A summary of Markov processes.

1. **Probabilistic definition.** $X = X(t)$, $t \geq 0$, on $\big(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\big)$ and with values in a complete separable metric space E (`phase/state space`) is called `Markov` if

$$\mathbb{P}\big(X(t) \in A|\mathcal{F}_s\big) = \mathbb{P}\big(X(t) \in A|X(s)\big)$$

for all $A \in \mathcal{B}(E)$ and $t > s \geq 0$ [$\mathcal{B}(E)$ is the Borel sigma-algebra on E].

Note. Typically, $\mathcal{F}_t = \sigma\big(X(s),\ s \leq t\big)$. Occasionally, situations when $\sigma\big(X(s),\ s \leq t\big) \subset \mathcal{F}_t$ do appear.

Fact. The `transition probability function`

$$P(t, A; s, y) = \mathbb{P}\big(X(t) \in A|X(s) = y\big) \tag{1}$$

(transition from y at time s to A at time t) satisfies the `Chapman-Kolmogorov equation`

$$P(t, A; s, y) = \int_E P(t, A; r, z)P(r, dz; s, y),\ s < r < t. \tag{2}$$

2. **Analytical definition.** The function $P(t, A; s, y)$, $t \geq s \geq 0$, $A \in \mathcal{B}(E)$, $y \in E$, is called the `transition kernel/function/semigroup/etc.` if [in addition to the usual measurability] $P(t, \cdot; s, y)$ is a probability measure on $\big(E, \mathcal{B}(E)\big)$ [for each (t, s, y)],

$$P(t, A; t, y) = \chi_A(y) = \begin{cases} 1, & y \in A \\ 0, & y \notin A. \end{cases} \tag{3}$$

and (2) holds. *The function from* (1) *has all these properties.*

3. **Notations.**

$$P_{t,s}[f](x) = \int_E f(z)P(t, dz; s, x),\ f \text{ is boubded measurable}$$
$$P^*_{t,s}[\nu](A) = \int_E P(t, A; s, y)\nu(dy),\ \nu \text{ is a probability measure.}$$

In particular, property (3) implies that $P_{t,t}[f] = f$. Also, by (2), with $s < r < t$,

$$\int_E f(x)P(t, dx; s, y) = \int_E \left(\int_E f(x)P(t, dx; r, z)\right) P(r, dz; s, y),$$
$$\int_E P(t, A; s, y)\nu(dy) = \int_E P(t, A; r, z) \left(\int_E P(r, dz; s, y)\nu(dy)\right),$$

and we get the *semigroup properties*

$$P_{t,s}[f] = P_{r,s}[P_{t,r}[f]]; \tag{4}$$
$$P^*_{t,s}[\nu] = P^*_{t,r}[P^*_{r,s}[\nu]]; \tag{5}$$

note the "reverse" order of the operators on the right of (4).

Probabilistic interpretations: $P_{t,s}[f](x) = \mathbb{E}\Big(f\big(X(t)\big)|X(s) = x\Big)$; $P^*_{t,s}[\mu]$ is the distribution of $X(t)$ if the distribution of $X(s)$ is ν: $P^*_{t,s}[\nu](A) = \mathbb{P}\big(X(t) \in A\big)$, assuming $\mathbb{P}\big(X(s) \in B\big) = \nu(B)$, for all $A, B \in \mathcal{B}(E)$.

4. **Generator** $\mathcal{L}_t : f \mapsto \mathcal{L}_t f$ is defined by

$$(\mathcal{L}_t f)(x) = \lim_{h\to 0+} \frac{P_{t+h,t}[f](x) - f(x)}{h} = \lim_{h\to 0+} \frac{\mathbb{E}\Big(f\big(X(t+h)\big)|X(t) = x\Big) - f(x)}{h}$$

(for whatever functions the limit exists).

Note:

- The definition makes sense only in continuous time;
- $\mathcal{L}_t 1 = 0$;
- If x_0 is a point of local maximum of f, then $(\mathcal{L}_t f)(x_0) \leq 0$ (weak maximum principle).
- The fundamental theorem of calculus immediately implies

$$\mathbb{E}f\big(X(t)\big) = \mathbb{E}f\big(X(s)\big) + \mathbb{E}\int_s^t \big(\mathcal{L}_r f\big)\big(X(r)\big)dr,\ t > s. \tag{6}$$

The `Dynkin formula` states that, under some extra (minor) technical conditions,

$$f\big(X(t)\big) - f\big(X(s)\big) - \int_s^t \big(\mathcal{L}_r f\big)\big(X(r)\big)dr$$

is a (local) martingale for $t > s$ [that is, (6) actually holds for all (bounded) *stopping times*].

