Math 106 – Notes
Week 5: February 3, 2026

Ethan Levien

General continuous time Markov processes (5.3)

Now we return to Markov processes. We have already seen that some Gaussian processes are Markov Processes as well. Soon, we will see how these can be represented with the same forward/backward equations as Q-processes. Along they way we will encounter other Markov processes. A Markov process is defined as follows.

Definition 1 (Markov processes). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A process X_t is a Markov processes if it is $\mathcal{F}_t$-adopted processes and for $s \leq t$ and $B \in \mathcal{R}$*

$$\mathbb{P}(X_t \in B|\mathcal{F}_s) = \mathbb{P}(X_t \in B|X_s) \tag{1}$$

We will consider homogenous Markov processes which are defined by a transition function

$$p(t, x, B) = \mathbb{P}(X_t \in B|X_0 = x) \tag{2}$$

we also define an linear operator

$$T_t f(x) = \mathbb{E}[f(X_t)|X_0 = x] \tag{3}$$

If there is a density $\rho(t, x, y)$ such that $p(t, x, dy) = p(t, x, [y, y + dy)) \approx \rho(t, x, y)dy$, then this becomes

$$T_t f(x) = \int f(y)\rho(t, x, y)dy \tag{4}$$

A considerable amount of time will later be spent studying the PDE for ρ, but first we spend a bit more time on the general case.

Note that $p(t, x, B)$ and $T_t f(x)$ correspond respectively to the notation $P_{i,j}(t)$ and $h_i(t)$ which was introduced for Q-processes. Let f be bounded and measurable. Using time–homogeneity and the Markov property,

$$\begin{aligned} T_t(T_s f)(x) &= \mathbb{E}\big[T_s f(X_t) \mid X_0 = x\big] & (5)\\ &= \mathbb{E}\big[\mathbb{E}[f(X_{t+s}) \mid X_t] \mid X_0 = x\big] & (6)\\ &= \mathbb{E}[f(X_{t+s}) \mid X_0 = x] & (7)\\ &= T_{t+s} f(x). & (8) \end{aligned}$$

Moreover $T_0 f(x) = f(x)$. Thus $\{T_t\}_{t\geq 0}$ satisfies

$$T_0 = I, \qquad T_{t+s} = T_t T_s, \tag{9}$$

and is a semigroup of linear operators.

A brief detour: semigroup theory

We now pause to introduce the basic semigroup language needed later.

Definition 2 (Strongly continuous semigroup DF.1)**.** *Let B be a Banach space (a complete normed vector space). A family of linear operators $\{T_t\}_{t\geq 0}$ on B is called a* strongly continuous operator semigroup*, or C_0-semigroup if*

$$T_0 = I, \tag{10}$$

$$T_{t+s} = T_t T_s, \qquad t, s \geq 0, \tag{11}$$

$$\lim_{t\downarrow 0} \|T_t f - f\| = 0 \quad \textit{for all } f \in B. \tag{12}$$

The continuity condition (12) *is called* strong continuity*.*

Remark 1. *Recall that for a bounded linear operator $T : X \to X$,*

$$\|T\| \;=\; \sup_{f\neq 0} \frac{\|Tf\|}{\|f\|}. \tag{13}$$

Strong continuity implies that each T_t is a bounded linear operator, so boundedness need not be assumed separately (this is related to the Uniform boundedness principle).

The Markov operators $\{T_t\}$ constructed above always satisfy the algebraic properties (10)–(11). Strong continuity (12) is an additional regularity assumption, which holds for many Markov processes of interest (e.g. diffusions) and expresses the idea that the process evolves continuously in time when tested against observables f.

Example: Translation Semigroup

A canonical example is the translation semigroup on functions on $\mathbb{R}$,

$$(T_t f)(x) = f(x + vt). \tag{14}$$

In words, T_t takes a function and translates it horizontally with velocity v. From the perspective of stochastic processes, we can view T_t as the semigroup associated with the deterministic process

$$\frac{d}{dt} X_t = v. \tag{15}$$

Even though X_t is deterministic, we can still formally take the expectation which just gives $\mathbb{E}[f(X_t)|X_0 = x] = f(X_t) = f(x + vt)$.

Lemma 1. *The translation semigroup is strongly continuous on the Banach space*

$$C_0(\mathbb{R}) \coloneqq \left\{ f : \mathbb{R} \to \mathbb{R} : f \textit{ is continuous and } \lim_{x\to\pm\infty} = 0 \right\} \tag{16}$$

with the norm $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$.

Proof. Take $v = 1$ for simplify. Note that

$$\|T_t f\|_\infty = \sup_{x\in\mathbb{R}} |f(x+t)| = \sup_{y\in\mathbb{R}} |f(y)| = \|f\|_\infty,$$

so T_t is a bounded linear operator with $\|T_t\| = 1$. The semigroup identities follow immediately from the definition. Thus it remains to show strong continuity. Fix $f \in C_0(\mathbb{R})$ and $\varepsilon > 0$. Since f vanishes at infinity, there exists $R > 0$ such that $|f(x)| < \varepsilon/3$ for $|x| \geq R$. On the compact set $I = [-R-1, R+1]$, f is uniformly continuous (this is the Heine-Cantor Theorem), meaning that for all $\varepsilon > 0$ there exists t such that

$$\sup_{x \in I} |f(x+t) - f(x)| < \varepsilon/3 \tag{17}$$

Moreover, when $|x| > R$ and $|t| < 1$ small, both $|f(x)|$ and $|f(x+t)|$ are $< \varepsilon/3$. Thus

$$\|T_t f - f\|_\infty = \sup_{x \in \mathbb{R}} |f(x+t) - f(x)| < 2 \sup_{x:|x|>R} |f(x)| + \sup_{x \in I} |f(x+t) - f(x)| \leq \varepsilon \tag{18}$$

which proves the result. □

Why strong, not uniform, continuity?

