Markov chains and Markov decision processes

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

A great deal of the political-economy modeling that uses time-series structure does so in the form of a stochastic process whose future depends on the past only through the present. A voter updating her belief about a candidate's competence after each electoral observation does not, for the next round of updating, need to remember the entire history of past elections — her current posterior summarizes everything relevant. A two-party system whose voter-share dynamics are driven by generational party-affiliation transmission carries no memory of the affiliation patterns three generations back; only the current distribution matters for the next generation. An authoritarian regime's survival probability under stochastic mobilization shocks depends on the current legitimacy stock and current repression intensity, not on the path by which today's legitimacy stock arose. In each case the analyst is implicitly modeling the dynamics as a *Markov chain*: the present state is a sufficient statistic for the future, and the dependence on history collapses to dependence on the present.

Three structural insights organize this handout. First, the *Markov property* reduces the analysis of a stochastic process to the analysis of a transition matrix $\mathbf{P}$ on a state space S, with the entire dynamic governed by repeated application of $\mathbf{P}$ (§2–§3). Second, under irreducibility and aperiodicity, a finite-state Markov chain has a unique *stationary distribution* π to which it converges from any starting point, with the rate of convergence governed by the spectral gap — the gap between the dominant eigenvalue at 1 and the second-largest in modulus (§4–§5). The convergence theorem is one of the central structural results of the theory and the formal underpinning of essentially all long-run claims about Markov-modeled political-economy dynamics. The Perron–Frobenius machinery of the eigenvalue handout supplies the proof. Third, when the agent has decisions to make and the transitions depend on those decisions, the framework generalizes to a *Markov decision process* (§6), and the stochastic Bellman equation is a contraction in the same Banach sense as the dynamic-optimization handout's deterministic Bellman equation.

The handout closes the loop on the stochastic generalization that the dynamic-optimization handout deferred (its §3 parenthetical and Exercise 9). It also threads two earlier handouts: the eigenvalue-and-quadratic-forms handout (#19), whose Perron–Frobenius theorem is the structural input on the chain side, and the dynamic-optimization handout (#23), whose contraction-mapping argument extends with minimal modification to the stochastic case. The applied clusters that follow — game theory most immediately, where Markov-perfect equilibria are the strategic generalization of MDPs — will lean on this material directly.

2 Markov chains and the Markov property

What is the simplest workable structural model of a political process whose dynamics are stochastic? A voter is forming beliefs about a candidate's competence across many elections; a two-party system is shedding and gaining members across generations; a regime is accumulating or losing legitimacy

across years of stochastic events; a network of legislators is exchanging information about the policy environment across sessions. The single most useful modeling assumption for capturing this kind of process — and the one that the formal-theory literature reaches for first — is the assumption that the future depends on the past only through the present: the current state is a sufficient statistic for what comes next, and the entire history of the process is irrelevant once the current state is known. This is the *Markov property*, and it is the foundation of the framework this handout develops.

Throughout the handout, S is a finite or countable state space. Generalizations to uncountable state spaces are sketched in §7.

Definition 1. A discrete-time stochastic process $\{X_t\}_{t\geq 0}$ taking values in S is a *Markov chain* if for every $t \geq 0$ and every history $x_0, x_1, \ldots, x_{t+1} \in S$,

$$\mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = \mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t),$$

whenever the conditioning event has positive probability. The chain is *time-homogeneous* if the right-hand side does not depend on t.

Time-homogeneity is the standing assumption in this handout (and in essentially all of the applied political-economy literature using Markov-chain machinery). Under it, the entire dynamic is encoded by the *transition matrix* $\mathbf{P}$ with entries

$$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i), \qquad i, j \in S.$$

Each row of $\mathbf{P}$ is a probability distribution over S (*row-stochastic*: $\sum_j P_{ij} = 1$ for every i, with $P_{ij} \geq 0$).

The chain's evolution from a known initial distribution μ_0 is now a matter of matrix arithmetic. Writing μ_t for the distribution of X_t as a row vector, $\mu_{t+1} = \mu_t \mathbf{P}$, hence $\mu_t = \mu_0 \mathbf{P}^t$.[1] The *$n$-step transition probabilities* $P_{ij}^{(n)} = \mathbb{P}(X_{t+n} = j \mid X_t = i)$ are the entries of $\mathbf{P}^n$.