5. **The forward Kolmogorov equation** (or Fokker-Planck equation, with Smoluchowksi, Liouville, Krylov and Bogolyuobov also having something to do with it...) describes the evolution in t of the measure $P(t, \cdot; s, y)$ for fixed s, y, that is, the evolution of the transition kernel in the *forward variables.* The equation follows directly from (6), putting $X(s) = y$ and differentiating with respect to t:

$$\frac{\partial P_{t,s}[f]}{\partial t} = P_{t,s}[\mathcal{L}_t f],\ t > s,\ P_{s,s}[f] = f. \tag{7}$$

IF

$$P(t, A; s, y) = \int_A V(t, x; s, y)d\ell(x) \tag{8}$$

[that is, the measure $P(t, \cdot; s, y)$ has a density with respect to some reference measure ℓ (positive, sigma-finite, ... say, Lebesgue ...) on $\mathcal{B}(E)$], then

$$\frac{\partial V(t, x; s, y)}{\partial t} = \mathcal{L}_t^{*,(x)} V(t, x; s, y),\ t > s,\ V(s, x; s, y) = \delta(x - y), \tag{9}$$

where $\mathcal{L}_t^*$ is the adjoin of $\mathcal{L}_t$ *relative to the reference measure* ℓ, that is

$$\int_E (\mathcal{L}_t f) g\, d\ell = \int_E (\mathcal{L}_t^* g) f\, d\ell, \tag{10}$$

and $\mathcal{L}_t^{*,(x)}$ means that the operator is acting on the x variable in $V(t, x; s, y)$

6. **The backward Kolmogorov equation** describes the evolution of the function $P(t, A; s, y)$ in the *backward variables* s, y. Alternatively, given a bounded measurable f, we consider the function $u(s, y) = \mathbb{E}\Big(f\big(X(t)\big)|X(s) = y\Big) = P_{t,s}[f](y)$ for fixed $t \geq s$. Then, leaving all the technicalities and using (4) with $r = s$ (and $s = s - h$),

$$\begin{aligned}
\frac{\partial u(s, y)}{\partial s} &= -\lim_{h\to 0+} \frac{P_{t,s-h}[f](y) - P_{t,s}[f](y)}{h} = -\lim_{h\to 0+} \frac{P_{s,s-h}[P_{t,s}f](y) - P_{t,s}[f](y)}{h} \\
&= -\lim_{h\to 0+} \frac{P_{s,s-h}[u(s, \cdot)](y) - u(s, y)}{h} = -\lim_{h\to 0+} \frac{P_{s+h,s}[u(s+h, \cdot)](y) - u(s+h, y)}{h} \\
&= -\big(\mathcal{L}_s u(s, \cdot)\big)(y),
\end{aligned}$$

where the last equality follows from the definition of the generator; the main technicality is confirming that the function u is indeed in the domain of the generator.

In other words,

$$\frac{\partial u(s, y)}{\partial s} + \big(\mathcal{L}_s u(s, \cdot)\big)(y) = 0,\ s < t;\ u(t, y) = f(y). \tag{11}$$

Taking $f=\chi_A$ (indicator function of a set), we get

$$\frac{\partial P(t,A;s,y)}{\partial s}+\big(\mathcal{L}_s P(t,A;s,\cdot)\big)(y)=0.$$

7. **Note.** The forward Kolmogorov equation is easier to derive and is somewhat more "natural" from physical point of view, but it is somewhat hard to use in the form (7), whereas (9) requires additional assumptions.

