Stat205B: Probability Theory (Spring 2003) **Lecture: 27**

Feller Processes and Semigroups

Lecturer: Rui Dong *Scribe: Rui Dong* ruidong@stat.berkeley.edu

For convenience, we can have a look at the list of materials contained in this lecture first. Note here we always consider the time-homogenous Markov processes.

- Transition kernels $\mu_t, t \geq 0 \longleftrightarrow$ transition operators semigroup $T_t, t \geq 0$.
- Feller semigroup $T_t, t \geq 0$
- Feller processes
  - existence
  - every Feller process has a cadlag version
    - regularization theorem for submartingale
  - strong Markov property
  - Blumental 0-1 law
  - Lévy processes
- Generator of Feller semigroup
  - existence
  - Dynkin's formula
    - extended generator for Markov process
  - generator of Lévy process
    - three basic building blocks
  - generator of Feller process
  - application (heat equation)

Now we start from the definition of the transition operator T_t. For any Markov transition kernels $\mu_t(\cdot, \cdot)$ on $(S, \mathcal{S})$, $f : S \to \mathbb{R}$ bounded or nonnegative, define **transition operator** T_t:

$$T_t f(x) := \int \mu_t(x, dy) f(y)$$

Notice

(i) It's a **positive contraction operator** i.e. $0 \leq f \leq 1$ implies $0 \leq T_t f \leq 1$,

(ii) The **identity operator** I corresponds to kernels $\mu(x, \cdot) = \delta_x$

The following lemma shows you the connection between Markov property and semigroup property:

Lemma 27.1 (semigroup property). Probability kernels $\mu_t, t \geq 0$ satisfy C-K relation iff the corresponding operators $T_t, t \geq 0$ have the semigroup property;

$$T_{s+t} = T_s T_t \qquad s, t \geq 0$$

Proof. Recall C-K equation: $\mu_t \mu_s = \mu_{s+t}$, i.e. for any bounded or nonnegative f, $\int \mu_t(x, dy) \int \mu_s(y, dz) f(z) = \int \mu_{s+t}(x, dz) f(z)$.

For any $B \in \mathcal{S}$,

$$\begin{aligned}
(T_t T_s) 1_B(x) &= \int \mu_s(x, dy) \mu_t(y, B) \\
&= \mu_{s+t}(x, B) \qquad\qquad C - K \\
&= T_{s+t} 1_B(x)
\end{aligned}$$

□

Now, Markov property can be equivalent with semigroup property. Naturally, we have two questions;

- Add what kind of properties to the semigroup can we get Strong Markov property?
- Add what kind of properties to the semigroup can we guarantee the cadlag modification?

because without these two things, it's hard to go further. There is an example which is a continuous Markov process but not a Strong Markov process:

Example 27.2 (a continuous Markov process without Strong Markov property). $(B_t, t \geq 0)$ is a Brownian motion not necessarily starting from 0. Let

$$X_t = B_t 1_{B_0 \neq 0} = \begin{cases} B_t & \text{if } B_0 \neq 0 \\ 0 & \text{if } B_0 = 0 \end{cases}$$

whose transition kernels are

$$\mu_t(x, dy) = \begin{cases} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} dy & \text{if } x \neq 0 \\ \delta_0 dy & \text{if } x = 0 \end{cases}$$

It's a Markov process because for any Borel set B,

$$\begin{aligned}
E[1_B(X_{t+s}) | \mathcal{F}_s] &= E[1_B(X_{t+s}) \cdot 1_{B_0 \neq 0} | \mathcal{F}_s] + E[1_B(X_{t+s}) \cdot 1_{B_0 = 0} | \mathcal{F}_s] \\
&= 1_{B_0 \neq 0} \cdot \int_B \frac{1}{\sqrt{2\pi t}} e^{-\frac{(X_s - y)^2}{2t}} dy + 1_{B_0 = 0} \cdot 1_B(0) \\
&= 1_{X_s \neq 0} \cdot \int_B \frac{1}{\sqrt{2\pi t}} e^{-\frac{(X_s - y)^2}{2t}} dy + 1_{X_s = 0} \cdot 1_B(X_s) + 1_{(B_0 \neq 0, X_s = 0)} \cdot (\int_B \frac{1}{\sqrt{2\pi t}} e^{-\frac{(X_s - y)^2}{2t}} dy - 1_B(0)) \\
&= E 1_B(X_s)
\end{aligned}$$

where we use $\{B_0 \neq 0\} = \{B_0 \neq 0\, X_s \neq 0\} + \{B_0 \neq 0, X_s = 0\}$, $\{X_s = 0\} = \{B_0 \neq 0\, X_s = 0\} + \{B_0 = 0, X_s = 0\}$, $\{B_0 = 0, X_s = 0\} = \{B_0 = 0\}$, $\{B_0 \neq 0, X_s \neq 0\} = \{X_s \neq 0\}$ and $1_{(B_0 \neq 0, X_s = 0)} = 0$ a.s.

While it's not a strong Markov process because if we look $\tau = \inf t > 0, X_t = 0$, then $\forall\, x > 0$,

$$0 = P_x(X_1 \neq 0, \tau \leq 1) = P_x(\tau \leq 1) > 0$$

by $P_x(X_1 = 0) = 0$.

And you will see among the two conditions required for Feller semigroup, here this example doesn't satisfy (F_1).

