Continuous-time markov processes, martingale problem, and diffusion approximation

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: August 1, 2023

1 Martingales and local martingales

Let $(\Omega, \mathcal{F}, P)$ a probability space. A collection of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ indexed by $[0, \infty)$ is said to be a filtration if $\mathcal{F}_t \subseteq \mathcal{F}_{t+s}$ for any $t, s \in [0, \infty)$. A filtration is said to be right-continuous if for every $t \geq 0$, $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$. A filtration is complete if $(\Omega, \mathcal{F}, P)$ is complete and $\{A \in \mathcal{F} : P(A) = 0\} \subseteq \mathcal{F}_0$. A right-continuous and complete filtration is said to be standard (or to satisfy the usual conditions). A random variable T with values in $[0, \infty]$ is said to be a stopping with respect to a filtration $(\mathcal{F}_t)$ if $\{T \leq t\} \in \mathcal{F}_t$ for every $t \geq 0$. A process $X = (X_t)_{t\geq 0}$ is said to be adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ if for every $t \geq 0$, X_t is $\mathcal{F}_t$-measurable. All processes we now consider take values in $\mathbb{R}$ or $\mathbb{R}^d$.

Definition 1 (Martingale)**.** A process X is said to be a martingale with respect to a filtration $(\mathcal{F}_t)$ if:
(i) X is adapted to $(\mathcal{F}_t)$;
(ii) X is integrable, i.e., $\mathbb{E}[|X_t|] < \infty$ for all $t \geq 0$;
(iii) $\mathbb{E}[X_{t+s}|\mathcal{F}_t] = X_t$ for all $t, s \geq 0$.

It can be shown that all martingales with respect to standard filtrations have a version with cadlag sample path (see Lalley's notes). In what follows, we always consider martingales with cadlag sample path.

If X is a martingale, the optional stopping theorem states that if T is a bounded stopping time, then the stopped process $X^T := (X_{t\wedge T})_{t\geq 0}$ is also a martingale. "Stopping can be used [...] to truncate the path of a process in order to gain more integrability or tightness while keeping adaption and continuity" (D. Chafai). This prompts us to introduce a category of processes that generalize martingales.

Definition 2 (Local Martingale)**.** A cadlag process adapted to a standard filtration is said to be a local martingale if there exists a sequence of stopping time $(T_n)_{n\in\mathbb{N}}$ with $T_1 \leq T_2 \leq \ldots$ and $T_n \to \infty$ a.s. such that X^{T_n} is a martingale for every $n \in \mathbb{N}$.

Every martingale is a local martingale, but the converse is false. A martingale can be understood as the fortune of a player in a fair game. A strictly local martingale can be understood as the fortune of a player in a game that is only locally fair – it is akin to a financial bubble. Strict local martingales do not exist in discrete-time.

2 Stochastic integration

[REFs. Le Gall BMMSC (2013), Etheridge's lecture notes CMSC (2018), Berestycki's lecture notes SCA (2010)]

In this section we establish the notion of stochastic integrals (in the sense of Itô) for continuous martingales (i.e., martingales with continuous sample path). We would like to define an integral of the form

$$\int_0^t f(s)\, dM_s := \int_0^t f(\omega, s)\, dM_s(\omega)$$

as a Stieltjes integral, but the variations of the sample path of M_s are too wild for that. For instance, the sample path of a Brownian motion is a.s. nowhere differentiable, and so exhibits a.s. infinite total variation. The trick to rigorously define stochastic integrals is to leverage the fact that continuous martingales have finite quadratic variations: stochastic integrals can be defined as limits in some appropriately defined L^2 space of (well-defined) integrals of some simple processes that takes only finitely many (bounded) values.

Denote $\mathcal{M}^2$ the set of all cadlag martingales bounded in L^2 and $\mathcal{M}^2_c$ the set of all continuous martingales bounded in L^2.

Proposition 3. *Let $X \in \mathcal{M}^2$. There exists $X_\infty \in L^2$ such that*

$$X_t \to X_\infty \quad \text{in } L^2 \text{ and a.s., as } t \to \infty.$$

Moreover, $X_t = \mathbb{E}\left[X_\infty | \mathcal{F}_t\right]$ a.s. for all $t \geq 0$.

The set $\mathcal{M}^2_c$ is a normed linear space which we endow with the norm $\|M\|^2 := \mathbb{E}\left[M^2_\infty\right]$.

Definition 4 (Simple Process)**.** A simple process is any map $H \colon \Omega \times (0, \infty) \to \mathbb{R}$ of the form

$$H(\omega, t) := \sum_{i=0}^{m} Z_i(\omega) \mathbb{1}_{\{(t_i, t_{i+1}]\}}(t)$$

where $n \in \mathbb{N}$, $0 = t_0 < \cdots < t_{m+1} < \infty$, and Z_i is a bounded $\mathcal{F}_{t_i}$-measurable random variable for all i. The stochastic integral for H with respect to $M \in \mathcal{M}^2$ is defined as

$$\int_0^t H_s\, dM_s := \sum_{i=0}^{m} Z_i (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}).$$

We denote the set of simple processes by $\mathcal{E}$.

Theorem 5 (Quadratic Variation)**.** *Let M be a continuous local martingale. There exists a unique (up to indistinguishability) adapted continuous non-decreasing process denoted by $[M] := ([M]_t)_{t \geq 0}$ such that $[M]_0 = 0$ a.s. and the process given by $M_t^2 - [M]_t$ is a continuous local martingale. The process $[M]$ is said to be the quadratic variation of M.*

Proposition 6. *Let M be a continuous local martingale. For any $T > 0$ and any sequence of partitions $\pi_n = \{0 = t_0^n \leq \cdots \leq t^n_{n(\pi_n)} = T\}$ with $\delta(\pi_n) := \sup_{1 \leq i \leq n(\pi_n)} (t_i^n - t_{i-1}^n) \to 0$ as $n \to \infty$,*

$$[M]_T = \lim_{n \to \infty} \sum_{i=1}^{n(\pi_n)} \left(M_{t_i^n} - M_{t_{i-1}^n} \right)^2$$

in probability.