8. **Time homogenous case** is one time variable instead of two:

$$P(t,A;s,y)=p(t-s,A;y),\ p_t[f](x)=\int_E f(z)p(t,dz;x),\ p_t^*[\nu](A)=\int_E p(t,A;y)\nu(dy);$$

$$(\mathcal{L}_t f)(x)=(\mathcal{L}f)(x)=\lim_{t\to 0+}\frac{p_t[f](x)-f(x)}{t}=\lim_{t\to 0+}\frac{\mathbb{E}\Big(f\big(X(t)\big)|X(0)=x\Big)-f(x)}{t}.$$

Also defined in this case are the *carré du champ operator*

$$\Gamma(f,g)=\frac{1}{2}\left(\mathcal{L}(fg)-f\mathcal{L}g-g\mathcal{L}f\right)$$

and, when there is a unique invariant measure $\bar{\mu}$ (that is, $p_t^*[\bar{\mu}]=\bar{\mu}$), the *energy/Dirichlet* form

$$\mathcal{E}(f,g)=\int_E \Gamma(f,g)d\bar{\mu}.$$

The symmetry of the generator $\mathcal{L}$ on $L_2(E,\bar{\mu})$ is closely related to *reversibility* of the process X, although, as a rule, $\mathcal{L}\neq\mathcal{L}^*$: the adjoint $\mathcal{L}^*$ is computed with respect to the *reference measure* that is usually NOT the same as $\bar{\mu}$; compare (10) with

(12) $$\int_E (\mathcal{L}f)g\, d\bar{\mu}=\int_E(\mathcal{L}g)f\, d\bar{\mu}.$$

The forward Kolmogorov equation (7) can now be written in a compact (but potentially confusing) operator form

(13) $$\frac{\partial p}{\partial t}=p\mathcal{L}.$$

After time reversal $u(T-s,y)=U(s,y)$, the backward Komogorov equation (11) can be written as

$$\frac{\partial U}{\partial s}=\mathcal{L}U,\ t>0;\ \ U|_{s=0}=f(x),$$

or

$$\frac{\partial p(t,A;y)}{\partial t}=\mathcal{L}p(t,A;y),\ p(0,A;y)=\chi_A(y);$$

the analogue of (13) is

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

which is even more confusing, given that you would expect the generator to commute with the semigroup it generates.

IF $E\subset\mathbb{R}^{\mathrm{d}}$ and IF $p(t,A;y)=\int_A v(t,x;y)dx$, then the `transition density` v satisfies

$$\frac{\partial v(t,x;y)}{\partial t}=\mathcal{L}^{(y)}v(t,x;y),\ t>0,\ \ v(0,x;y)=\delta(x-y);$$

notation $\mathcal{L}^{(y)}$ indicates that the operator $\mathcal{L}$ is acting on the y variable.

Furthermore, IF there is a unique invariant measure $\bar{\mu}(dx)$ with $\bar{\mu}(dx) = \rho(x)dx$ and IF $\mathcal{L}$ has a complete orthonormal basis of eigenfunctions φ_k in $L_2(E;\bar{\mu})$ [$\mathcal{L}\varphi_k = -\lambda_k\varphi_k$], with the corresponding eigenvalues $-\lambda_k < 0$, $k \geq 1$, $\lambda_k \to +\infty$, $k \to \infty$, THEN, keeping in mind that

$$(f,g)_{L_2(E,\bar{\mu})} = \int_E f(y)g(y)\rho(y)dy,$$

we find

$$v(t,x;y) = \sum_{k=0}^{\infty} e^{-\lambda_k t}\big(v(0,x;\cdot),\varphi_k\big)_{L_2(E,\bar{\mu})}\varphi_k(y) = \rho(x)\sum_k e^{-\lambda_k t}\varphi_k(x)\varphi_k(y). \tag{14}$$

From (14), using $\lambda_0 = 0$ [because $\mathcal{L}1 = 0$], we *hope to get ergodicity*:

$$\lim_{t\to+\infty} v(t,x;y) = \rho(x)$$

for every y; the rate of this convergence should depend on λ_1 [the reason we cannot claim this convergence from (14) right away is that we do not know the behavior of $\varphi_k(x)$ for fixed x as $k \to +\infty$.]

9. **Examples.**

Discrete time and discrete space is the basic Markov chain, characterized by the transition probability matrix.

Discrete time and arbitrary space is a Markov sequence, such as $X_{n+1} = a(X_n) + b(X_n)\xi_{n+1}$ with iid ξ_n; the phase space of ξ_n determines the phase space of X_n. The generic `random walk` corresponds to $a(x) = x$, $b(x) = 1$: $X_n = \sum_{k=1}^n \xi_k$.