Proposition 2 (Chapman–Kolmogorov)**.** *For all $n, m \geq 0$ and $i, j \in S$,*

$$P_{ij}^{(n+m)} = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)}.$$

This is the matrix identity $\mathbf{P}^{n+m} = \mathbf{P}^n \mathbf{P}^m$ written out elementwise, and the probabilistic content is the natural one: to go from i to j in $n+m$ steps, the chain must pass through some intermediate state k at step n, and the probabilities sum over the choice of k.

Example 3 (Two-party generational party-affiliation chain)**.** Let $S = \{D, R\}$ and let $\rho \in (0, 1)$ be the per-generation probability that a child inherits the parent's party. The transition matrix

$$\mathbf{P} = \begin{pmatrix} \rho & 1-\rho \\ 1-\rho & \rho \end{pmatrix}$$

encodes the dynamic. Starting from $\mu_0 = (1, 0)$ (a fully Democratic founding generation), after t generations $\mu_t = \mu_0 \mathbf{P}^t$. The two-party Markov chain is the simplest non-trivial setting for the questions the handout pursues: long-run distribution, rate of convergence, mixing.

[1] The choice of row vectors and right-multiplication by $\mathbf{P}$ is the standard convention for stochastic dynamics — the rows are probability distributions, and $\mathbf{P}$ acts on them on the right. Some references adopt the column-vector convention, with $\mathbf{P}$ acting on the left; the substance is identical, but the eigenvalue conventions of the spectral theorem (handout #19) match the row-vector convention here when one works with left eigenvectors of $\mathbf{P}$.

3 Reachability, recurrence, and absorbing states

Several substantive political-economy questions about a stochastic dynamic turn out, on inspection, to be questions about the chain's reachability structure. Can a voter's belief reach any value, or are some belief-states unreachable from others? Does a regime trajectory have absorbing states from which there is no recovery — collapse, locked-in autocracy, a hardened partisan affiliation, a radicalized ideological commitment? Are there cyclic patterns in legislator turnover or electoral-cycle dynamics that prevent convergence to a long-run distribution? Each of these maps onto a question in the formal classification of Markov-chain states, and this section develops the organizing concepts — communicating classes, absorbing classes, recurrence, and periodicity — that the rest of the handout will use to characterize long-run behavior.

Definition 4. State j is *accessible* from i, written $i \to j$, if $P_{ij}^{(n)} > 0$ for some $n \geq 0$. States i and j *communicate*, written $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Communication is an equivalence relation on S: reflexivity is the convention $P_{ii}^{(0)} = 1$; symmetry is by definition; transitivity follows from Chapman–Kolmogorov.

Definition 5. The chain is *irreducible* if all states communicate, i.e., S is a single communicating class. A class C is *closed* if $\sum_{j \in C} P_{ij} = 1$ for every $i \in C$ (no escape from C). A state i is *absorbing* if $P_{ii} = 1$ (so $\{i\}$ is itself a closed class).

Definition 6. The *period* of state i is $d(i) = \gcd\{n \geq 1 : P_{ii}^{(n)} > 0\}$, with $d(i) = \infty$ if the set is empty. State i is *aperiodic* if $d(i) = 1$.

Periodicity is a class property: states in the same communicating class share a period. The chain itself is called *aperiodic* if all its states are aperiodic. A standard sufficient condition for aperiodicity is the existence of a self-loop $P_{ii} > 0$ at any state i; in applied political-economy models this is usually trivial (a voter can stay with her current party affiliation; a regime can stay at its current legitimacy level).

Definition 7. State i is *recurrent* if $\mathbb{P}(\text{the chain returns to } i \mid X_0 = i) = 1$, and *transient* otherwise. Equivalently, i is recurrent iff $\sum_{n \geq 0} P_{ii}^{(n)} = \infty$ and transient iff this sum is finite.

The equivalence is a standard counting argument: the expected number of visits to i starting from i equals $\sum_n P_{ii}^{(n)}$, and this is infinite exactly when the return probability is 1. Recurrence is also a class property.