If B is a standard Brownian motion, it is a continuous local martingale, and $[B]_t = t$.

The stochastic integral of a simple process with respect to a continuous martingale bounded in L^2 is also a continuous martingale (bounded in L^2), and in particular a continuous local martingale. It can be shown that

$$\left[\int_0^t H_s \, dM_s\right]_t = \sum_{i=1}^m Z_i^2 \left([M]_{t_{i+1}\wedge t} - [M]_{t_i \wedge t}\right) = \int_0^t H_s^2 \, d[M]_s.$$

The idea is then to use $\mathbb{E}\,(\int_0^\infty H_s^2 \, d[M]_s)$ as a guide to define the L^2 space we are after. It will then turn out that the map $H \mapsto (\int_0^t H_s \, dM_s)_{t\geq 0}$ is a linear isometry. If simple processes are dense in the considered L^2 space, then we will be able to define stochastic integrals for all processes in this L^2 space via a limiting procedure.

Definition 7 ($L^2(M)$ Space)**.** Given $M \in \mathcal{M}_c^2$, we denote by $L^2(M)$ the space of previsible processes K such that

$$\|K\|^2_{L^2(M)} := \mathbb{E}\left(\int_0^\infty K_s^2 d[M]_s\right) < \infty.$$

The space $L^2(M)$ is standard L^2 space (and hence an Hilbert space) for the measure μ given by

$$\mu(A \times (s,t]) := \mathbb{E}\,(\mathbb{1}_{\{A\}}([M]_t - [M]_s))$$

for all $s < t$, $A \in \mathcal{F}_s$. We have that $\mathcal{E} \subseteq L^2(M)$ for any $M \in \mathcal{M}_c^2$.

Proposition 8. *Let $M \in \mathcal{M}_c^2$. Then $\mathcal{E}$ is dense in $L^2(M)$.*

Theorem 9 (Itô's Isometry)**.** *Let $M \in \mathcal{M}_c^2$. The map $H \mapsto (\int_0^t H_s \, dM_s)_{t\geq 0}$ from $\mathcal{E}$ to $\mathcal{M}_c^2$ has a unique extension to a linear isometry from $L^2(M)$ to $\mathcal{M}_c^2$ which we denote*

$$K \mapsto \left(\int_0^t K_s \, dM_s\right)_{t\geq 0}.$$

The isometry property rewrites as

$$\left\|\left(\int_0^t K_s \, dM_s\right)_{t\geq 0}\right\|^2 = \mathbb{E}\left(\left(\int_0^\infty K_s \, dM_s\right)^2\right) = \mathbb{E}\left(\int_0^\infty K_s^2 d[M]_s\right) = \|K\|^2_{L^2(M)}.$$

Stopping at suitable stopping times, the Itô integral can then be extended from $\mathcal{M}_c^2$ to continuous local martingales. The integrators can then be extended to locally bounded previsible processes (which include continuous adapted process). A previsible process K is locally bounded if there exists increasing stopping times S_n with $S_n \to \infty$ a.s. such that $K\mathbb{1}_{(0,S_n]}$ is bounded for all $n \in \mathbb{N}$.

Proposition 10. *Let K a locally bounded previsible process for stopping times S_n and M a continuous local martingale. Consider the stopping times $S'_n := \inf\{t \geq 0 : |M_t| \geq n\}$ and set $T_n := S_n \wedge S'_n$. Then for all $t \geq T_n$,*

$$\int_0^t K_s \mathbb{1}_{(0,S_n]}(s) \, d[M_n^T]_s$$

is well-defined and is denoted $\int_0^t K_s \, dM_s$.

3 Stochastic differential equations

[Le Gall, Oksendal, Klenke, Etheridge's lecture notes, Revuz&Yor and Ethier&Kurz (for definition of Markov processes)]

A stochastic differential equation is simply a standard differential equation augmented by a noise term, which is typically of the form σB_t where B_t is a Brownian motion and σ is a constant corresponding to the intensity of the noise. "[T]he use of Brownian motion [...] is justified due to its property of independent increments[:] the random perturbations affecting disjoint time intervals are assumed to be independent". In other words, a standard ODE

$$y_t = y_0 + \int_0^t b(y_s)\, ds$$

is augmented to

$$Y_t = Y_0 + \int_0^t b(Y_s)\, ds + \sigma B_t.$$

By allowing the intensity to depend on the state of the system at time t, we obtain an equation of the form

$$Y_t = Y_0 + \int_0^t b(Y_s)\, ds + \int_0^t \sigma(Y_s)\, dB_s$$

where the rightest integral is a stochastic integral as previously defined. The equation can be further extended to the inhomogeneous case by allowing b and σ to depend on t, that is,

$$Y_t = Y_0 + \int_0^t b(s, Y_s)\, ds + \int_0^t \sigma(s, Y_s)\, dB_s,$$

which rewrites in differential form as

$$dY_t = b(t, Y_t)\, dt + \sigma(t, Y_t)\, dB_t.$$

It is standard to call b the drift coefficient and σ the diffusion coefficient.

Definition 11 (Solution SDEs)**.** Let d and m be positive integers. Let b and σ be locally bounded measurable functions from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}^{d\times m}$ and $\mathbb{R}^d$. A solution of the stochastic differential equation $E(\sigma, b)$

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

consists of:

(i) a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ where the filtration is assumed complete;

(ii) an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $B = (B^1, \ldots, B^m)$ taking values in $\mathbb{R}^m$;

(iii) an $(\mathcal{F}_t)_{t\geq 0}$-adapted process $X = (X^1, \ldots, X^d)$ with values in $\mathbb{R}^d$ and continuous sample paths such that

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s,$$

meaning that for every $i = 1, \dots, d$,

$$X_t^i = X_0^i + \int_0^t b_i(s, X_s)\, ds + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, X_s)\, dB_s^j.$$

When, in addition, $X_0 = x \in \mathbb{R}^d$, X is said to be a solution started from x.

