

Markov Chains in Continuous Time

Alexander Plakhin, Nikita Kiselev

HSE FES Probability Theory Club

November 25, 2022

Plan

1. Discrete state space
2. Continuous state space

Discrete State Space

CTMC, discrete states

Assume discrete state space \mathcal{S} . Suppose now that whenever a chain enters state $i \in \mathcal{S}$, independent of the past, the length of time spent in state i is a continuous, strictly positive (and proper) random variable H_i called the **holding time** in state i . When the holding time ends, the chain transitions to another state j . Our goal is to place conditions on the holding times to ensure that the continuous time process satisfies the Markov property.

Definition

A stochastic process $\{X_t, t \geq 0\}$ with discrete state space \mathcal{S} is called a (homogeneous) continuous-time Markov chain (CTMC) if for all $t, s \geq 0$, $i, j \in \mathcal{S}$

$$\mathbb{P}(X_{s+t} = j | X_s = i, \{X_u : 0 \leq u < s\}) = \mathbb{P}(X_{s+t} = j | X_s = i) = P_{ij}(t)$$

$P_{ij}(t)$ is the probability the chain will end up in state j , t time units from now, given it is in state i now.

Definition

For each $t \geq 0$ there is a transition matrix

$$P(t) = \{P_{ij}(t)\},$$

and $P(0) = I$ - the identity matrix.

Note that as in discrete-time Markov chains, $P_{ii} > 0$ is allowed, though such transition creates a new random variable H_i , independent of the past.

Holding times

Proposition

Holding times must have memoryless properties and thus be exponentially distributed.

Suppose $X_t = i$. Time t lies somewhere in the middle of the holding time H_i for state i . In order for the future to be independent of the past given $X_t = i$, we deduce that the remaining holding time must only depend (in distribution) on i and be independent of its age - the memoryless property follows. Now notice that the memoryless property is equivalent to the restriction that holding times are exponentially distributed.

Let's consider $H_i \sim \exp(a_i)$. □

Result: CTMC can be fully described by a transition matrix $P = \{P_{ij}\}$ and a set of rates $\{a_i : i \in \mathcal{S}\}$, the holding time rates.

Transition rate matrix (infinitesimal generator)

Proposition

Assume $P_{ii} = 0$ for all $i \in \mathcal{S}$. Then a_i can be interpreted as transition rate out of state i given $X_t = i$ - that is the exponential holding time will end, independent of the past, in the next dt units of time with probability $a_i dt$.

Let $N_i(t)$ denote a Poisson counting process with rate a_i . Then

$$\mathbb{P}\{N_i(h) > 0\} = \mathbb{P}\{N_i(h) = 1\} = a_i h + o(h)$$

Then

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}\{X_h \neq i | X_0 = i\}}{h} = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}\{N_i(h) = 1\}}{h} = \lim_{h \rightarrow 0^+} \frac{a_i h + o(h)}{h} = a_i$$



Transition rate matrix (infinitesimal generator)

More generally, for $i \neq j$ denote by P_{ij} the probability (independent of the past) of chain going to state j after leaving state i . Then

$$P'_{ij}(0) = \lim_{h \rightarrow 0^+} \frac{P_{ij}(h)}{h} = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}\{N_i(h) = 1\}P_{ij}}{h} = a_i P_{ij}$$

Therefore, $a_i P_{ij}$ can be interpreted as the transition rate from state i to state j given that the chain is currently in state i .

When $i = j$, $P_{ii}(h) = 1 - \mathbb{P}\{X_h \neq i | X_0 = i\}$ and so

$$P'_{ii}(0) = \lim_{h \rightarrow 0^+} \frac{P_{ii}(h) - 1}{h} = \lim_{h \rightarrow 0^+} \frac{-\mathbb{P}\{N_i(h) = 1\}}{h} = -a_i$$

Transition rate matrix (infinitesimal generator)

Definition

The matrix $Q = P'(0)$ given explicitly by equalities on the previous slide is called the transition rate matrix (infinitesimal generator) of the Markov chain.

Example

Let $S = \{0, 1, 2, 3\}$, then

$$Q = \begin{pmatrix} -a_0 & a_0 P_{01} & a_0 P_{02} & a_0 P_{03} \\ a_1 P_{10} & -a_1 & a_1 P_{12} & a_1 P_{13} \\ a_2 P_{20} & a_2 P_{21} & -a_2 & a_2 P_{23} \\ a_3 P_{30} & a_3 P_{31} & a_3 P_{32} & -a_3 \end{pmatrix}$$

Notice that since we assumed $P_{ii} = 0$ for all i , each row of Q sums to 0.