In fact, (F_2) is guaranteed by right continuous path; for (F_1), if we let $f(x)$ to be a pseudo-indicator, i.e. be 1 on $[a, b]$, be 0 on $(-\infty, 0] \cup [a+b, \infty)$, and be connected continuously on the gaps, here $a, b > 0$. Then by definition, $T_t f(x) = E_x f(X_t)$, notice when $x \neq 0$, $X_t \sim N(x, t)$, so $T_t f(x)$ changes continuously and reaches maximum at $x = (a+b)/2$; but when $x = 0$, $T_t f(0) = E_0 f(X_t) = f(0) = 0$. Thus $T_t f(x)$ is not continuous (at 0).

The answer for both questions is Feller semigroup:

Definition 27.3 (Feller semigroup). Let S to be a locally compact, separable metric space; $C_0 := C_0(S)$ to be all the continuous functions $f : S \to \mathbb{R}$ and $f(x) \to 0$ when $x \to \infty$.(Notice C_0 is a Banach space given $||f|| = \sup_x |f(x)|$.)

A semigroup of positive contraction linear operator $T_t, t \geq 0$ on C_0 is called **Feller semigroup** if it has the following regularity conditions:

(F_1) $T_t C_0 \subset C_0$, $t \geq 0$;

(F_2) $T_t f(x) \to f(x)$, $t \downarrow 0$, $\forall\ f \in C_0$, $x \in S$.

Remark. In fact, $(F_1) + (F_2)$ + semigroup property $\Rightarrow$ (F_3) $||T_t f - f|| \to 0$, $t \downarrow 0$, $\forall\ f \in C_0$(**strong continuity**).

• How to find a nice Markov process associated with a Feller semigroup?

In order to get probability kernels, we may need (T_t) to be conservative.

Definition 27.4 (conservative). T is **conservative** if $\forall\ x \in S$, $\sup_{f \leq 1} Tf(x) = 1$

But in many cases, the (T_t) we know may not be conservative, for example, the particles may die out or disappear suddenly, while the following compactification may help:

$\hat{S} := S \cup \{\Delta\}$ is the one-point compactification of S. Δ is usually called "cemetry", it can be treated as point at infinite. $\hat{C} := C(\hat{S})$ is the set of continuous functions on $\hat{S}$.

Remark. $\forall\ f \in C_0$, f has a continuous extension to $\hat{S}$ by letting $f(\Delta) = 0$.

Now, we can define a new Feller semigroup $\hat{T}_t, t \geq 0$ on $\hat{C}$ which is conservative.

Lemma 27.5 (compactification). Any Feller semigroup $T_t, t \geq 0$ on C_0 admits an extension to a conservative Feller semigroup $\hat{T}_t, t \geq 0$ on $\hat{C}$ by letting

$$\hat{T}_t f := f(\Delta) + T_t(f - f(\Delta)) \qquad t \geq 0, f \in \hat{C}$$

Proof. The only point needing some proof is the positivity: for any $f \in \hat{C} \geq 0$, let $g := f(\Delta) - f$. Since $g \leq f(\Delta)$, we have

$$T_t g \leq T_t g^+ \leq ||g^+|| \leq f(\Delta)$$

so $\hat{T}_t f = f(\Delta) - T_t g \geq 0$.

To see the contraction and conservation, just notice $\hat{T}1 = 1$. To verify the Feller semigroup property:

(F_1): $\forall\ f \in \hat{C}$, $\hat{T}_t f = f(\Delta) + T_t(f - f(\Delta)) \in \hat{C}$, since $T_t g \in C_0$;

(F_2): $||\hat{T}_t f - f|| = ||T_t g - g|| \to 0$, $t \downarrow 0$, since $g \in C_0$. □

Remark. This extension is consistent, i.e. $\forall\ f \in C_0$, $\hat{T}_t f = T_t f$.

Next step, we want to construct an associated semigroup of Markov transition kernels μ_t on $\hat{S}$ satisfying

$$T_t f(x) = \int f(y)\mu_t(x, dy) \qquad \forall\ f \in C_0 \tag{$*$}$$

Theorem 27.6 (existence of Markov kernels). For any Feller semigroup $T_t, t \geq 0$ on C_0, there exists a unique semigroup of Markov transition kernels $\mu_t, t \geq 0$ on $\hat{S}$ satisfying $(*)$, and s.t. Δ is absorbing for $\mu_t, t \geq 0$.

Proof. $\forall\ x \in S$, $t \geq 0$, $f \longrightarrow \hat{T}_t f(x)$ is a positive linear functional on $\hat{C}$, norm 1. By Reisz's representation theorem, there exist probability measures $\mu_t(x, \cdot)$ on $\hat{S}$ satisfying

$$\hat{T}_t f(x) = \int \mu_t(x, dy) f(y) \qquad f \in \hat{C},\ x \in \hat{S},\ t \geq 0$$

Notice $x \longrightarrow \int \mu_t(x, dy) f(y) \in \hat{C}$, hence measurable. By DCT(or MCT) we can show, for any B measurable, $\mu_t(x, B)$ is measurable. So together with the semigroup property, we can say $\mu_t(\cdot, \cdot)$ are Markov transition kernels. $(*)$ follows from the definition of μ_t. And $\forall\ f \in C_0$,

$$\int f(y)\mu_t(\Delta, dy) = \hat{T}_t f(\Delta) = f(\Delta) = 0$$

so Δ is absorbing. Uniqueness also follows from definition. □

By Kolmogrov's existence theorem, for any probability measur π on $\hat{S}$, there exists a Markov process (X_t) in $\hat{S}$ with initial distribution π and transition kernels μ_t.