As implicit in the definition, when we speak of a solution of $E(\sigma, b)$, the filtered probability space and the Brownian motion B need not be fixed a priori but can be constructed at the same time as the process X. This possibility leads to two notions of solution: weak and strong. "Intuitively, a strong solution corresponds to solving the SDE for a given Brownian motion, while in producing a weak solution we are allowed to construct the Brownian motion and the solution at the same time" (Durrett). Moreover, as made explicit in next definition, there are at least two reasonable notions of uniqueness for solutions of SDEs.

Definition 12. Let $E(\sigma, b)$ the SDE in the previous definition. We say that:

(i) $E(\sigma, b)$ has a solution if for each $x \in \mathbb{R}^d$, there exists a solution of $E(\sigma, b)$ started from x;

(ii) there is uniqueness in law if all solutions of $E(\sigma, b)$ started from x has same distribution;

(iii) there is pathwise uniqueness if, when $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ and B are fixed, any two solutions X and X' satisfying $X_0 = X_0'$ a.s. are indistinguishable;

(iv) a solution X for $E(\sigma, b)$ started from x is a strong solution if X is adapted to the natural filtration of B.

4 Markov processes and diffusion processes

Definition 13 (Transition Family)**.** Let $(E, \mathcal{E})$ be a measurable space. A family of probability kernels $(\kappa_{s,t})_{0\leq s<t}$ on $(E, \mathcal{E})$ is said to be a transition function (t.f.) if for all $s < t < v$,

$$\int \kappa_{s,t}(x, dy)\kappa_{t,v}(y, A) = \kappa_{s,v}(x, A)$$

for all $x \in E$ and all $A \in \mathcal{E}$. This is known as the Chapman–Kolmogorov equation. The t.f. is said to be homogeneous if $\kappa_{s,t}$ depends on s and t only through $t - s$. In that case, we write κ_t for $\kappa_{0,t}$ and the Chapman–Kolmogorov equation rewrites as

$$\kappa_{s+t}(x, A) = \int \kappa_s(x, dy)\kappa_t(y, A)$$

for all $s, t \geq 0$; in other words, the family $(\kappa_t)_{t\geq 0}$ forms a semigroup which is called transition semigroup.

Definition 14 (Continuous-Time Markov Process)**.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space. An adapted process X is said to be a Markov process with respect to $(\mathcal{F}_t)_{t\geq 0}$ with transition function $\kappa_{s,t}$ if for any positive Borel function f and any $s < t$,

$$\mathbb{E}\left(f(X_t)|\mathcal{F}_s\right) = \int_E f(y)\kappa_{s,t}(X_s, dy), \quad P \text{ a.s.}.$$

The probability measure $X_0(Q)$ is called the initial distribution of X. The process is said to be homogeneous if the t.f. is homogeneous, in which case the above equation reads as

$$\mathbb{E}\left(f(X_{s+t})|\mathcal{F}_s\right) = \int_E f(y)\kappa_t(X_s, dy), \quad P \text{ a.s..}$$

Definition 15 (Strong Markov Property)**.** Let X be a cadlag (time-homogeneous) Markov process, $(\mathcal{F}_t)_{t\geq 0}$ the usual enlargement of the natural filtration, and τ an $\mathcal{F}_t$-stopping time. Then X is said to be strong Markov at τ if for every positive Borel function f,

$$\mathbb{E}\left(f(X_{t+\tau})|\mathcal{F}_\tau\right) = \int_E f(y)\kappa_t(X_\tau, dy), \quad P \text{ a.s. on } \tau < \infty,$$

for all $t \geq 0$. The process is said to be strong Markov if it is strong Markov for all stopping times of $(\mathcal{F}_t)_{t\geq 0}$.

Strong Markov (time-homogeneous) processes in $\mathbb{R}^d$ with continuous sample paths are called diffusion processes (but the definition is not standardized in the literature – see Protter). As shown in next section, these processes arise naturally as solutions of (time-homogeneous) SDEs (under regularity conditions on the coefficients). Since (under these conditions) solutions of SDEs can be characterized by some martingale properties involving infinitesimal generators, this justifies some authors to define diffusion processes directly in terms of these properties. On the other hand, some other authors define diffusion processes as solutions of SDEs (without necessarily the strong Markov property). These distinctions are often immaterial since, at least in the homogeneous case, standard (sufficient) conditions for existence of solutions of SDEs are sufficient for the martingale characterization and the strong Markov property.

Definition 16 (Diffusion Process)**.** A process with values in $\mathbb{R}^d$ is said to be a diffusion process if:

(i) it has continuous sample path;

(ii) it is a time-homogeneous strong Markov process.

Proposition 17. *Suppose that the functions σ and b are continuous on $\mathbb{R}_+ \times \mathbb{R}^d$ and Lipschitz in the variable x, i.e., there is a constant K such that for every $t \geq 0$, $x, y \in \mathbb{R}^d$,*

$$\|\sigma(t,x) - \sigma(t,y)\| \leq K\|x - y\|,$$
$$\|b(t,x) - b(t,y)\| \leq K\|x - y\|.$$

Then $E(\sigma, b)$ admits pathwise unique strong solutions.

If, in addition, the coefficients σ and b are homogeneous, i.e., $\sigma(t,x) = \sigma(x)$ and $b(t,x) = b(x)$, then the solutions of $E(\sigma, b)$ are Markov processes with transition semigroup given by

$$\kappa_t f(x) = \mathbb{E}\left(f(X_t^x)\right)$$

where X^x is an arbitrary solution of $E(\sigma, b)$ started from x. The semigroup is Feller and its generator is such that $C_K^2(\mathbb{R}^d) \subseteq D(L)$ and, for every $f \in C_K^2(\mathbb{R}^d)$,

$$Lf(x) = \frac{1}{2}\sum_{i,j=1}^{d}(\sigma\sigma^T)_{ij}(x)\frac{\partial^2 f}{\partial x_i x_j}(x) + \sum_{i=1}^{d} b_i(x)\frac{\partial f}{\partial x_i}(x).$$

Moreover, the solutions of $E(\sigma, b)$ satisfy the strong Markov property.