Definition

Let τ_n be the time of n th transition. Then $X_n = X_{\tau_n+}$ (the state right after the n th transition) defines the underlying discrete-time Markov chain called the embedded Markov chain.

$\{X_n, n \in \mathbb{N}\}$ keeps track of the states visited right after each transition, and in our previous notation, it moves between states according to the one-step transition probabilities $P_{ij} = \mathbb{P}\{X_{n+1} = j | X_n = i\}$. This embedded transition matrix $\{P_{ij}\}$ together with holding time rates $\{a_i\}$ determines the CTMC.

Chapman-Kolmogorov equations

Proposition

For all $t, s \geq 0$

$$P(t + s) = P(t)P(s)$$

In other words, for all $t, s \geq 0, i, j \in \mathcal{S}$

$$P_{ij}(t + s) = \sum_{k \in \mathcal{S}} P_{ik}(t)P_{kj}(s)$$

Can be easily derived from using conditional probability and combining it with the Markov property. □

Examples of CTMC's

Example (Poisson counting process)

Let $\{N_t : t \geq 0\}$ be the counting process for the Poisson process $\{\varphi_t\}$ with rate λ . Then $\{N_t\}$ forms a CTMC with $\mathcal{S} = \{0, 1, \dots\}$, $P_{i,i+1} = 1$, $a_i = \lambda$. Notice that if $N_0 = 0$, then $X_n = N_{t_n+} = n$ - the embedded Markov chain is deterministic and non-decreasing.

Example (Rat in a maze)

The rat in a closed maze is equally likely to move to one of the two neighboring cells at each transition. The holding time in cell i equals H_i - exponentially distributed with rate $a_i = i$, $i = 1, 2, 3, 4$. Let X_t denote the state at which rat is located at time t . One of the interests is to calculate the long-run proportion of time spent at each cell (notice that the average holding time in each cell is different):

$$P_i := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t I\{X_s = i\} ds, \quad i = 1, 2, 3, 4$$

Communication classes, irreducibility and recurrence

For CTMC's we can also say that states i and j are *reachable* if there exists $s \geq 0$ such that $\mathbb{P}\{X_s = j | X_0 = i\} = P_{ij}(s) > 0$. Similarly, we can partition the state space into disjoint communication classes, $S = C_1 \cup C_2 \cup \dots$.

Irreducible chain is such a chain that consists of one communication class.

Proposition

A CTMC is irreducible if and only if its embedded Markov chain is irreducible.

The recurrence and transience of states can also be defined in a known way. Let T_{ii} denote the amount of time until the chain revisits state i starting from $X_0 = i$. Note that $T_{ii} \geq H_i$. State i is called *recurrent* if $\mathbb{P}\{T_{ii} < \infty\} = 1$ and *transient* otherwise.

Proposition

A state i is recurrent/transient for a CTMC if and only if it is recurrent/transient for the embedded discrete-time chain.

Because communication classes are determined by the embedded chain as well, using the above proposition we can conclude that in a CTMC communication classes all have the same types of states - they are all either recurrent or transient.

Positive recurrence and the existence a limiting distribution

Definition

State i is called *positive recurrent* if, in addition to being recurrent, $\mathbb{E}(T_{ii}) < \infty$.

State i is called *null recurrent* if, in addition to being recurrent, $\mathbb{E}(T_{ii}) = \infty$.

Unlike recurrence, positive (or null) recurrence is not equivalent to that for the embedded chain: it is possible for a CTMC to be positive recurrent while its embedded chain is null recurrent (and vice versa).

But positive and null recurrence are still class properties - in particular, in an irreducible CTMC all states are together transient, positive recurrent or null recurrent.

