A Summary of (Abstract) Ergodic Theory for (Abstract) Markov Processes.

Main objects.

State Space $\mathbf{E}$: A Polish metric space (complete separable metric space) with metric $\rho = \rho(x, y)$;
Metric ball in $\mathbf{E}$: $\mathrm{B}(y, r) = \{x \in \mathbf{E} : \rho(x, y) < r\}$, center y, radius r;
Transition Probability Function (Transition Semigroup) : $P = P_t(x, A),\ t \geq 0,\ x \in \mathbf{E},\ A \in \mathcal{B}(\mathbf{E})$;
(Real – valued) Bounded measurable functions on $\mathbf{E}$: $\mathbb{B}_b(\mathbf{E})$;
(Real – valued) Bounded continuous functions on $\mathbf{E}$: $\mathcal{C}_b(\mathbf{E})$;
Probability measures on $\big(\mathbf{E}, \mathcal{B}(\mathbf{E})\big)$: $\mathcal{M}_1(\mathbf{E})$.

Basic Assumptions on P.

(1) For every $t \geq 0$ and $x \in \mathbf{E}$, the mapping $A \mapsto P_t(x, A)$ is a probability measure on $\mathcal{B}(\mathbf{E})$;
(2) For every $t \geq 0$ and $A \in \mathcal{B}(\mathbf{E})$, the mapping $x \mapsto P_t(x, A)$ is measurable [from $\big(\mathbf{E}, \mathcal{B}(\mathbf{E})\big)$ to $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$].
(3) Semigroup Property or Chapman-Kolmogorov equation:

$$P_{t+s}(x, A) = \int_{\mathbf{E}} P_t(x, dy) P_s(y, A) = \int_{\mathbf{E}} P_s(x, dy) P_t(y, A), \quad t, s > 0,\ A \in \mathcal{B}(\mathbf{E}). \tag{1}$$

(4) Starting from identity:

$$P_0(x, A) = 1(x \in A). \tag{2}$$

In terms of a Markov process: if $\mathbb{X} = \big(X(t), t \geq 0\big)$ is defined on $\big(\Omega, \mathcal{F},\ (\mathcal{F})_{t\geq 0}, \mathbb{P}\big)$ so that $X(t) \in \mathbf{E}$ is $\mathcal{F}_t$-measurable, the Markov property

$$\mathbb{E}\big(\varphi(X(t))|\mathcal{F}_s\big) = \mathbb{E}\big(\varphi(X(t))|X(s)\big),\ t > s \geq 0,\ \varphi \in \mathbb{B}_b(\mathbf{E}),$$

holds, and the process is (time-)homogeneous:

$$\mathbb{E}\big(\varphi(X(t))|X(s)\big) = \mathbb{E}\big(\varphi(X(t-s))|X(0)\big),\ t > s \geq 0,\ \varphi \in \mathbb{B}_b(\mathbf{E}),$$

then

$$\mathbb{P}\big(X(t) \in A | X(s) = x\big) = P_{t-s}(x, A),\ t > s \geq 0,\ x \in \mathbf{E},\ A \in \mathcal{B}(\mathbf{E}).$$

Note that these constructions work the same for both discrete and continuous time parameter.

Basic Notations:

- Total variation distance:

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{B}(\mathbf{E})} |\mu(A) - \nu(A)|, \quad \mu, \nu \in \mathcal{M}_1(\mathbf{E});$$

- Action on functions:

$$P_t[\varphi](x) = \int_{\mathbf{E}} \varphi(y) P_t(x, dy), \quad x \in \mathbf{E},\ \varphi \in \mathbb{B}_b(\mathbf{E});$$

- Action on measures:

$$P_t^*[\nu](A) = \int_{\mathbf{E}} P_t(x, A) \nu(dx), \quad A \in \mathcal{B}(\mathbf{E}),\ \nu \in \mathcal{M}_1(\mathbf{E});$$

In terms of the Markov process $\mathbb{X}$: $P_t[\varphi](x) = \mathbb{E}\Big(\varphi\big(X(t)\big)|X(0)=x\Big)$; $P^*_t[\nu](A) = \mathbb{P}\big(X(t)\in A\big)$ if the distribution of $X(0)$ is ν, that is, $\mathbb{P}\big(X(0)\in C\big)=\nu(C)$.