5 Martingale problem and weak solutions of SDEs

[REFs. Stroock&Varadhan (1979), Berestycki's lecture notes, Etheridge's lecture notes, Durrett (1996), Ethier&Kurtz (1986), Karatzas&Shreve, Rogers&Williams, Revuz&Yor]

In 1948 Lévy proved that the standard Brownian motion could be characterized as the only continuous process B such that $B(t)$ and $B^2(t) - t$ are both martingales. This connection was further explored by Varadhan and Stroock: they leveraged the existence of a larger correspondence between weak solutions of SDEs and processes satisfying some martingale properties to derive important results for solutions of SDEs. Indeed, the martingale characterization proves more suitable for continuity and weak convergence arguments: a number of important existence and approximation results can be derived from the martingale side. (See, e.g., S.5.4. in Karatzas&Shreve or R.19.8. in Rogers&Williams.)

Definition 18 (Martingale Problem)**.** Let $(\sigma_{i,j}(x))_{1\leq i,j\leq d}$ and $(b_i(x))_{1\leq i\leq d}$ be families of real-valued measurable functions. Define $a(x) := \sigma(x)\sigma(x)^T$. A continuous process X with values in $\mathbb{R}^d$, together with a standard filtered probability space, is said to solve the martingale problem $M(a, b)$ if for all $1 \leq i, j \leq d$,

$$Y^i := \left(X_t^i - \int_0^t b_i(X_s)\,ds\right)_{t\geq 0}$$

and

$$\left(Y_t^i Y_t^j - \int_0^t a_{ij}(X_s)\,ds\right)_{t\geq 0}$$

are local martingales.

When $d = 1$, the processes are $Y_t := X_t - \int_0^t b(X_s)\,ds$ and $Y_t^2 - \int \sigma^2(X_s)\,ds$.

Suppose that the SDE given by

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t,$$

admits a solution X. Then, since the stochastic integral is a local martingale, the process given by $Y_t := X_t - X_0 + \int_0^t b(X_s)\,ds$ is a mean zero local martingale. Squaring and using the Itô isometry for stochastic integrals, we have that $Y_t^2 - \int \sigma^2(X_s)\,ds$ is a local martingale. That is, X solves the martingale problem $M(a, b)$. It turns out that the converse is true: there is a one-to-one correspondence between (distributions of) weak solutions to stochastic differential equations and solutions to martingale problems. In particular, existence and uniqueness for solutions of SDEs imply well-posed martingale problems and well-posed martingale problems imply existence and uniqueness for solutions of SDEs.

Theorem 19. *Let X be a solution to $M(a, b)$. Then there exists a Brownian motion B, possibly defined on an enlarged probability space, such that (X, B) solves*

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t.$$

Proof. T.5.4.5. in Durrett (1996) p.199 or T.6.1. Berestycki's lecture notes SCA (2010) p.87. □

Proposition 20. *If a and b are locally bounded and $M(a, b)$ is well-posed, then the solution satisfies the strong Markov property.*

Proof. T.5.4.5. in Durrett (1996) □

If the coefficients are Lipschitz, we know that there is a unique strong solution to the corresponding SDE, which implies that the corresponding martingale problem is well-posed. Weaker conditions for the existence of weak solutions can be derived from the martingale side. In particular, Stroock and Varadhan derived the

Theorem 21. *Suppose that a and b are measurable and that:*
(i) a is continuous;
(ii) $a(x)$ is strictly definite positive for each $x \in \mathbb{R}^d$;
(iii) there exists $K < \infty$ such that for all $1 \leq i, j \leq d$ and all $x \in \mathbb{R}^d$,

$$|a_{ij}(x)| \leq K(1 + \|x\|^2) \quad \text{and} \quad |b_i(x)| \leq K(1 + \|x\|).$$

Then $M(a, b)$ is well-posed.

Proof. T.7.2.1. in Stroock&Varadhan, stated in the form of T.V.24.1 in Rogers&Williams. □

Note that the martingale problem defined above is not the standard martingale problem considered by Stroock–Varadhan (and most of the textbooks) but a simplified version (from Durrett) that is enough for the correspondence argument to carry over. The non-reduced martingale problem (which implies the reduced martingale problem – see lecture notes) involves the infinitesimal generator. Define for all $f \in C_K^2(\mathbb{R}^d)$,

$$Lf(x) = \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x).$$

Under regularity conditions on the coefficients, the operator L emerges as the infinitesimal generator of solutions of (homogeneous) SDEs. That is, $Lf(x)$ represents the infinitesimal expected change in $f(X)$ given that $X_t = x$ defined as

$$\lim_{\varepsilon \to 0} \mathbb{E}\left(\frac{f(X_{t+\varepsilon}) - f(X_t)}{\varepsilon} \,\middle|\, \mathcal{F}_t, X_t = x \right) = Lf(x).$$

Definition 22 ((SV) Martingale Problem)**.** A continuous process with values in $\mathbb{R}^d$ is said to the solve the (SV) martingale problem if for all $f \in C_K^\infty(\mathbb{R}^d)$,

$$M_t^f := f(X_t) - f(X_0) - \int_0^t Lf(X_s)\, ds$$

is a local martingale.

By considering functions $f(x) = x_i$ and $f(x) = x_i x_j$, it can be shown that if X is solution to the (SV) martingale problem for $a = \sigma\sigma^T$, then it is a solution to $M(a, b)$.

6 Diffusion approximation of Markov chains

Lemma 23. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with values in $C([0,1],\mathbb{R}^d)$. Then $(X_n)_{n\in\mathbb{N}}$ is tight if and only if for all $\varepsilon > 0$, there exists $n_0, M \in \mathbb{N}$ and $\delta > 0$ such that:*

(i) $P(\|X^n(0)\| > M) \le \varepsilon$ for all $n \ge n_0$;

(ii) $P(\mathrm{osc}_\delta > \varepsilon) \le \varepsilon$;

where

$$\mathrm{osc}_\delta(\omega) := \sup\{\|\omega(s) - \omega(t)\| : |s - t| \le \delta\}.$$

Similar conditions based on an extension of the Arzelà–Ascoli theorem (using an extended notion of δ-modulus for functions which may have jumps) can be derived for characterizing tightness in $D([0,1],\mathbb{R}^d)$. For our purpose, we note that the previous characterization in C in terms of the standard notion of δ-modulus of continuity provides a sufficient condition for tightness in D. Moreover, if the conditions hold, the subsequential limits are in C a.s.. (See T.15.5. in CPM Billingsley (1968)).