Definition 27.7 (Feller process). A Markov process associated by a Feller semigroup transition operators is called a **Feller semigroup**.

Now, we come to show any Feller process has a cadlag version.

Theorem 27.8 (regularization for Feller process). Let (X_t) be a Feller process in $\hat{S}$ with arbitrary initial distribution π. Then (X_t) has a cadlag version $(\tilde{X}_t)$.

And $X_t = \Delta$ or $X_{t-} = \Delta$ implies $\tilde{X}_t \equiv \Delta$ on $[t, \infty)$. If $T_t, t \geq 0$ is conservative and $\pi(S) = 1$, then $(\tilde{X}_t)$ can be shown to be cadlag in S.

Remark. The first sentence in the theorem means for any initial π, there exists a cadlag process $(\tilde{X}_t)$ on $\hat{S}$, s.t. $X_t = \tilde{X}_t$ a.s.-P_π, $\forall\ t \geq 0$.

The idea of the proof is, first we have regularization theorem for submartingales, then we can find a large class of continuous functions of Feller process which are all supermartingales, thus we can apply the theorem for submartingales to get the result.

Lemma 27.9 (regularization for submartingales). If $(X_t, t \geq 0)$ is a submartingale, then for a.e. ω, for each $t > 0$, $\lim_{r\uparrow t, r\in\mathbb{Q}} X_r(\omega)$ exists and for each $t \geq 0$, $\lim_{r\downarrow t, r\in\mathbb{Q}} X_r(\omega)$ exists, i.e. the right and left limits exist along $\mathbb{Q}$.

Proof. The crucial idea is to use **upcrossing inequality** to control the time of upcrossing. The theorem says, if (X_t) is a submartingale, T is a countable index set, for any $a < b$, define the upcrossing time on T by

$$U_T := \sup_{F finite, F\subset T} \{U_F, \text{time of upcrossing along F}\}$$

then

$$(b-a)EU_T \le \sup_{t\in T} E(X_t - a)^+$$

Now, to prove the lemma, it's enough to show $\forall\ t \in I$ the lemma is true, where I is a compact subset of T(here take $T = \mathbb{Q}$), t_d is the right end of I. $\forall\ t \in I$, $X_t \le E(X_{t_d}|\mathcal{F}_{t_d}) \le E(X^+_{t_d}|\mathcal{F}_{t_d})$, so $X^+_t \le E(X^+_{t_d}|\mathcal{F}_{t_d})$ i.e. $EX^+_t \le EX^+_{t_d}$, which implies

$$E(X_t - a)^+ \le EX^+_t + a^- \le EX^+_{t_d} + a^- < \infty$$

i.e.

$$EU_{I\cap T} \le \sup_{t\in I\cap T} (X_t - a)^+ < \infty$$

We have $U_{I\cap T} < \infty$ a.s.

This shows you the right and left limit at t must exist along $\mathbb{Q}$, otherwise, take a, b to be the different limits of different sequences, then the upcrossing times will be infinite. □

To find a large class of continuous functions of Feller process which are supermartingales, we need

Definition 27.10 (resolvent). $\forall\ \lambda > 0$, **resolvent** R_λ is defined as

$$R_\lambda f(x) := \int_0^\infty e^{-\lambda t} T_t f(x) dt$$

Remark. $\forall\ t, \lambda > 0$, $T_t R_\lambda = R_\lambda T_t$ and $||\lambda R_\lambda f - f|| \to 0$, $\lambda \to \infty$.

Lemma 27.11 (resolvents). If $f \in C^+_0$, $(X_t, t \ge 0)$ is a Feller process, then the process $Y^\lambda_t := e^{-\lambda t} R_\lambda f(X_t), t \ge 0$ is a supermartingale for any $\lambda > 0$ under P_π, $\forall\ \pi$.

Remark. In fact, repeat the proof here, you can see the lemma is true for any nonnegative measurable f.

Proof. Denote the filtration induced by (X_t) as $(\mathcal{G}_t)$, then $\forall\ t, h \ge 0$,

$$\begin{aligned} E(Y^\lambda_{t+h}) &= E(e^{-(t+h)\lambda} R_\lambda f(X_{t+h})|\mathcal{G}_t) \\ &= e^{-(t+h)\lambda} T_h R_\lambda f(X_t) \\ &= e^{-(t+h)\lambda} R_\lambda T_h f(X_t) \\ &= e^{-(t+h)\lambda} \int_0^\infty e^{-\lambda s} T_{s+h} f(X_t) ds \\ &= e^{-\lambda t} \int_h^\infty e^{-\lambda s} T_s f(X_t) ds \\ &\le Y^\lambda_t. \end{aligned}$$

□

Proof of regularization of Feller processes. Let's fix any initial π first. By the separability of C^+_0, we can choose (f_n) to be a sequence in C^+_0 which separate points, i.e. $\forall\ x \ne y$, $x, y \in \hat{S}$, there exists n s.t. $f_n(x) \ne f_n(y)$. Since $||\lambda R_\lambda f - f|| \to 0$ uniformly when $\lambda \to \infty$, we have the countable set $\mathcal{H} := \{R_\lambda f_n : \lambda \in \mathbb{N}, n \in \mathbb{N}\}$ also separates points and $\mathcal{H} \subset C^+_0$.