Positive recurrence and the existence a limiting distribution

Definition

A CTMC is called positive recurrent if it is irreducible and it is positive recurrent. We define (if they exist and independent of $X_0 = i$) *limiting probabilities* $\{P_j\}$ as the long-run proportion of time spent in each state $j \in \mathcal{S}$:

$$P_j = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t I\{X_s = j | X_0 = i\} ds$$

Taking expectations of both sides, we obtain

$$P_j = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t P_{ij}(s) ds$$

Positive recurrence and the existence a limiting distribution

When each P_j exists and $\sum_{j \in \mathcal{S}} = 1$, then $\vec{P} = (P_1, P_2, \dots)$ (row-vector) is the *limiting* (stationary) distribution of CTMC. Denote

$$P^* = \begin{pmatrix} \vec{P} \\ \vec{P} \\ \vdots \end{pmatrix}$$

- the matrix with all rows equal to limiting probabilities. Then we can rewrite the previous equality in a nice way:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t P(s) ds = P^*$$

Proposition

If X_t is a positive recurrent CTMC, then the limiting probability \vec{P} exists, it is unique and given by

$$P_j = \frac{\mathbb{E}H_j}{\mathbb{E}T_{jj}} = \frac{1}{a_j \mathbb{E}T_{jj}}$$

Moreover, convergence holds:

$$P_j = \lim_{t \rightarrow \infty} P_{ij}(t), \quad i, j \in \mathcal{S}$$

Finally, if CTMC is either null recurrent or transient, no limiting distribution exists (strictly speaking, $\forall j \in \mathcal{S} P_j = 0$).

Positive recurrence and the existence a limiting distribution

For a fixed state j , we can break up the whole process into i.i.d. (due to Markov property) cycles corresponding to returning to j . Let $\tau_n(j)$ denote the n th time at which the chain returned to j and $\tau_0(j) = 0$. Let $T_n(j) = \tau_n(j) - \tau_{n-1}(j)$ denote the cycle length - they are also i.i.d. distributed as T_{jj} . $\{\tau_n(j)\}$ forms a renewal process and let $N_t(j)$ be the corresponding counting process. For such processes we know

$$\lim_{t \rightarrow \infty} \frac{N_t(j)}{t} = \frac{1}{\mathbb{E} T_{jj}}$$

Let

$$J_n = \int_{\tau_{n-1}(j)}^{\tau_n(j)} I\{X_s = j\} ds$$

- the amount of time spent in state j during n th cycle. Again, J_n are i.i.d. and obviously distributed as holding time H_j . Then $\mathbb{E} J_n = \mathbb{E} H_j$. □

Positive recurrence and the existence a limiting distribution

We can approximate

$$\int_0^t I\{X_s = j\} ds \approx \sum_{n=1}^{N_j(t)} J_n$$

Then

$$\frac{1}{t} \int_0^t I\{X_s = j\} ds = \frac{N_j(t)}{t} \cdot \frac{1}{N_j(t)} \sum_{n=1}^{N_j(t)} J_n$$

Taking the limit, we obtain

$$P_j = \frac{\mathbb{E}H_j}{\mathbb{E}T_{jj}}$$

Uniqueness follows by the unique representation of P_j .



Arbitrary random initial value

Assume $X_0 \sim \nu$, then the distribution of X_t is given by $X_t \sim \nu P(t)$.
Recalling the definition of P^* , we derive for any distribution ν

$$\nu P^* = \vec{P}$$

Therefore, we can left-multiply both sides of equality from slide 17 to get

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \nu P(s) ds = \vec{P}$$

Logic - the limiting distribution is not affected by the initial distribution.

Limiting distribution \leftrightarrow stationary distribution

Proposition

For a positive recurrent CTMC limiting distribution is stationary - if $X_0 \sim \vec{P}$, then $X_t \sim \vec{P}$, that is $\vec{P}P(t) = \vec{P}$.

We can equivalently show $P^*P(t) = P^*$ for any $t \geq 0$:

$$\begin{aligned} P^*P(t) &= \left(\lim_{u \rightarrow \infty} \frac{1}{u} \int_0^u P(s) ds \right) P(t) = \lim_{u \rightarrow \infty} \frac{1}{u} \int_0^u P(s)P(t) ds = \\ &= \lim_{u \rightarrow \infty} \frac{1}{u} \int_0^u P(s+t) ds = \lim_{u \rightarrow \infty} \frac{1}{u} \int_0^u P(s) ds = \\ &= P^* \end{aligned}$$



Kolmogorov backward equations

How to compute transition probabilities $P_{ij}(t) = \mathbb{P}\{X_t = j | X_0 = i\}$?