Theorem 24 (Stroock–Varadhan Approximation Theorem)**.** *Let $Y^h := (Y^h_n)_{n\in\mathbb{N}}$ be a rescaled (discrete) Markov chain taking values in a set $S_h \subseteq \mathbb{R}^d$ with $h > 0$ a scaling parameter. The transition probabilities of Y^h are given by a Markov kernel Π_h, that is, $\Pi_h(x, A) = P(Y^h_{n+1} \in A | Y^h_n = x)$. Define the process X^h on $[0,1]$ by*

$$X^h_t := Y^h_{\lfloor t/h \rfloor}$$

so that X^h is a.s. right-continuous and constant between two successive jumps of the chain, which may happen every h units of time for X^h. Define

$$K_h(x, dy) := \frac{1}{h}\Pi_h(x, dy).$$

For $1 \le i, j \le d$, define

$$\begin{aligned} a^h_{ij} &:= \int_{|y-x|\le 1} (y_i - x_i)(y_j - x_j) K_h(x, dy) \\ b^h_i &:= \int_{|y-x|\le 1} (y_i - x_i) K_h(x, dy) \\ \Delta^h_\varepsilon &:= K_h(x, B(x,\varepsilon)^c). \end{aligned}$$

Suppose that:

(i) for every $1 \le i, j \le d$, every $R > 0$, and every $\varepsilon > 0$,

$$\begin{aligned} &\lim_{h\to 0} \sup_{|x|\le R} |a^h_{ij}(x) - a_{ij}(x)| = 0, \\ &\lim_{h\to 0} \sup_{|x|\le R} |b^h_i(x) - b_i(x)| = 0, \\ &\lim_{h\to 0} \sup_{|x|\le R} \Delta^h_\varepsilon(x) = 0, \end{aligned}$$

for some functions a_{ij} and b_i;

(ii) the limit coefficients a_{ij} and b_i are continuous functions on $\mathbb{R}^d$;

(iii) for each $x \in \mathbb{R}^d$, the martingale problem $M(a,b)$ is well-posed, i.e., it has a unique (in distribution) solution X with initial value $X_0 = x$ a.s..

If $X_0^h = x_h \to x_0$ as $h \to 0$ for some $x_0 \in \mathbb{R}^d$, then X^h converges weakly to X as $h \to 0$ in the space $D([0,1],\mathbb{R}^d)$ of cadlag functions from $[0,1]$ to $\mathbb{R}^d$ endowed with the Skorokhod topology, where X is the (unique) solution to the martingale problem $M(a,b)$ with initial value $X_0 = x_0$. In particular, the linear interpolations of Y^h converge weakly in the space $C([0,1],\mathbb{R}^d)$ of continuous functions from $[0,1]$ to $\mathbb{R}^d$ endowed with the supremum norm.

Reference. T.11.2.3. in Stroock&Varadhan (1979) p.272 or T.7.1. in Durrett (1996) p.297 or T.6.5. in Berestycki's lecture notes SCA (2010) p.91.

Proof. The conditions (i) of the theorem ensure that: the infinitesimal mean and variance converges to those of the diffusion limit; there is no jump in the limit. The idea of the proof is a standard application of Prokhorov's theorem. We first prove tightness, then ensure of the uniqueness of subsequential limits. Before that, a localization argument can be used to replace the conditions (i) by the stronger conditions

$$\lim_{h\to 0} \sup_{x\in R^d} |a_{ij}^h(x) - a_{ij}(x)| = 0,$$
$$\lim_{h\to 0} \sup_{x\in R^d} |b_i^h(x) - b_i(x)| = 0,$$
$$\lim_{h\to 0} \sup_{x\in R^d} \Delta_\varepsilon^h(x) = 0,$$

a_{ij}^h, b_i^h, and Δ^h are uniformly bounded in h and x.

We start by proving tightness. Let f be a bounded measurable function and define

$$L^h f(x) := \int (f(y) - f(x))K_h(x,dy).$$

This can be interpreted as the infinitesimal generator of the discrete-time process considered. Since $X_0^h = x_h$ is nonrandom and converges towards a fixed x_0, the condition (i) in Lemma 23 is verified. To prove tightness, we now verify condition (ii). Using Lemma 26, it suffices to prove that for all $x \in \mathbb{R}^d$ and for all δ sufficiently small,

(a) $P_x(\theta > \varepsilon/4) < \varepsilon/2$ as $h \to 0$,

(b) $P_x(\sigma < \delta) < \varepsilon/2$ as $h \to 0$.

To prove (a), note that there are at most $1/h$ time steps in $[0,1]$, so by union bound,

$$P_x(\theta > \varepsilon) \le \frac{1}{h}\sup_y \Pi_h(y, B(y,\varepsilon)^c) \le \sup_y \Delta_\varepsilon^h(y) \to 0.$$

In particular, for h sufficiently small, $P_x(\theta > \varepsilon/4) < \varepsilon/2$.
To prove (b), we start by estimating $P_x(\tau_1 \le u)$ for small u. By Lemma 25,

$$\mathbb{E}\left(f_{x,\varepsilon}(Y_{k+1}^h) - f_{x,\varepsilon}(Y_k^h) - hC_\varepsilon\right) \ge 0,$$

so that

$$f_{x,\varepsilon}(Y_k^h) + C_\varepsilon hk, \quad k = 0,1,\ldots$$

is a submartingale. Set $\tau := \inf\{k \ge 1 : \|Y_k^h - x\| > \varepsilon\}$, so that $\tau_1 = h\tau$. Using the optional stopping theorem at $\tau \wedge u'$ with $u' = u/h$, we get

$$\mathbb{E}_x(f_{x,\varepsilon}(Y^h_{\tau\wedge u'}) + C_\varepsilon h(\tau \wedge u')) \ge \mathbb{E}_x(f_{x,\varepsilon}(Y^h_0)) = 1.$$