$\forall\ h \in \mathcal{H}$, by the lemma, $h(X_t)$ has right limit along $\mathbb{Q}$ for a.e. ω. Since $\mathcal{H}$ is countable, we can choose the null set to be independent with the choice of h. Now, the statement is, for a.e. ω, $h(X_t(\omega))$ has right limit along

$\mathbb{Q}$ for every $h \in \mathcal{H}$. This can imply that for a.e. ω, $X_t(\omega)$ has right limit along $\mathbb{Q}$. If not, you can choose a, b to be the different limits of different sequences from the right, then choose $h \in \mathcal{H}$ which separates a, b, thus we can get two different limits $h(a)$ and $h(b)$ of two different sequences for $h(X_t)$, which is a contradiction with the existence of right limit for $h(X_t)$.

Now, set $\tilde{X}_t(\omega) := \lim_{r\downarrow t, r\in\mathbb{Q}} X_r(\omega)$ for those ω the right limit exists; $\tilde{X}_t(\omega) \equiv 0$ for those ω the right limit doesn't exist.

I claim that $\forall\ t$, $\tilde{X}_t = X_t$ a.s. To prove this, take any bounded $g, h \in \hat{C}$,

$$\begin{aligned} E_\pi(g(X_t)h(\tilde{X}_t)) &= \lim_{s\downarrow t, s\in\mathbb{Q}} E_\pi(g(X_t)h(X_s)) \\ &= \lim_{s\downarrow t, s\in\mathbb{Q}} E_\pi(g(X_t)T_{s-t}h(X_t)) \qquad \text{(condition on } \mathcal{F}_t) \\ &= E_{pi}(g(X_t)h(X_t)). \end{aligned}$$

So, for any positive Borel function $f(x, y)$ on $\hat{S} \times \hat{S}$, we have $E_\pi f(X_t, \tilde{X}_t) = E_\pi f(X_t, X_t)$. By the property of $\hat{S}$, $1_{x\neq y}$ is such a function, so $\tilde{X}_t = X_t$ a.s. for any t.

Now, $\forall\ h \in \mathcal{H}$, use the lemma again, we have for a.e. ω, $h(\tilde{X}_t)$ is right continuous and has left limits along $\mathbb{Q}$ for every $h \in \mathcal{H}$ because $e^{-\lambda t}R_\lambda f(\tilde{X}_t)$ is again right continuous supermartingale. This implies for a.e. ω, the $\tilde{X}^t(\omega)$ is right continuous and has the left limit along $\mathbb{Q}$. By some soft calculus arguments, $\tilde{X}_t(\omega)$ must be cadlag for a.e. ω.

To see those two remainders, notice when X_t or $X_{t-} = \Delta$, those supermartingales must be 0 after then, which means X_t has to be Δ after then, so is $\tilde{X}_t$. If $T_t, t \geq 0$ are all conservative, $\pi(S) = 1$, then $X_t \in S$ for any t, so is $\tilde{X}_t$, and all the regularity doesn't change. □

Now, we can take Ω to be space of all $\hat{S}$ valued cadlag function s.t. Δ is absorbing. Under any P_π, $(X)_t$ is a Markov process with initial π, transition kernels μ_t, and has cadlag path. Particularly, $X \equiv \Delta$ on $[\zeta, \infty]$, where

$$\zeta := \inf\{t \geq 0 : X_t = \Delta \text{or} X_{t-} = \Delta\}$$

Take $(\mathcal{F}_t)$ to be right continuous filtration induced by (X_t), shift operators θ_t.

Definition 27.12 (canonical Feller process). Process $(X_t, t \geq 0)$ with distribution P_π, filtration $(\mathcal{F}_t)$, shift operators θ_t, and cadlag path is called **the canonical Feller process** with semigroup $(T_t, t \geq 0)$.

To show the strong Markov property, we use the similar arguments as for Brownian motion.

Theorem 27.13 (strong Markov property). For any canonical Feller process $(X_t, t \geq 0)$ with initial π, stopping time τ, and r.v. Y,

$$E_\pi(Y \circ \theta_\tau | \mathcal{F}_\tau) = E_{X_\tau} Y \qquad \text{a.s.-}P_\pi \text{ on } \{\tau < \infty\}$$

Proof. Let $(\mathcal{G}_t)$ be the filtration induced by (X_t). Let

$$\tau_n := \frac{[2^n\tau] + 1}{2^n}$$

then, all τ_n are $\mathcal{G}$-stopping time and $\mathcal{F}_\tau \subset \mathcal{G}_{\tau_n}$ for any n. Notice τ_n only takes countably many value, so we have strong Markov property for each τ_n, i.e.

$$E_\pi(Y \circ \theta_{\tau_n}; A) = E_\pi(E_{X_{\tau_n}}; A) \qquad \forall\ A \in \mathcal{F}_\tau, n \in \mathbb{N}$$