Continuous time and discrete space is a continuous time Markov chain, such as Poisson process or a birth-death process. The corresponding generator is known as the `jump intensity matrix` and has zero row sums. The size of the jump intensity matrix can be infinite.

Continuous time and continuous space is an (Itô) diffusion process

$$dX(t) = b(t,X(t))dt + \sqrt{2}\,\sigma(t,X(t))dW(t) \tag{15}$$

with generator

$$(\mathcal{L}_t f)(x) = b(t,x)f'(x) + \sigma^2(t,x)f''(x). \tag{16}$$

Note:

- *Stationary increments* imply *time homogenous* (but not necessarily the other way around: Think geometric Brownian motion.)
- *Independent increments* imply Markov property, but (obviously) not necessarily the other way around.
- `The Lévy process` is a process with independent and stationary increments. Prescribing the distribution of the increments specifies the process; the name of the distribution, such as Poisson, Gamma, Pascal (same as Negative Binomial), is often used in the name of the process as well.
- Factoring out $\sqrt{2}$ in the diffusion coefficient in (15) saves a lot of factors of 2 in the subsequent formulas.

In the special case of the time-homogeneous scalar diffusion

$$dX(t) = b\big(X(t)\big)dt + \sqrt{2}\,\sigma\big(X(t)\big)dW(t)$$

such that $X(t) \in E \subset \mathbb{R}$ and

$$\int_E \frac{1}{\sigma^2(x)} \exp\left(\int \frac{b(x)}{\sigma^2(x)}dx\right)dx < \infty,$$

we deduce from (16) [using the Lebesgue measure as the reference measure] that

$$\mathcal{L}^* f = \Big(\big(\sigma^2(x) f(x)\big)' - b(x) f(x) \Big)'$$

and therefore, with a suitable number c, the function

$$\rho(x) = \frac{c}{\sigma^2(x)} \exp\left(\int \frac{b(x)}{\sigma^2(x)} dx \right)$$

is the *invariant density* for X:

$$\mathcal{L}^* \rho = 0.$$

The generator $\mathcal{L}$ is *symmetric* on $L_2(E; \rho(x)dx)$ because $(\sigma^2 \rho)' = b\rho$ [but obviously $\mathcal{L} \neq \mathcal{L}^*$].

If $b(x) = ax + b$ and $\sigma^2(x) = \sigma_0 + \sigma_1 x + \sigma_2 x^2$, then, with suitable choices of a, b, σ_k we can use (14) to get an explicit expression for the transition density $v(t, x; y)$ [that is, the pdf of $X(t)$ given $X(0) = y$] in terms of classical orthogonal polynomial (Hermite, Laguerre, and Jacobi).

Only the case of Hermite polynomials, corresponding to the Ornstein-Uhlenbeck process

$$dX = -X dt + \sqrt{2}\, dW(t),$$

with $b(x) = -x, \sigma(x) = 1$, leads to a closed-form expression for the transition density:

$$v(t, x; y) = \frac{1}{\sqrt{2\pi(1 - e^{-2t})}} \exp\left(-\frac{(x - ye^{-t})^2}{2(1 - e^{-2t})} \right) \sim \mathcal{N}(ye^{-t}, 1 - e^{-2t}).$$

Indeed, if $X(0) = y$, then

$$X(t) = ye^{-t} + \sqrt{2} \int_0^t e^{-(t-s)} dW(s) \sim \mathcal{N}(ye^{-t}, 1 - e^{-2t}).$$

A really brutal (but potentially illuminating...) exercise on partial differentiation is verifying that the function $V(t, x; s, y) = v(t - s, x; y)$ satisfies the forward Kolmogorov equation

$$\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2} + x \frac{\partial V}{\partial x} + V, \; t > s,$$

and the backward Kolmogorov equation

$$\frac{\partial V}{\partial s} + \frac{\partial^2 V}{\partial y^2} - y \frac{\partial V}{\partial y} = 0, \; s < t.$$

References.

[1] D. Bakry, I. Gentil, and M. Ledoux. Analysis and geometry of Markov diffusion operators [Part I]. Springer, 2014.

[2] G. DaPrato and J. Zabczyk. Ergodicity for infinite dimensional systems [Chapter 2]. Cambridge University Press, 1996.

[3] D. W. Stroock and S. R. S. Varadhan. Multidimensional diffusion processes [Chapter 0]. Springer, 1979.