On the event $(\tau \le u')$, we have that $\|Y^h_k - x\| \ge \varepsilon$, so that $f_{x,\varepsilon}(Y^h_{\tau\wedge u'}) = 0$, and since $\tau \wedge u' \le u'$, we have

$$P_x(\tau_1 \le u) \le P(\tau \le u') \le \mathbb{E}_x(1 - f_{x,\varepsilon}(Y^h_{\tau\wedge u'})) \le hC_\varepsilon u' = C_\varepsilon u$$

where the first inequality follows from the fact that the term in the expectation is non-negative and is equal to 1 if $\tau \le u'$. Let $p := P_x(\tau_1 \le u)$. For all $u > 0$,

$$\begin{aligned}\mathbb{E}_x(e^{-\tau_1}) &\le P(\tau_1 \le u) + e^{-u}P_x(\tau_1 \ge u)\\ &= p + e^{-u}(1-p) = e^{-u} + p(1-e^{-u})\\ &\le e^{-u} + pu \le 1 - u + C_\varepsilon u^2.\end{aligned}$$

Thus by choosing u small enough, we can find $\lambda < 1$, independent of x and δ, such that $\mathbb{E}_x(e^{-\tau_1}) \le \lambda$. By iterating and using the strong Markov property, we get $\mathbb{E}_x(e^{-\tau_n}) \le \lambda^n$. By Markov's inequality,

$$\begin{aligned}P_x(N > n) = P_x(\tau_n < 1) &\le P_x(e^{-\tau_n} \ge e^{-1})\\ &\le e\mathbb{E}_x(e^{-\tau_n}) \le e\lambda^n.\end{aligned}$$

Observe that for any $k \in \mathbb{N}$,

$$\begin{aligned}P_x(\sigma \le \delta) &\le k \sup_y P_y(\tau_1 \le \delta) + P_x(N > k)\\ &\le kC_\varepsilon\delta + e\lambda^k.\end{aligned}$$

By taking k large enough and δ small enough so that $e\lambda^k < \varepsilon/4$ and $kC_\varepsilon\delta < \varepsilon/4$, we get $P_x(\sigma \le \delta) < \varepsilon/2$. This concludes the proof for tightness.

We now prove the uniqueness of subsequential limits. Since the martingale problem $M(a,b)$ is assumed well-posed, it suffices to show that the limit of any weakly convergent subsequent solves the martingale problem $M(a,b)$. We first show that as $h \to 0$ the operator L^h converges in an appropriate sense to the infinitesimal generator L of the solution

$$Lf(x) = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 f}{\partial x_i x_j}(x) + \sum_{i=1}^d b_i(x)\frac{\partial f}{\partial x_i}(x).$$

More precisely, we show that for any $f \in C^2_K$, $L^h f(x)$ converges uniformly in $x \in \mathbb{R}^d$ to $Lf(x)$ as $h \to 0$. By Taylor's theorem for $h(t) := f(x + t(y-x))$ on $t \in [0,1]$, there exists $c_{xy} \in [0,1]$ such that

$$\begin{aligned}f(y) - f(x) &= h'(0) + h''(c_{xy})/2!\\ &= \sum_i (y_i - x_i)D_i f(x) + \sum_{i,j}(y_i - x_i)(y_j - x_j)D_{ij}f(z_{xy})\end{aligned}$$

where $z_{xy} = x + c_{xy}(y-x)$. Then integrating over $\|y - x\| \le 1$ with respect to $K_h(x, dy)$, we get

$$
\begin{aligned}
L^h f(x) = &\sum_i b_i^h(x) D_i f(x) \\
&+ \int_{\|y-x\|\le 1} \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) K_h(x, dy) \\
&+ \int_{\|y-x\|>1} (f(y) - f(x)) K_h(x, dy).
\end{aligned}
$$

Recalling the definition of Δ_1^h, the final term goes to 0 uniformly in x by assumption (i.3). For the first term, note that

$$
\left| \sum_i b_i^h(x) D_i f(x) - \sum_i b_i(x) D_i f(x) \right| \le \sup_i |b_i^h(x) - b_i(x)| \|D_i f\|_\infty,
$$

which converges to 0 uniformly in x by assumption (i.2) since $f \in C_K^2$. For the intermediate term, note first that

$$
\begin{aligned}
&\left| \int_{\|y-x\|\le 1} \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) K_h(x, dy) - \sum_{i,j} a_{ij}(x) D_{ij} f(x) \right| \\
&\quad \le \left| \sum_{i,j} a_{ij}^h(x) D_{ij} f(x) - \sum_{i,j} a_{ij}(x) D_{ij} f(x) \right| \\
&\quad + \left| \int_{\|y-x\|\le 1} \sum_{i,j} (y_i - x_i)(y_j - x_j)(D_{ij} f(z_{xy}) - D_{ij} f(x)) K_h(x, dy) \right|.
\end{aligned}
$$

The first term converges to 0 uniformly in x by assumption (i.1) since the derivatives of f are uniformly bounded. The second term can be split in an integral over$\|y - x\| > \varepsilon$ and one over $\|y - x\| \le \varepsilon$. The first one converges to 0 uniformly in x by assumption (i.3) since the integrand is bounded. For the second term, define

$$
\Gamma(\varepsilon) := \sup_{i,j} \sup_{\|y-x\|\le\varepsilon} |D_{ij} f(z_{xy}) - D_{ij} f(x)|.
$$