To extend the property to τ, it's enough to see $Y = f_1(X_{t_1})f_2(X_{t_2})\cdots f_m(X_{t_m})$, where $f_1, f_2, \dots, f_m \in C_0$, $t_1 < t_2 < \cdots < t_m$. Then when $n \to \infty$, by the right continuity, the left hand side

$$Y \circ \theta_{\tau_n} \to Y \circ \theta_\tau$$

To see the right hand side, let $h_k = t_k - t_{k-1}$, $t_0 = 0$, then

$$\begin{aligned} E_{X_{\tau_n}} Y &= T_{h_1}(f_1 T_{h_2}(\cdots(f_{m-1}T_{h_m}f_m)\cdots))(X_{\tau_n}) \\ &= T_{h_1}(f_1 T_{h_2}(\cdots(f_{m-1}T_{h_m}f_m)\cdots))(X_\tau) \\ &= E_{X_\tau} Y \end{aligned}$$

Then use DCT in both sides, DONE. □

Similarly, we have

Theorem 27.14 (Blumental 0-1 law). For any canonical Feller process, we have

$$P_x A = 0 \text{ or } 1 \qquad \forall\, x \in S, A \in \mathcal{F}_0$$

Proof. Let $\tau = 0$ in the last theorem, we get immediately

$$1_A = P_x(A|\mathcal{F}_0) = P_{X_0}A = P_x A \qquad \text{a.s.-}P_x.$$

□

Remark. Notice by the definition, $\mathcal{F}_0 := \mathcal{G}_{0+}$, so for any $\mathcal{F}$ stopping time τ,

$$P_x(\tau = 0) = 0 \text{ or } 1.$$

Now, let's have a look at Lévy processes. We'll see, Lévy process in fact is the Feller process generated by the convolution semigroup.

A convolution semigroup is a family of probability measures on $\mathbb{R}_d$ s.t.

(i) $\pi_t * \pi_s = \pi_{s+t}$, $\forall\ s, t \geq 0$;

(ii) $\pi_0 = \delta_0$ and $\lim_{t\downarrow 0} = \delta_0$ in vague topology, i.e. $\pi_t f \to f(0)$, $\forall$ continuous f with compact support.

Then the transition kernels and operators are:

$$\mu_t(x, A) = \int_{\mathbb{R}_d} 1_A(x+y)\pi_t(dy)$$

$$T_t f(x) = \int_{\mathbb{R}_d} f(x+y)\pi_t(dy).$$

It's easy to check (T_t) is a Feller semigroup.

Theorem 27.15. If transition kernels of (X_t) is given by a convolution semigroup, then (X_t) has stationary independent increments. The law of $X_t - X_s$ is π_{t-s}.

Proof. If starts at x,

$$E_x f(X_t - X_0) = E_x f(X_t - x) = \pi_t f \qquad \text{independent with } x$$

$$E_\pi(f(X_t - X_s)|\mathcal{F}_s) = E_{X_s} f(X_{t-s} - X_0) = \pi_{t-s} f \qquad P_\pi\text{-a.s.}$$

□

It's also clear that if a Feller process has stationary independent increments, then its transition kernels are given by a convolution semigroup, because

$$\mu_{t+s}(A) = P(X_{t+s} \in A) = P(X_{t+s} - X_t + X_t \in A) = \mu_s * \mu_t(A)$$

and

$$T_t f(x) = \int_{\mathbb{R}_d} f(x+y)\mu_t(dy)$$

$$\pi_t \to \delta_0 \Longleftrightarrow T_t f(x) \to f(x), \qquad t \downarrow 0.$$

Definition 27.16 (L*é*vy process). L*é*vy process is Feller process with stationary independent increments.

The general relations are:

$$\{\text{Markov processes}\} = \{\text{strong Markov processes}\} \cup \{\text{cadlag Markov processes}\} \cup \{\text{other Markov processes}\}$$

$$\{\text{Feller processes}\} \subset \{\text{strong Markov processes}\} \cap \{\text{cadlag Markov processes}\}$$

$$\{\text{diffusion}\} = \{\text{strong Markov processes}\} \cap \{\text{continuous Markov processes}\}$$

$$\{\text{Feller diffusion}\} = \{\text{Feller processes}\} \cap \{\text{continuous Markov processes}\}$$

$$\{\text{Lévy processes}\} \subset \{\text{Feller processes}\}$$

$$\{\text{Brownian motion}\} \subset \{\text{Lévy processes}\} \cap \{\text{continuous Markov processes}\} \subset \{\text{Feller diffusion}\} \subset \{\text{diffusion}\}.$$

We've seen how to construct a Markov process based on transition operators, but the problem is there aren't many transition operators which are explicitly known; moreover, what's known in most cases is the way in which the process moves from point to point. So we define

Definition 27.17 (generator). (X_t) is a Feller process, a function f in C_0 is said to belong to the **domain** $\mathcal{D}_A$ of the infinitesimal generator of X_t if the limit

$$Af := \lim_{t \downarrow 0} \frac{T_t f - f}{t}$$

exists in C_0. The operator $A : \mathcal{D}_A \to C_0$ thus defined is called the **infinitesimal generator** of the process (X_t) or of the semigroup (T_t).

Remark. To see the meaning of A, let f bounded and $f \in \mathcal{D}_A$,

$$E(f(X_{t+h}) - f(X_t)|\mathcal{F}_t) = T_h(X_t) - f(X_t) = hAf(X_t) + o(h)$$

Thus A appears as describing how the process moves from point to point in an infinitesimal small time interval.