Theorem (Kolmogorov backward equations)

For CTMC with transition rate matrix $Q = P'(0)$, the following linear differential equation is satisfied:

$$P'(t) = QP(t), \quad t > 0, \quad P(0) = I,$$

that is

$$P'_{ij}(t) = -a_i P_{ij}(t) + \sum_{k \neq i} a_i P_{ik} P_{kj}(t), \quad i, k \in \mathcal{S}, \quad t \geq 0$$

The unique solution is thus

$$P(t) = e^{Qt}, \quad t \geq 0$$

Kolmogorov backward equations

From the Chapman-Kolmogorov equation,

$$P(t+h) - P(t) = (P(h) - I)P(t) = (P(h) - P(0))P(t)$$

Dividing by h and limiting to h to 0 yields

$$P'(t) = P'(0)P(t) = QP(t)$$



Continuous State Space

We can intuitively (not precisely) define a continuous-time Markov process with state space \mathbb{R} as a stochastic process whose future depends on its present state and not on how it got there:

$$\mathbb{P}(X_{t+h} \in \Gamma | \{X_s, s \leq t\}) = \mathbb{P}(X_{t+h} \in \Gamma | X_t)$$

for a Borel sets Γ .

We will consider continuous-time Markov processes for which a conditional probability density exists:

$$\mathbb{P}(X_{t+h} \in \Gamma | X_t = x) = \int_{\Gamma} p(y, t+h, x, t) dy$$

Example of Markov process

Example

Brownian motion is a Markov process with conditional probability density given by the following formula:

$$\mathbb{P}(W_{t+h} \in \Gamma | W_t = x) = \int_{\Gamma} \frac{1}{\sqrt{2\pi h}} \exp\left(-\frac{(y-x)^2}{2h}\right) dy$$

The Markov property of Brownian motion follows from the fact that it has independent increments.

Precise definition of a Markov process

Definition

Let X_t be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mu)$ with values in \mathbb{R}^d , and let \mathcal{F}_t^X be the filtration generated by $\{X_t; t \in \mathbb{R}_+\}$. Then $\{X_t; t \in \mathbb{R}_+\}$ is a Markov process if

$$\mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X) = \mathbb{P}(X_t \in \Gamma | X_s)$$

for all $t, s \in T$ with $t \geq s$, and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$.

Transition function

With every continuous-time Markov process X_t (we always take $t \in \mathbb{R}^d$) defined in a probability space $(\Omega, \mathcal{F}, \mu)$ and state space $(\mathbb{R}^d, \mathcal{B})$, we can associate the transition function

$$P(\Gamma, t|X_s, s) := \mathbb{P}(X_t \in \Gamma | \mathcal{F}_s^X)$$

for all $t, s \in \mathbb{R}_+$ with $t \geq s$ and all $\Gamma \in \mathcal{B}(\mathbb{R}^d)$

For fixed $t, X_s = x, s$ we will consider the following function:

$$P(\Gamma, t|x, s) := \mathbb{P}(X_t \in \Gamma | X_s = x)$$

The Chapman–Kolmogorov equation

Proposition

For all $x \in \mathbb{R}^d$, $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, and $s, u, t \in \mathbb{R}_+$ with $s \leq u \leq t$ and transition function P :

$$P(\Gamma, t|x, s) = \int_{\mathbb{R}^d} P(\Gamma, t|y, u)P(dy, u|x, s)$$

The derivation of the Chapman–Kolmogorov equation is based on the Markovian assumption, properties of conditional probability and the fact that for $s < t \Rightarrow \mathcal{F}_s^X \subset \mathcal{F}_u^X$.

$$\begin{aligned} P(\Gamma, t|x, s) &= P(X_t \in \Gamma | X_s = x) = P(X_t \in \Gamma | \mathcal{F}_s^X) = \\ &= \mathbb{E}(\mathbb{I}_\Gamma(X_t) | \mathcal{F}_s^X) = \mathbb{E}(\mathbb{E}(\mathbb{I}_\Gamma(X_t) | \mathcal{F}_s^X) | \mathcal{F}_u^X) = \mathbb{E}(\mathbb{E}(\mathbb{I}_\Gamma(X_t) | \mathcal{F}_u^X) | \mathcal{F}_s^X) = \\ &= \mathbb{E}(P(X_t \in \Gamma | X_u = y) | X_s = x) = \int_{\mathbb{R}^d} P(\Gamma, t|y, u)P(dy, u|x, s) \end{aligned}$$

Intuition of the Chapman–Kolmogorov equation

$$P(\Gamma, t|x, s) = \int_{\mathbb{R}^d} P(\Gamma, t|y, u)P(dy, u|x, s)$$

In words, the Chapman–Kolmogorov equation tells us that for a Markov process, the transition from x at time s to the set Γ at time t can be done in two steps: first, the system moves from x to y at some intermediate time u . Then it moves from y to Γ at time t . In order to calculate the probability for the transition from x at time s to Γ at time t , we need to sum (integrate) the transitions from all possible intermediate states y .