Since z_{xy} lies on the segment between x and y and $D_{ij} f$ is continuous and hence uniformly continuous on compact sets, we have that $\Gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$. By the Cauchy–Schwarz inequality,

$$
\begin{aligned}
&\left| \int_{\|y-x\|\le\varepsilon} \sum_{i,j} (y_i - x_i)(y_j - x_j)(D_{ij} f(z_{xy}) - D_{ij} f(x)) K_h(x, dy) \right| \\
&\qquad \le \Gamma(\varepsilon) \int_{\|y-x\|\le\varepsilon} \|y - x\|^2 K_h(x, dy),
\end{aligned}
$$

which concludes the proof that $\|L^h f - Lf\|_\infty \to 0$. Now, fix a sequence $h_n \to 0$ such that $X^{h_n} \to X$ weakly in D as $n \to \infty$. Fix $s < t$. By definition of L^h,

$$
f(X_{kh_n}^{h_n}) - \sum_{j=0}^{k-1} h_n L^{h_n} f(X_{jh_n}^{h_n}), \quad k = 0, 1, \ldots
$$

is a (discrete-time) martingale. The martingale property implies that for any $\mathcal{F}_s$-measurable function $g\colon D \to \mathbb{R}$,

$$\mathbb{E}_x\left(g(X^{h_n})\left(f(X^{h_n}_{l_n h_n}) - f(X^{h_n}_{k_n h_n}) - \sum_{j=k_n}^{l_n-1} h_n L^{h_n} f(X^{h_n}_{j h_n})\right)\right) = 0$$

where $k_n = \lceil s/h_n \rceil$ and $l_n = \lceil t/h_n \rceil$. Using the Skorokhod representation theorem, there exists Y^n such that $Y^n =_d X^{h_n}$ and $Y^n \to Y$ a.s., where $Y =_d X$. Using that $\|L^h f - Lf\|_\infty \to 0$ and the bounded convergence theorem, it follows by taking the limit as $n \to \infty$ that

$$\mathbb{E}_x\left(g(X)\left(f(X_t) - f(X_s) - \int_s^t Lf(X_u)\,du\right)\right) = 0.$$

Since g is arbitrary, it follows that

$$f(X_t) - \int_0^t Lf(X_u)\,du\,, \quad t \geq 0$$

is a martingale for all $f \in C^2_K$. Hence X is a solution to the (SV) martingale problem, and hence to $M(a,b)$. Since $M(a,b)$ has a unique solution, this concludes the proof. □

Lemma 25. *Let $g\colon \mathbb{R} \to \mathbb{R}$ such that $g \in C^2$, $0 \leq g \leq 1$, $g(x) = 0$ if $x \geq 1$, $g(0) = 1$. Define for $x \in \mathbb{R}^d$, $f_\varepsilon(x) := g(\|x\|^2/\varepsilon^2)$. Define for $a \in \mathbb{R}^d$, $f_{a,\varepsilon}(x) := f_\varepsilon(x-a)$. Then there exists $C_\varepsilon < \infty$, independent of h, such that $\|L^h f_{a,\varepsilon}(x)\| \leq C_\varepsilon$ for all $a, x \in \mathbb{R}^d$.*

Proof. By Taylor's theorem for $\phi(t) := f_{a,\varepsilon}(x + t(y-x))$ on $t \in [0,1]$, there exists $c_{xy} \in [0,1]$ such that

$$\begin{aligned} f_{a,\varepsilon}(y) - f_{a,\varepsilon}(x) &= \phi'(0) + \phi''(c_{xy})/2! \\ &= \sum_i (y_i - x_i) D_i f(x) + \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f(z_{xy}) \end{aligned}$$

where $z_{xy} = x + c_{xy}(y-x)$. Then integrating with respect to $K_h(x,dy)$, we get

$$\begin{aligned} L^h f_{a,\varepsilon}(x) &= \int (f_{a,\varepsilon}(y) - f_{a,\varepsilon}(x)) K_h(x,dy) \\ &\leq \left|\nabla f_{a,\varepsilon}(x) \cdot \int_{\|y-x\|\leq 1} (y-x) K_h(x,dy)\right| \\ &\quad + \left|\int_{\|y-x\|\leq 1} \sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{a,\varepsilon}(z_{xy}) K_h(x,dy)\right| \\ &\quad + 2\|f_{a,\varepsilon}\|_\infty K_h(x, B(x,1)^c). \end{aligned}$$

Define $A_\varepsilon := \sup_x \|\nabla f_{a,\varepsilon}(x)\|$ and $B_\varepsilon := \sup_y \|D f_{a,\varepsilon}(y)\|$. We have

$$\left|\sum_{i,j} (y_i - x_i)(y_j - x_j) D_{ij} f_{a,\varepsilon}(z_{xy})\right| \leq \|y - x\|^2 B_\varepsilon,$$

hence

$$L^h f_{a,\varepsilon}(x) \le A_\varepsilon \| |b^h(x)\| + B_\varepsilon \int_{\|y-x\|\le 1} \|y-x\|^2 K_h(x,dy) + 2K_h(x, B(x,1)^c).$$

Since $\int_{\|y-x\|\le 1} \|y-x\|^2 K_h(x,dy) = \sum_i a^h_{ii}(x)$, the uniform boundedness of the stronger assumptions yields the result. □

Lemma 26. *Define the random variables*

$$\begin{aligned} \tau_0 = 0, \quad \tau_n &:= \inf\{t \ge \tau_{n-1} : \|X^h_t - X^h_{\tau_{n-1}}\| \ge \varepsilon\}, \\ N &:= \min\{n : \tau_n > 1\}, \\ \sigma &:= \min\{\tau_n - \tau_{n-1} : 1 \le n \le N\}, \\ \theta &:= \max\{\|X^h(t) - X^h(t^-)\| : 0 < t \le 1\}. \end{aligned}$$

If $\sigma > \delta$ and $\theta < \varepsilon$, then $\operatorname{osc}_\delta(X^h) \le 4\varepsilon$.