The first thing we should mind is the **existence**. Let's first see the motivation of proving the existence:

Motivation: If $p_t = e^{ta}$, in order to get the value of a, there're two methods, by either differentiation:

$$\frac{p_t - 1}{t} \to a, \qquad t \downarrow 0$$

or integration

$$\int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda - a}, \qquad \lambda > 0.$$

Motivated by the later formula, introduce **resolvent** R_λ, $\forall\ \lambda > 0$,

$$R_\lambda f(x) := \int_0^\infty e^{-\lambda t} T_t f(x) dt, \qquad f \in C_0$$

this definition makes sense since $\forall\ x$, $T_t f(x)$ is bounded and right continuous on $[0, \infty)$.

Theorem 27.18 (resolvent and generator). (T_t) is a Feller semigroup on C_0 with resolvent R_λ, $\lambda > 0$. Then λR_λ are **injective contractions** on C_0, $||\lambda R_\lambda f - f|| \to 0$, $\lambda \to \infty$.

The range $\mathcal{D}_A := R_\lambda C_0$ is independent of λ and dense in C_0. There exists and operator A on C_0 with domain $\mathcal{D}_A$ s.t. $R_\lambda^{-1} = \lambda - A$ on $\mathcal{D}_A$, $\forall\ \lambda > 0$.

On $\mathcal{D}_A$, $AT_t = T_t A$, $\forall\ t \geq 0$.

Proof. See [1] theorem 19.4. □

Proposition 27.19 (generator). There're following properties about the generator;

(i) $\mathcal{D}_A$ is dense in C_0, A is a closed operator;

(ii) $T_t \mathcal{D}_A \subset \mathcal{D}_A$, $\forall\ t \geq 0$;

(iii) $\forall\ f \in \mathcal{D}_A$, $T_t f$ is differentiable

$$\frac{d}{dt} T_t f = AT_t f = T_t A f$$

$$T_t f - f = \int_0^t T_s A f ds = \int_0^t AT_s f ds$$

(iv) (positive maximum principle) If $f \in \mathcal{D}_A$, and if $\sup\{f(x) : x \in S\} = f(x_0) \geq 0$, then $Af(x_0) \leq 0$.

Proof. (iv):

$$Af(x_0) = \lim_{t \downarrow 0} \frac{T_t f(x_0) - f(x_0)}{t}$$

$$T_t f(x_0) - f(x_0) \leq f(x_0)(\mu_t(x_0, S) - 1) \leq 0.$$

□

Remark. Take an example, for Brownian motion, $Af = \frac{1}{2} f''$, $\mathcal{D}_A = C_0^2$, then if $f(x_0)$ is maximal, we have $Af(x_0) = \frac{1}{2} f''(x_0) \leq 0$.

Now focus on the probabilistic significance of A:

Theorem 27.20 (Dynkin's formula). (X_t) is a Feller process. If $f \in \mathcal{D}_A$, define

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

then $(M_t^f, t \geq 0)$ is a $(\mathcal{F}_t, P_\pi)$ martingale for any π.

Remark. Again take Brownian motion for example, in this case, $Af = \frac{1}{2} f''$, for $f \in C_0^2$, $B_0 = 0$,(notice the following things only make sense locally, because we require $f(x) \to 0$ when $x \to \infty$)

$$\begin{aligned}
&f(x) = x, \quad M_t^f = B_t; \\
&f(x) = x^2, \quad M_t^f = B_t^2 - t; \\
&f(x) = x^3, \quad M_t^f = B_t^3 - 3\textstyle\int_0^t B_s ds; \\
&\vdots
\end{aligned}$$

Also, Itô's formula says, $\forall\ f \in C^2$,

$$f(B_t) = f(B_0) + f'(B) \cdot B + \frac{1}{2} f''(B) \cdot [B]$$

while for Brownian motion, $[B]_t = t$, i.e. $(\frac{1}{2}f''(B) \cdot [B])_t = \int_0^t Af(B_s)ds$, so in fact, for Brownian motion,

$$M_t^f = (f'(B) \cdot B)_t$$

which is a stochastic integration, and is surely again a martingale.

Proof. First, M_t^f is integrable. For any $t, h \geq 0$,

$$M_{t+h}^f - M_t^f = f(X_{t+h}) - f(X_t) - \int_t^{t+h} Af(X_s)ds$$

$$\begin{aligned}
E(M_{t+h}^f - M_t^f | \mathcal{F}_t) &= E(f(X_{t+h}) - f(X_t) - \int_t^{t+h} Af(X_s)ds | \mathcal{F}_t) \\
&= E_{X_t}(f(X_h) - f(X_0) - \int_0^h Af(X_s)ds) \\
&= T_h f(X_t) - f(X_t) - \int_0^h T_s Af(X_t)ds \\
&= 0
\end{aligned}$$

by **proposition** 27.19 (iii). □

Proposition 27.21 (reverse of Dynkin's). If $f \in C_0$, and there exists a function $g \in C_0$, s.t.

$$f(X_t) - f(X_0) - \int_0^t g(X_s)ds$$

is a $(\mathcal{F}_t, P_x)$ martingale for any $x \in S$, then $f \in \mathcal{D}_A$ and $Af = g$.