Another form of the Chapman–Kolmogorov equation

We consider Markov processes whose transition function has a density:

$$P(\Gamma, t|x, s) = \int_{\Gamma} p(y, t|x, s) dy$$

We will refer to $p(y, t|x, s)$ as the transition probability density. Notice that:

$$P(\Gamma, t|x, s) = \int_{\Gamma} p(y, t|x, s) dy = \int_{\mathbb{R}^d} \int_{\Gamma} p(y, t|z, u) p(z, u|x, s) dz dy$$

We obtain the Chapman–Kolmogorov equation for the transition probability density

$$p(y, t|x, s) = \int_{\mathbb{R}^d} p(y, t|z, u) p(z, u|x, s) dz$$

Time-homogeneous Markov Process

We will say that a Markov process is time-homogeneous if the transition function $P(\cdot, t|\cdot, s)$ depends only on the difference $t - s$.

$$P(\Gamma, t|x, s) = P(\Gamma, t - s|x, 0) = P(t - s, x, \Gamma)$$

We can fix the initial time, $s = 0$. The Chapman–Kolmogorov equation for a time-homogeneous Markov process then becomes:

$$P(t + s, x, \Gamma) = \int_{\mathbb{R}^d} P(s, x, dz) P(t, z, \Gamma) dz$$

The Chapman–Kolmogorov equation for densities:

$$p(t + s, x, y) = \int_{\mathbb{R}^d} p(s, x, z) P(t, z, y) dz$$

Example

Example

Example of a time-homogeneous Markov process is Brownian motion. The transition function is the Gaussian:

$$P(t, x, dy) = \gamma_{t,x}(y)dy, \quad \gamma_{t,x}(y) = \frac{1}{\sqrt{2\pi t}} \exp -\frac{|x - y|^2}{2t}$$

The Generator of a Markov Process

Let X_t denote a time-homogeneous Markov process. The Chapman–Kolmogorov equation suggests that a time-homogeneous Markov process can be described through a semigroup of operators, i.e., a one-parameter family of linear operators with the properties

$$P_0 = I, \quad P_{t+s} = P_s \circ P_t \quad \text{for all } t, s \geq 0$$

Let $f \in C_b(\mathbb{R}^d)$, the space of continuous bounded functions on \mathbb{R}^d , and define the operator:

$$(P_t f)(x) := \mathbb{E}(f(X_t) | X_0 = x) = \int_{\mathbb{R}^d} f(y) P(t, x, dy)$$

This is a linear operator with:

$$(P_0 f)(x) = \mathbb{E}(f(X_0) | X_0 = x) = f(x) \Rightarrow P_0 = I$$

The Generator of a Markov Process

Furthermore,

$$\begin{aligned}(P_{t+s}f)(x) &= \int_{\mathbb{R}^d} f(y)P(t+s, x, dy) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y)P(s, z, dy)P(t, x, dz) = \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y)P(s, z, dy) \right) P(t, x, dz) = \int_{\mathbb{R}^d} (P_s f)(z)P(t, x, dz) = \\ &= (P_t \circ P_s f)(x)\end{aligned}$$

Consequently,

$$P_{t+s} = P_t \circ P_s f$$

The semigroup P_t is an example of a Markov semigroup. We can study properties of a time-homogeneous Markov process X_t by studying properties of the Markov semigroup P_t .

The Generator of a Markov Process

Let now X_t be a time-homogeneous Markov process in \mathbb{R}^d , and let P_t denote the corresponding semigroup that we defined. We consider this semigroup acting on continuous bounded functions and assume that $P_t f$ is also a $C_b(\mathbb{R}^d)$ function. We define by $\mathcal{D}(\mathcal{L})$ the set of all $f \in C_b(\mathbb{R}^d)$ such that the strong limit:

$$\mathcal{L}f := \lim_{t \rightarrow 0} \frac{P_t f - P_0 f}{t} = \lim_{t \rightarrow 0} \frac{P_t f - f}{t}$$

The operator $\mathcal{L} : \mathcal{D}(\mathcal{L}) \rightarrow C_b(\mathbb{R}^d)$ is called the (infinitesimal) generator of the operator semigroup P_t . We will also refer to \mathcal{L} as the generator of the Markov process X_t .