Proof. By definition of $\operatorname{osc}_\delta$, we need to show that for all $s, t \in [0,1]$ with $|s-t| \le \delta$, $|X^h(s) - X^h(s)| \le 4\varepsilon$. Since $|s-t| \le \delta < \sigma$, s and t can only span at most one of the intervals $[\tau_{n-1}, \tau_n]$, and the definition of the stopping times τ_n ensures that X^h does not vary by more than 2ε on such an interval. More precisely: if $\tau_{n-1} \le s < t < \tau_n$, then $|X^h(s) - X^h(t)| \le 2\varepsilon$; if $\tau_{n-1} \le s < \tau_n \le t$, then

$$\begin{aligned} |X^h(s) - X^h(t)| \le& |X^h(s) - X^h(\tau_{n-1})| + |X^h(t) - X^h(\tau_n)| \\ &+ |X^h(\tau_n) - X^h(\tau_n^-)| + |X^h(\tau_n^-) - X^h(\tau_{n-1})| \\ \le& 4\varepsilon. \end{aligned}$$

□

Example 27 (Ehrenfest Chain)**.** Two urns contain a total of $2n$ balls. At each time m, we pick one ball uniformly at random from the $2n$ balls and move it to the other urn. We expect the number of balls in the left urn to be about $n + C\sqrt{n}$, so we define Z_m the number of balls in the left urn and consider the rescaled process

$$Y^{1/n}_m := \frac{Z_m - n}{\sqrt{n}}.$$

Suppose that $Z_0 = n$. Define

$$X^{1/n}_t := Y^{1/n}_{\lfloor tn \rfloor} = \frac{Z_{\lfloor tn \rfloor} - n}{\sqrt{n}}.$$

Then, as $n \to \infty$, the process $X^{1/n}$ converges weakly to an Ornstein–Uhlenbeck diffusion process X with unit viscosity, that is, the pathwise unique solution to the SDE given by

$$dX_t = -X_t\,dt + d\,B_t\,, \quad X_0 = 0.$$

To see this, note first that the space of $Y^{1/n}$ is $E_n = \{k/\sqrt{n} : -n \le k \le n\}$ and that the transition kernel $K_{1/n}(x, dy)$ is given by $n\Pi_{1/n}(x, dy)$ where

$$\Pi_{1/n}(x, x + n^{-1/2}) = \frac{n - x\sqrt{n}}{2n}, \quad \Pi_{1/n}(x, x - n^{-1/2}) = \frac{n + x\sqrt{n}}{2n}.$$

Indeed, if $Y_m^{1/n} = x$, then the number Z_m of particles in the first container is $n+x\sqrt{n}$, and if one particle is chosen, which happens with probability $(n+x\sqrt{n})/2n$, then $Z_{m+1} = Z_m - 1$ and $Y_{m+1} = x - n^{-1/2}$. Since the chain can only make jumps of size $1/\sqrt{n}$, then condition (i.3) holds trivially. Moreover, we have

$$b^{1/n}(x) = \int (y - x)K_{1/n}(x, dy) = n\left(\frac{1}{\sqrt{n}}\frac{n - x\sqrt{n}}{2n} - \frac{1}{\sqrt{n}}\frac{n + x\sqrt{n}}{2n}\right) = -x,$$

and

$$a^{1/n}(x) = \int (y - x)^2 K_{1/n}(x, dy) = n\left(\frac{1}{n}\frac{n - x\sqrt{n}}{2n} + \frac{1}{n}\frac{n + x\sqrt{n}}{2n}\right) = 1,$$

hence the coefficients are independent of n and so satisfy the conditions (i.1-2). Moreover, the limiting coefficients $b(x) = -x$ and $a(x) = 1$ are Lipschitz continuous and so the martingale problem is well-posed. The result then follows directly from Theorem 24.

Example 28 (Branching Processes)**.** Consider a branching process where Z_m denotes the size of the population at time m. At each period m, each individual in the population has an independent and identically distributed number of offspring in the next period $m + 1$. Suppose the probability of having k children is p_k with mean $1 + \beta_n/n$ and variance σ_n^2. Define the rescaled variable

$$Y_m^{1/n} := \frac{Z_m}{n}$$

and consider the process $X^{1/n}$ given by

$$X_t^{1/n} := Y_{\lfloor tn \rfloor}^{1/n} = \frac{Z_{\lfloor tn \rfloor}}{n}.$$

Suppose that as $n \to \infty$,

$$\begin{aligned}
&\beta_n \to \beta \in (-\infty, \infty), \\
&\sigma_n \to \sigma \in (0, \infty), \\
&\text{for any } \delta > 0, \sum_{k > \delta n} k^2 p_k^n \to 0.
\end{aligned}$$

Then $X^{1/n}$ converges weakly to Feller's branching diffusion X, that is, the solution to the SDE given by

$$dX_t = \beta X_t \, dt + \sigma \sqrt{X_t} \, dB_t.$$

References

Barbour, A. D. (1990): "Stein's method for diffusion approximations," *Probability theory and related fields*, 84(3), 297–322.

Berestycki, N. (2010): "Stochastic Calculus," *Lecture notes at Cambridge*.

Billingsley, P. (1968): *Convergence of Probability Measures*.

Durrett, R. (1996): *Stochastic calculus: a practical introduction*, vol. 6. CRC press.

Etheridge, A. (2018): "Continuous martingales and stochastic calculus," *Lecture notes at Oxford*.

Ethier, S. N., and T. G. Kurtz (2009): *Markov processes: characterization and convergence*. John Wiley & Sons.

Feigin, P. D. (1976): "Maximum likelihood estimation for continuous-time stochastic processes," *Advances in Applied Probability*, 8(4), 712–736.

Karatzas, I., and S. Shreve (1991): *Brownian motion and stochastic calculus*, vol. 113. Springer Science & Business Media.

Lalley, S. (2016): "Brownian Motion and Stochastic Calculus," *Lecture notes for Statistics 385 at UChicago*.

Le Gall, J.-F. (2016): *Brownian motion, martingales, and stochastic calculus*. Springer.

Revuz, D., and M. Yor (2013): *Continuous martingales and Brownian motion*, vol. 293. Springer Science & Business Media.

Rogers, L. C. G., and D. Williams (2000): *Diffusions, Markov processes and martingales*. Cambridge University Press.

Stein, C. (1986): "Approximate computation of expectations," IMS.

Stroock, D. W., and S. S. Varadhan (1997): *Multidimensional diffusion processes*, vol. 233. Springer Science & Business Media.