Proof. $\forall\, x \in S$,

$$T_t f(x) - f(x) - \int_0^t T_s g(x)ds = 0$$

hence,

$$||\frac{1}{t}(T_t f - f) - g|| = ||\frac{1}{t}\int_0^t (T_s g - g)ds|| \leq \frac{1}{t}\int_0^t ||T_s g - g||ds \to 0, \qquad t \downarrow 0.$$

□

Motivated by the last two results, there's a heuristic extension of the theory, we won't go further here. That is, we use the martingales in Dynkin's formula to define the extended generator for general Markov processes.

Definition 27.22 (general generator). (X_t) is a Markov process, a Borel function f is said to belong domain $\mathcal{D}_A$ of the extended infinitesimal generator if there exists a Borel function g s.t. a.s. $\int_0^t |g(X_s)|ds < \infty$ for each t, and

$$f(X_t) - f(X_0) - \int_o^t g(X_s)ds$$

is a $(\mathcal{F}_t, P_x)$ right continuous martingale for every x.

Of course this definition is consistent with that for Feller processes. If we define $Af := g$ in this case, A is called **extended infinitesimal generator**. This definition also makes perfect sense for Markov processes which are not Feller processes. Most of the results here can be extended to the more general case keeping the same probability significance.

Well, after the general theory, now we come to see some fundamental examples:

There're few cases where $\mathcal{D}_A$ and A are completely known, generally the subspace of $\mathcal{D}_A$ is satisfactory. Below we start with real valued Lévy processes. Let $\mathcal{A}$ be the space of infinitely differentiable functions on the real line s.t. $\lim_{|x|\to\infty} f^k(x)P(x) = 0$ for any polynomial $P(x)$, $\forall\ k \in \mathbb{Z}^+$. Fourier transform is one-to-one on $\mathcal{A}$ to itself. $\forall\ f, g \in \mathcal{A}$, define the inner product $< f, g >:= \int f(x)g(x)dx$.

Theorem 27.23 (generator of Lévy process). Let (X_t) be real valued Lévy process, then $\mathcal{A} \subset \mathcal{D}_A$, and $\forall\ f \in \mathcal{A}$,

$$Af(x) = \beta f'(x) + \frac{\sigma^2}{2} f''(x) + \int (f(x+y) - f(x) - \frac{y}{1+y^2} f'(x))\nu(dy).$$

Proof. Recall Lévy-Kintchin formula, use $\hat{\mu}_t$ denote the Fourier transform for X_t, then

$$\hat{\mu}_t(\mu) = e^{t\psi(\mu)}$$

$$\psi(\mu) = i\beta\mu - \frac{\sigma^2\mu^2}{2} + \int (e^{i\mu y} - 1 - \frac{i\mu y}{1+y^2})\nu(dy)$$

where $\beta \in \mathbb{R}$, $\sigma \geq 0$, ν is a Radon measure on $\mathbb{R}\backslash\{0\}$, s.t. $\int \frac{x^2}{1+x^2}\nu(dx) < \infty$.

I claim, $|\psi|$ increases at most like $|\mu|^2$ at infinite. To see this, notice

$$|\int_{[-1,1]^c} (e^{i\mu x} - 1 - \frac{i\mu x}{1+x^2})\nu(dx)| \leq 2\nu[-1,1]^c + |\mu| \int_{[-1,1]^c} \frac{|x|}{1+x^2}\nu(dx)$$

$$|\int_{-1}^{1} (e^{i\mu x} - 1 - \frac{i\mu x}{1+x^2})\nu(dx)| \leq |\mu| \int_{-1}^{1} |\frac{x}{1+x^2} - x|\nu(dx) + \int_{-1}^{1} |e^{i\mu x} - 1 - i\mu x|\nu(dx)$$

while $|e^{i\mu x} - 1 - i\mu x|$ is majored by $c|x|^2|\mu|^2$ for a constant c.

For any $f \in \mathcal{A}$, there exists unique $g \in \mathcal{A}$, s.t. $f(x) = \int e^{ivx}g(v)dv := \int g_x(v)dv =< 1, g_x >$, and then

$$\begin{aligned} < \hat{\mu}_t, g_x > &= \int (e^{ixv}g(v))(\int e^{iyv}\mu_t(dy))dv \\ &= \int \mu_t(dy) \int e^{i(x+y)v}g(v)dv \\ &= \int f(x+y)\mu_t(dy) \\ &= T_t f(x). \end{aligned}$$

So, use Taylor expansion,

$$\begin{aligned} T_t f(x) =< \hat{\mu}_t, g_x > &= \int e^{t\psi(v)}g_x(v)dv \\ &=< 1, g_x > +t < \psi, g_x > + \frac{t^2}{2}H(t,x) \end{aligned}$$

where $|H(x,t)| \leq \sup_{0\leq s\leq t} | < \psi^2 e^{s\psi}, g_x > | \leq < |\psi|^2, |g_x| >$.