To note: by (2), $P_0[\varphi]=\varphi$, $P^*_0[\nu]=\nu$; also, $P_t[1]=1$.

Advanced definitions.

(1) $\boldsymbol{\mu}\in\mathcal{M}_1(\mathbf{E})$ is called INVARIANT MEASURE for P if $P^*_t[\boldsymbol{\mu}](A)=\boldsymbol{\mu}(A)$ for all $A\in\mathcal{B}(\mathbf{E})$ and $t\geq 0$. *Equivalently,*

$$\int_{\mathbf{E}} P_t[\varphi](x)\boldsymbol{\mu}(dx) = \int_{\mathbf{E}} \varphi(x)\boldsymbol{\mu}(dx),\ \varphi\in\mathbb{B}_b(\mathbf{E}). \tag{3}$$

(2) P is called FELLER if

$$\lim_{t\to 0+} P_t\big(x,\mathrm{B}(x,r)\big)=1,\ r>0,\ x\in\mathbf{E}; \tag{4}$$

$$\varphi\in\mathcal{C}_b(\mathbf{E})\ \Rightarrow\ P_t[\varphi]\in\mathcal{C}_b(\mathbf{E}),\ t>0. \tag{5}$$

(3) P is called STRONG FELLER at $t_0>0$ if it is Feller and

$$\varphi\in\mathbb{B}_b(\mathbf{E})\ \Rightarrow\ P_{t_0}[\varphi]\in\mathcal{C}_b(\mathbf{E}). \tag{6}$$

(4) P is called REGULAR at $t_1>0$ if the measures $P_{t_1}(x,\cdot)$ are equivalent for all $x\in\mathbf{E}$;

(5) P is called IRREDUCIBLE at $t_2>0$ if $P_{t_2}(x,A)>0$ for all $x\in\mathbf{E}$ and all *open* sets $A\subseteq\mathbf{E}$.

To Note:

- Condition (4) is equivalent to $\lim_{t\to 0+}P_t[\varphi]=\varphi$ for every bounded *Lipschitz continuous* $\varphi:\mathbf{E}\to\mathbb{R}$.
- If $\boldsymbol{\mu}\in\mathcal{M}_1(\mathbf{E})$ and P_t is *symmetric* on $L_2(\mathbf{E};\boldsymbol{\mu})$, that is
$$\int_{\mathbf{E}} P_t[\varphi](x)\,\psi(x)\,\boldsymbol{\mu}(dx) = \int_{\mathbf{E}} P_t[\psi](x)\,\varphi(x)\,\boldsymbol{\mu}(dx),\ t\geq 0,\ \varphi,\psi\in L_2(\mathbf{E};\boldsymbol{\mu}),$$
then $\boldsymbol{\mu}$ is invariant for P [take ψ identically equal to 1 and get (3)].

The Ergodic Theorem. If P has a unique invariant measure $\boldsymbol{\mu}$ and $f:\mathbf{E}\to\mathbb{R}$ is a function such that $\int_{\mathbf{E}}|f(x)|\,\boldsymbol{\mu}(dx)<\infty$, then, for the corresponding Markov process $\mathbb{X}$, the following convergence takes place, both with probability one and in L_1:

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f\big(X(t)\big)\,dt = \int_{\mathbf{E}} f(x)\,\boldsymbol{\mu}(dx)\ \ \text{(continuous time)}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\big(X(k)\big) = \int_{\mathbf{E}} f(x)\,\boldsymbol{\mu}(dx)\ \ \text{(discrete time).}$$

Moreover, under some additional conditions, the Central Limit Theorem holds for a suitably normalized difference.

The main results.

(1) **Krylov-Bogolyubov:** If P is Feller and there exist $\nu, \boldsymbol{\mu} \in \mathcal{M}_1(\mathbf{E})$ such that, for some sequence $t_n \nearrow +\infty$,

$$\lim_{n\to\infty} \frac{1}{t_n}\int_0^{t_n} P_t^*[\nu]\,dt = \boldsymbol{\mu} \text{ weakly,}$$

then $\boldsymbol{\mu}$ is invariant for P.