Theorem 8 (Recurrence in finite chains)**.** *In a finite Markov chain, every state in a closed communicating class is recurrent. Equivalently, transient states all lie in non-closed classes (classes from which the chain can escape with positive probability and never return).*

Proof sketch. The chain visits some state infinitely often (pigeonhole on a finite state space and infinite time), and the recurrent states form the closed communicating classes. States in non-closed classes are transient: with positive probability the chain escapes the class on each visit, and once it leaves it never returns. □

Example 9 (Opinion dynamics with absorbing extremes)**.** Consider a voter's ideological position $X_t \in \{\text{far-left}, \text{moderate-left}, \text{moderate-right}, \text{far-right}\}$ evolving across periods, with the two

extremes absorbing (a voter who reaches a far position commits and never moderates) and moderate states freely communicating with each other and stochastically transitioning to the extremes. The closed classes are {far-left} and {far-right}, both absorbing. The moderate states form a non-closed class and are therefore transient. Long-run prediction: every voter eventually commits to an extreme, with probabilities depending on initial position and on the relative pull of each extreme. The political-economy reading is that any model with absorbing extremes implies eventual full polarization, regardless of the moderation rate, as long as transitions to extremes have positive probability.

4 Stationary distributions and convergence

When a working political scientist says "in the long run, the partisan share stabilizes at $X\%$" or "opinion eventually settles into its equilibrium distribution" or "the share of legislators in each ideological cluster converges," she is making a *stationary-distribution claim* about an underlying stochastic dynamic. The right structural object is the distribution that the chain leaves invariant under its own dynamics; the right structural result is the convergence theorem that pins down when the chain reaches that distribution from arbitrary initial conditions, and how fast. Both follow directly from the Perron–Frobenius machinery of the eigenvalue handout (#19) applied to the transition matrix, and the spectral gap of #19 is exactly what governs the rate.

Definition 10. A probability distribution π on S is *stationary* for $\mathbf{P}$ if $\pi\mathbf{P} = \pi$, i.e., $\sum_i \pi_i P_{ij} = \pi_j$ for every $j \in S$.

In matrix terms, π is a left eigenvector of $\mathbf{P}$ with eigenvalue 1, normalized to be a probability distribution. The Perron–Frobenius theorem of the eigenvalue handout (#19) is the structural input that pins down when such a π exists, when it is unique, and when the chain converges to it from any starting point.

Theorem 11 (Existence, uniqueness, and convergence)**.** *Let* $\mathbf{P}$ *be the transition matrix of a finite-state Markov chain.*

(a) If the chain is irreducible, $\mathbf{P}$ *has a unique stationary distribution* π*, and* $\pi_i > 0$ *for every* $i \in S$*.*

(b) If the chain is irreducible and aperiodic, then for every initial distribution μ_0*,*

$$\mu_0\mathbf{P}^n \to \pi \quad \text{as } n \to \infty,$$

in the sense that $(\mu_0\mathbf{P}^n)(j) \to \pi_j$ *for every* $j \in S$*.*

Proof sketch. Both claims are corollaries of Perron–Frobenius. $\mathbf{P}$ is row-stochastic, so 1 is an eigenvalue (with right eigenvector the all-ones vector $\mathbf{1}$, since $\mathbf{P1} = \mathbf{1}$). Irreducibility is the Perron–Frobenius hypothesis that the non-negative matrix has the dominant eigenvalue simple, with a strictly-positive left eigenvector π that we normalize to a probability distribution. Aperiodicity is the additional condition that all other eigenvalues are strictly smaller in modulus than 1; this is what makes $\mathbf{P}^n$ converge (rather than oscillate among multiple equal-modulus eigenvalues at 1). Then $\mathbf{P}^n$ converges to the rank-one matrix $\mathbf{1}\pi^\top$, and $\mu_0\mathbf{P}^n \to \mu_0\mathbf{1}\pi^\top = \pi$ since $\mu_0\mathbf{1} = 1$ for any probability distribution. □

The rate of convergence is governed by the second-largest eigenvalue of $\mathbf{P}$ in modulus, written λ_2 with $|\lambda_2| < 1$ under aperiodicity. The *spectral gap* is $1 - |\lambda_2|$; a larger spectral gap means faster convergence to stationarity.

Proposition 12 (Geometric convergence rate)**.** *Under the hypotheses of Theorem 11(b), there is a constant $C > 0$ depending only on $\mathbf{P}$ such that for every initial distribution μ_0 and every $n \geq 0$,*

$$\|\mu_0 \mathbf{P}^n - \pi\|_{\mathrm{TV}} \leq C \cdot |\lambda_2|^n,$$

where $\|\nu - \pi\|_{\mathrm{TV}} = \frac{1}{2} \sum_{i \in S} |\nu_i - \pi_i|$ is the total-variation distance between probability distributions.

