \in I$? In fact, the time at $t_n' = nT + \delta$ can be viewed as the Markov chain starting with initial value $X(\delta)$. The same estiamtes given that

$$|\mathbb{E}f(X(t_n')) - \int f\, d\pi| \le Br^n$$

for some B related to $X(\delta)$. Since $\delta < T$, B can be uniformly controlled.