(2) **Doob:** If P has an invariant measure $\boldsymbol{\mu}$ and there exists a $t_0 > 0$ such that P is regular at t_0, then
- $\boldsymbol{\mu}$ is the unique invariant measure for P;
- $\boldsymbol{\mu}$ is equivalent to $P_t(x,\cdot)$ for all $x \in \mathbf{E}$ and $t \ge t_0$;
- as $t \to \infty$, $P_t(x,\cdot)$ converges to $\boldsymbol{\mu}$ in *total variation* for every $x \in \mathbf{E}$.

(3) **Khasminskii:** If P is strong Feller at $t_0 > 0$ and irreducible at $t_1 > 0$, then P is regular at $t_0 + t_1$.

(4) **Doeblin:** If there exist $\nu_0 \in \mathcal{M}_1(\mathbf{E})$, $\varepsilon > 0$ and $t_0 > 0$ so that, for all $A \in \mathcal{B}(\mathbf{E})$,

$$P_{t_0}^*[\nu](A) \ge \varepsilon \nu_0(A),\ \ \nu \in \mathcal{M}_1(\mathbf{E}),$$

then there is a unique invariant measure $\boldsymbol{\mu}$ for P and

$$\|P_t^*[\nu] - \boldsymbol{\mu}\|_{\mathrm{TV}} \le \frac{1}{1-\varepsilon} e^{-\lambda t}\ \|\nu - \boldsymbol{\mu}\|_{\mathrm{TV}}, \quad \lambda = -\frac{\ln(1-\varepsilon)}{t_0}.$$

(5) **Bonus:** A Feller Markov process is strong Markov (that is, can be re-started at a stopping time), and a Feller semigroup is a transition semigroup of a strong Markov family [Theorems 31.11. and 31.12 in Fristedt and Gray, *A Modern Approach to Probability Theory*, Birkhäuser/Springer, 1997]

Now assume that time is discrete: $t = 0, 1, 2, \ldots$; $P_1(x, A) = P(x, A)$ is the one-step transition probability function: $P(x, A) = \mathbb{P}\big(X(1) \in A | X(0) = x\big)$; $P_n(x, A)$ is the n-step transition probability $P_n(x, A) = \mathbb{P}\big(X(n) \in A | X(0) = x\big)$. Even though there is a certain abuse of notations, with P denoting the whole transition semigroup and the one-step transition function, note that, by Chapman-Kolmogorov, one-step transition probabilities determine everything else: $P_2(x, A) = \int_{\mathbf{E}} P(x, dy)P(y, A)$, etc.

Doeblin condition(s).

[MGD] **More general**: there are $\nu_0 \in \mathcal{M}_1(\mathbf{E})$, $n_0 \ge 1$, $\varepsilon \in (0, 1)$ so that, for all $A \in \mathcal{B}(\mathbf{E})$ with $\nu_0(A) \le \varepsilon$ and all $x \in \mathbf{E}$, we have $P_{n_0}(x, A) \le 1 - \varepsilon$.

[LGD] **Less general**: there are $\nu_0 \in \mathcal{M}_1(\mathbf{E})$, $n_0 \ge 1$, and $\delta \in (0, 1)$ such that, for all $x \in \mathbf{E}$, we have $P_{n_0}(x, A) \ge \delta \nu_0(A)$.

Note that

(1) condition [MGD] always holds when $\mathbf{E}$ is finite, with N elements: take ν to be uniform on $\mathbf{E}$ and $\varepsilon < 1/N$ so that $\nu(A) \le \varepsilon$ means A is an empty set;

(2) condition [LGD] implies [MGD] with the same ν_0 and n_0 and with $\varepsilon = 1/2$;

(3) condition [LGD] holds for finite state irreducible aperiodic Markov sequences, with ν equal to the invariant distribution and n_0 is whatever ensures $p_{ij}^{(n_0)} > 0$ for all $i, j \in \mathbf{E}$;

(4) condition [LGD] ensures *geometric ergodicity*: existence and uniqueness of the invariant measure $\boldsymbol{\mu}$ such that

$$\|P_n(x,\cdot) - \boldsymbol{\mu}\|_{\mathrm{TV}} \le \big(1-\delta\big)^{(n/n_0)-1}, \quad x \in \mathbf{E}.$$