The Generator of a Markov Process

The semigroup property and the definition of the generator of a Markov semigroup imply that formally, we can write:

$$P_t = e^{t\mathcal{L}}$$

Consider the function $u(x, t) := (P_t f)(x) = \mathbb{E}(f(X_t) | X_0 = x)$. We calculate its partial derivative:

$$\frac{\partial u}{\partial t} = \frac{d}{dt}(P_t f) = \frac{d}{dt}(e^{t\mathcal{L}} f) = \mathcal{L}(e^{t\mathcal{L}} f) = \mathcal{L}P_t f = \mathcal{L}u$$

Furthermore, $u(x, 0) = P_0 f(x) = f(x)$. Consequently, $u(x, t)$ satisfies the initial value problem:

$$\frac{\partial u}{\partial t} = \mathcal{L}u$$

$$u(x, 0) = f(x)$$

The backward Kolmogorov equation

Before we got the initial value problem. It is the backward Kolmogorov equation.

$$\frac{\partial u}{\partial t} = \mathcal{L}u$$

$$u(x, 0) = f(x)$$

It governs the evolution of the expectation of an observable $f \in C_b(\mathbb{R}^d)$. When the Markov process is the solution of a stochastic differential equation, then the generator is a second-order elliptic differential operator, and the backward Kolmogorov equation becomes an initial value problem for a parabolic partial differential equation.

Example

Example

As an example, consider Brownian motion in one dimension. The transition function, the fundamental solution of the heat equation in one dimension. The corresponding Markov semigroup is the heat semigroup $P_t = \exp \frac{t}{2} \frac{d^2}{dx^2}$. The generator of the one-dimensional Brownian motion is the one-dimensional Laplacian $\frac{t}{2} \frac{d^2}{dx^2}$. The backward Kolmogorov equation is the heat equation:

$$\frac{\partial u}{\partial t} = \frac{t}{2} \frac{\partial^2 u}{\partial x^2}$$

The Adjoint Semigroup

The semigroup P_t acts on bounded continuous functions. We can also define the adjoint semigroup P_t^* , which acts on probability measures:

$$P_t^* \mu(\Gamma) = \int_{\mathbb{R}^d} \mathbb{P}(X_t \in \Gamma | X_0 = x) d\mu(x) = \int_{\mathbb{R}^d} \mathbb{P}(t, x, \Gamma) d\mu(x)$$

The image of a probability measure μ under P_t^* is again a probability measure. The operators P_t and P_t^* are (formally) adjoint in the L^2 -sense:

$$\int_{\mathbb{R}} P_t f(x) d\mu(x) = \int_{\mathbb{R}} f(x) d(P_t^* \mu)(x)$$

We can write adjoint of the generator of the process:

$$P_t^* = e^{t\mathcal{L}^*}$$

$$\int \mathcal{L} f h dx = \int f \mathcal{L}^* h dx$$

Forward Kolmogorov equation

Let X_t be a Markov process with generator \mathcal{L} with $X_0 \sim \mu$, and let P_t^* denote the adjoint semigroup. We define

$$\mu_t := P_t^* \mu$$

This is the law of the Markov process. An argument similar to that used in the derivation of the backward Kolmogorov equation enables us to obtain an equation for the evolution of μ_t :

$$\frac{\partial \mu_t}{\partial t} = \mathcal{L}^* \mu_t, \quad \mu_0 = \mu$$

Forward Kolmogorov equation

Assuming that the initial distribution μ and the law of the process t each have a density with respect to Lebesgue measure, denoted by $\rho_0(\cdot)$ and $\rho(\cdot, t)$ respectively, this equation becomes:

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^* \rho, \quad \rho(y, 0) = \rho_0(y)$$

This is the forward Kolmogorov equation. When the Markov process starts at $X_0 = x$, deterministic, the initial for the Fokker–Planck equation becomes $\rho_0 = \delta(x - y)$. As with the backward Kolmogorov equation, this equation is still formal, since we do not have a formula for the adjoint of the generator of the Markov process X_t .

1. Karl Sigman, IEOR 6711: Continuous-Time Markov Chains
(<http://www.columbia.edu/~ks20/stochastic-I/stochastic-I-CTMC.pdf>)
2. Grigorios A. Pavliotis: Stochastic Processes and Applications