We know $<|\psi|,|g_x|>$ and $<|\psi|^2,|g_x|>$ are finite, the bounds are independent with x, hence when $t\downarrow 0$,

$$\frac{1}{t}(T_tf(x)-f(x))\to <\psi,g_x> \qquad \text{uniformly on } x$$

So

$$Af(x)=<\psi,g_x>=<i\beta\mu-\frac{\sigma^2\mu^2}{2}+\int(e^{i\mu y}-1-\frac{i\mu y}{1+y^2})\nu(dy),g_x(\mu)>$$

and observe

$$f'(x)=i\int \mu g_x(\mu)d\mu=<i\mu,g_x(\mu)>;$$

$$f''(x)=i^2\int \mu^2 g_x(\mu)d\mu=<-\mu^2,g_x(\mu)>;$$

$$\begin{aligned}<\int(e^{i\mu y}-1-\frac{i\mu y}{1+y^2})\nu(dy),g_x(\mu)> &= \int\nu(dy)(\int(e^{i\mu y}-1-\frac{i\mu y}{1+y^2})g_x(\mu)d\mu)\\ &= \int(f(x+y)-f(x)-\frac{y}{1+y^2}f'(x))\nu(dy).\end{aligned}$$

□

There are three fundamental cases, here we focus on dimension $d=1$:

Proposition 27.24. (I) $X_t=\sigma B_t$, (B_t) is Brownian motion, then $\mathcal{D}_A=C_0^2$, $Af=\frac{\sigma^2}{2}f''$;

(II) Translation at speed β, $\mathcal{D}_A=$ absolutely continuous function in C_0^1, $Af=\beta f'(x)$;

(III) Poisson processes with rate λ, $\mathcal{D}_A=C_0$, $Af=\lambda(f(x+1)-f(x))$.

Proof. (I) see STAT205B lecture notes 18;

(II) $T_tf(x)=E_xf(X_t)=f(x+\beta t)$, for any $f\in C_0^1$, absolutely continuous,

$$Af(x)=\lim_{t\downarrow 0}\frac{f(x+\beta t)-f(x)}{t}=\beta f'(x);$$

(III) $T_tf(x)=e^{-\lambda t}f(x)+\sum_{i=1}^{\infty}\frac{(\lambda t)^i}{i!}e^{-\lambda t}f(x+i)$, for any $f\in C_0$,

$$\frac{d}{dt}T_tf(x)|_{t=0}=\lambda(f(x+1)-f(x)).$$

□

In all these cases, we can describe the whole $\mathcal{D}_A$, but this is rather unusual, and usually one can only describe the subspace of $\mathcal{D}_A$.

Heuristically speaking, from the last two results, a Lévy process is a mixture of a translation term, a diffusion term corresponding to $\frac{\sigma^2}{2}f''$ and a jump term, the jumps being described by the Lévy measure ν.

For general Feller processes in $\mathbb{R}^d$, we have the similar description as long as $\mathcal{D}_A\supset C_k^2$(compact support), but because these processes are no longer translation invariant as in stationary independent increments case, the translation, diffusion and jump terms will change with the position of the process.

Theorem 27.25 (generator for Feller processes). If (T_t) is a Feller semigroup on $\mathbb{R}^d$, $C_k^\infty \subset \mathcal{D}_A$, then

(i) $C_k^2 \subset \mathcal{D}_A$;

(ii) $\forall$ relatively compact open set U, there exists functions a_{ij}, b_i, c on U and a kernel N, s.t. $\forall\ f \in C_k^2$, $x \in U$,

$$\begin{aligned} Af(x) = c(x)f(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x) + \sum_{i,j} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \\ + \int_{\mathbb{R}^d\backslash\{x\}} (f(y) - f(x) - 1_U(x)\sum_i (y_i - x_i)\frac{\partial f}{\partial x_i}(x))N(x, dy) \end{aligned}$$

where $N(x,\cdot)$ is a Radon measure on $\mathbb{R}^d\backslash\{x\}$, the matrix $a(x) := (a_{ij}(x))$ is symmetric and nonnegative, $c \le 0$ and a, c don't depend on U.

If (X_t) has continuous path, then

$$Af(x) = c(x)f(x) + \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x) + \sum_{i,j} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x). \tag{$**$}$$

Remark. Intuitively, a process with above infinitesimal generator will move infinitesimally from a position x by adding a translation of vector $b(x)$, a Gaussian process with covariance $a(x)$ and jumps given by $N(x,\cdot)$; the term $c(x)f(x)$ corresponds to the possibility of being killed.

Also see [1] page 384, with some conditions, Feller process (x_t) has continuous path iff $(**)$ is satisfied. The resulting Markov process is called **canonical Feller diffusion**.

Here only one small application is shown:

Example 27.26 (Heat equation). $\forall\ f \in C_0^2$, we have for Brownian motion in $\mathbb{R}^d$,

$$\frac{d}{dt}T_t f = \frac{1}{2}\Delta T_t f$$

In fact, this is true for any bounded Borel function f and $t > 0$.

In the language of PDE, $T_t f$ are fundamental solutions of the heat equation:

$$\begin{cases} \frac{\partial}{\partial t}\mu(t,x) + \frac{1}{2}\Delta\mu(t,x) = 0 \\ \mu(0,x) = f(x). \end{cases}$$

[1] O. Kallenberg, Foundations of Modern Probability (2nd ed.), Springer-Verlag, 2002.

[2] M. Qian and G. Gong, Stochastic Processes (Chinese book, 2nd ed.), Peking university press, 1997.

[3] D. Revus and M. Yor, Continuous Martingales and Brownian motion (3rd ed.), Springer-Verlag, 1999