The Harris Chain. Here is one possible definition: The Markov sequence $\mathbb{X} = \big(X(n),\ n \geq 0\big)$ with $X(n) \in \mathbf{E}$, is called a HARRIS CHAIN if

$$\exists\ \varepsilon > 0\ \exists\ A_0 \in \mathcal{B}(\mathbf{E})\ \exists\ \nu \in \mathcal{M}_1(\mathbf{E})\ \forall\ A \in \mathcal{B}(\mathbf{E})\ \forall\ y \in A_0 :\ \ P(y, A) \geq \varepsilon\nu(A), \tag{7}$$

AND

$$\forall\ x \in \mathbf{E} : \mathbb{P}\Big(\inf_{n\geq 0}\big(X(n) \in A_0\big) < \infty \big| X(0) = x\Big) = 1. \tag{8}$$

The point is that, given a certain "stability" of the sequence $\mathbb{X}$, it should be possible to work with a "local version" of the Doeblin condition(s) such as (7), when something like [LGD] holds for x not in all of the state space; the stability of $\mathbb{X}$ is ensured by a suitable recurrence, such as (8), or by existence of a Lyapunov function. Then ergodicity can be proved under some additional conditions.

Here is an **example** (M. Hairer, J. Mattingly, 2011). Assume that there exist a measurable function $V : \mathbf{E} \to [0, +\infty)$ (the Lyapunov function) and numbers $\gamma \in (0, 1)$ and $K > 0$ such that $P[V](x) \leq \gamma V(x) + K,\ x \in \mathbf{E}$, and also condition (7) holds with $A_0 = \{x \in \mathbf{E} : V(x) \leq 2K/(1-\gamma)\}$. For a measurable real-valued function φ on $\mathbf{E}$ define

$$\|\varphi\| = \sup_{x\in\mathbf{E}} \frac{|\varphi(x)|}{1 + V(x)}.$$

Then P has a unique invariant measure $\boldsymbol{\mu}$. Moreover, $\int_{\mathbf{E}} V(x)\, \boldsymbol{\mu}(dx) < \infty$ and there exists a number $C > 0$ such that, for all φ with $\|\varphi\| < \infty$,

$$\|P_n[\varphi] - \bar{\varphi}\| \leq C\gamma^n \|\varphi - \bar{\varphi}\|,\ \ \bar{\varphi} = \int_{\mathbf{E}} \varphi(x)\, \boldsymbol{\mu}(dx). \tag{9}$$

Note that (9) suggests an alternative way to measure the distance between two probability measures μ, ν on $\mathcal{B}(\mathbf{E})$: given a suitable class $\mathfrak{F}$ of real-valued functions on $\mathbf{E}$,

$$\|\mu - \nu\|_{\mathfrak{F}} = \sup_{\varphi\in\mathfrak{F}} \left| \int_{\mathbf{E}} \varphi(x)\, \mu(dx) - \int_{\mathbf{E}} \varphi(x)\, \nu(dx) \right|.$$

In particular, the total variation distance corresponds to taking $\mathfrak{F}$ as the collection of indicator functions of sets from $\mathcal{B}(\mathbf{E})$, whereas the WASSERSTEIN DISTANCE (also known as the Kantorovich–Rubinstein metric) W_1 corresponds to taking $\mathfrak{F}$ as the collection of Lipschitz continuous functions on $\mathbf{E}$ with the Lipschitz constant at most one: $|\varphi(x) - \varphi(y)| \leq \rho(x, y)$, $x, y \in \mathbf{E}$, where ρ is the metric on $\mathbf{E}$.

The people.

- Nikolay Mitrofanovich Krylov (1878–1955); not to be confuse with Nicolai Vladimirovich Krylov (b. 1941), a professor at the University of Minnesota.
- Nikolay Nikolayevich Bogolyubov (1909–1992), a student of N. M. Krylov.
- Joseph Leo Doob (1910–2004), Harvard Ph.D. (1932), professor at UIUC (1935–1978), advisor of David Blackwell.
- Rafail Zalmanovich Khasminskii (b. 1931), student of Dynkin, eventually a professor at Wayne State University.
- Wolfgang Doeblin (1915–1940), died in WWII.
- Theodore Edward Harris (1919–2005), was a math professor at USC.
- William Feller (1906–1970), was born in Zagreb, Croatia, and his original name was Vilibald Srećko Feller.