A stronger notion would be *uniform continuity*, which requires

$$\|T_t - I\| \to 0 \quad \text{as } t \downarrow 0, \tag{19}$$

i.e. *every* function in the unit ball of B is moved only a small amount, uniformly over all such functions, for sufficiently small times. This requirement is too restrictive for most semigroups arising in probability, including the translation semigroup.

Lemma 2. *The translation semigroup is not uniformly continuous on $C_0(\mathbb{R})$.*

Infinitesimal generator and the resolvent operator

Just as was done for the Q-process, we would like characterize the process not by the transition operator T_t (for Q process this was just $P(t)$), but instead by something like the Q matrix. Following the book, we use $\mathcal{A}$ for the generalization of the Q matrix. Recall that for a Q- process on $\Omega = \{1, \ldots, I\}$, $f = (f(1), \ldots, f(I))$ is a vector and we already showed how

$$(\mathcal{A}f)_i = (Qf)_i = \lim_{t \downarrow 0} \frac{T_h f(i) - f(i)}{h} = \sum_{j \in \Omega} \lambda_{i,j} f(j) - \lambda_{j,j} f(j) \tag{20}$$

If T_t is known, we can calculate $\mathcal{A}$ by differentiating at $t = 0$. For example, in the case the translation operator,

$$\mathcal{A}f = \lim_{t \downarrow 0} \frac{T_h f(x) - f(x)}{h} = \frac{d}{dt}(T_t f)(x)\Big|_{t=0} = \frac{d}{dt} f(x + vt) = v f'(x). \tag{21}$$

This suggests $\mathcal{A} = vd/dt$, which is correct. However, we need to address the fact that f may not be differentiable, so Equation 21 does not make sense for all the functions in our Banach space. Moreover, $\mathcal{A}$ is not even a bounded linear operator. The Hille–Yosida theorem tells us when a bounded linear operator on a dense subset of a Banach spaces it is the generator of a strongly-continuous semigroup. For our purposes, we will generally be given an $\mathcal{A}$ which we know is the generator of a Markov process and the subtle issues about what space it is defined on will be handled with as needed, not in the abstract.

Note that more generally, translations in $\mathbb{R}^n$ are generated by ∇, and we have the same problem with differentiability.

Forward–backward structure and duality of Banach spaces

A continuous–time Markov process naturally gives rise to two complementary evolution equations corresponding to the Forward and Backward equations encountered in Q-processes. These two viewpoints are most clearly understood through the duality structure of Banach spaces.

Consider the backward viewpoint for the semigroup $\{T_t\}_{t\geq 0}$ acts by

$$T_t f(x) = \mathbb{E}_x[f(X_t)] = \mathbb{E}[f(X_t)|X_0 = x],$$

with infinitesimal generator $\mathcal{A}$. Formally, $u(t,x) = T_t f(x)$ satisfies the backward equation

$$\partial_t u(t,x) = \mathcal{A}u(t,x), \qquad u(0,x) = f(x),$$

Just as for Q-processes, this description is backward in the sense that one specifies a terminal observable and propagates it forward in time by conditioning on the present state.

The forward viewpoint describes the evolution of distributions. If μ is a probability measure on the state space, its time evolution is given by the pushforward under the transition kernel,

$$\mu_t(B) = \int p(t,x,B)\, \mu_0(dx).$$

If $\mu_0(B)$ can be expressed as a density $\mu_0(dy) = \rho_0(y)dy$, this becomes

$$\mu_t(B) = \int p(t,x,B)\, \rho_0(x)dx.$$

Now we introduce the pairing between observables and measures given by the usual inner product, which also happens to be the expectation when viewed as an operator on B; that is, for a measure μ and $f \in B$ we have

$$\langle f, \mu\rangle := \mathbb{E}[f(X)] = \int f(x)\, \mu(dx). \tag{22}$$

This inner product is viewed as a mapping $\langle \cdot, \cdot \rangle : B \times B^* \to \mathbb{R}$ where B^* is the dual space of B^*. We won't get into this too deeply, but it is good to know that there is a theorem – the Riesz representation theorem – which identifies the dual space. For example, when observables are taken in $B = C_0(\mathbb{R})$ it can be shown that $B^* = C_0(\mathbb{R})^*$ is the space of of finite signed Borel measures on $\mathbb{R}$. Modulo normalization these are our probability measures.

Using this duality structure, we can derive the Forward equation. First note that for $f \in B$,

$$\langle T_t f, \mu_0\rangle = \langle f, T_t^* \mu_0\rangle = \langle f, \mu_t\rangle. \tag{23}$$

Therefore, as we would expect, the adjoint semigroup T_t^* propagates probability measures in the same way that left multiplication by $P(t) = e^{tQ}$ propagates probability vectors for Q–processes.

Differentiating this identity at $t = 0$ and using the definition of the generator $\mathcal{A}$, we obtain

$$\frac{d}{dt}\langle T_t f, \mu_0\rangle\Big|_{t=0} = \langle \mathcal{A}f, \mu_0\rangle = \langle f, \mathcal{A}^*\mu_0\rangle \tag{24}$$

Since this identity holds for all test functions $f \in B$, it characterizes the forward (Kolmogorov) equation in weak form:

$$\partial_t \mu_t = \mathcal{A}^* \mu_t, \tag{25}$$

with initial condition $\mu_{t=0} = \mu_0$. Thus the generator $\mathcal{A}$ governs the backward evolution of observables, while its adjoint $\mathcal{A}^*$ governs the forward evolution of probability distributions.