Example 13 (Generational party-affiliation chain, revisited)**.** For the symmetric two-party chain of Example 3 with persistence $\rho \in (0, 1)$, solving $\pi\mathbf{P} = \pi$ with $\pi_D + \pi_R = 1$ gives $\pi = (\frac{1}{2}, \frac{1}{2})$ regardless of ρ. The eigenvalues of $\mathbf{P}$ are 1 and $\lambda_2 = 2\rho - 1$, so the spectral gap is $1 - |2\rho - 1| = \min(2\rho, 2(1-\rho))$. When ρ is close to 1 (high persistence), the gap is small and convergence to 50–50 is slow; when ρ is close to 0 (alternation), the gap is also small (but note $\lambda_2 < 0$, so the chain oscillates as it converges); when $\rho = \frac{1}{2}$ (full mixing each generation), $\lambda_2 = 0$ and the chain reaches stationarity in one step. The political-economy reading is that high-persistence dynamics imply slow long-run convergence, and the rate of convergence to the stationary share is what determines whether one observes long deviations from the long-run prediction in finite samples.

The convergence theorem extends naturally beyond the finite-state setting, and the general-state-space version is the structural foundation of one of the most-used computational technologies in modern political methodology.[2]

5 Reversibility, detailed balance, and mixing

How long does it take for the long-run prediction to be approximately right? When a polling organization treats today's snapshot as informative about a long-run partisan distribution, or when a measurement model built on social-influence dynamics is fit to a panel of survey waves, the validity of the inference depends on the underlying chain having *mixed* — having reached approximate stationarity. The technical name for "how long is long enough" is the *mixing time*, governed by a structural property of the chain known as the spectral gap. A closely related question, with its own substantive reading in political-economy modeling, is whether the dynamic is symmetric in time: does the process look the same running forward and backward conditional on starting at stationarity? This property is *reversibility*, formalized by the *detailed-balance* condition, and it is the technical handle that gives the cleanest mixing-time bounds and that underlies the entire computational machinery of Markov-chain Monte Carlo (forward-referenced in the §4 footnote).

[2] On a countably infinite state space the theorem holds with an additional condition (*positive recurrence*: the expected return time to any state is finite); on a general state space (uncountable) one needs the *Doeblin condition* or *Harris recurrence* to control the chain's escape behavior. The general-state-space convergence theory is the structural backbone of *Markov-chain Monte Carlo*, the algorithm class that generates samples from a target probability distribution by constructing a chain whose stationary distribution is the target. MCMC is the most-used applied technology in modern Bayesian political methodology — ideal-point estimation (the analyst's posterior over W-NOMINATE-style scaling), measurement models (DW-NOMINATE, IRT, structural topic models), survey-experimental analysis with hierarchical priors, and Bayesian-game model estimation all rest on it. The correctness of MCMC at this level of generality is exactly the convergence theorem of this section, applied in the general-state-space setting; the practitioner's responsibility is to verify the constructed chain is irreducible, aperiodic, and reversible with respect to the target π (§5), and the theorem then guarantees the sample averages converge to expectations under π. Meyn and Tweedie (2009) is the canonical reference on general-state-space convergence; Robert and Casella (2004) treats MCMC algorithmically.

Definition 14 (Detailed balance)**.** A probability distribution π on S satisfies *detailed balance* for $\mathbf{P}$ if

$$\pi_i P_{ij} = \pi_j P_{ji} \qquad \text{for every } i, j \in S.$$

The chain is *reversible* (with respect to π) if π satisfies detailed balance.

Proposition 15. *If π satisfies detailed balance for $\mathbf{P}$, then π is stationary.*

Proof. $(\pi\mathbf{P})_j = \sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$, where the second equality uses detailed balance and the last uses row-stochasticity. □

The converse can fail: a chain can have a stationary distribution without satisfying detailed balance. The structural difference is that detailed balance demands a pointwise symmetry — the flow $\pi_i P_{ij}$ from i to j matches the flow $\pi_j P_{ji}$ from j to i — whereas stationarity demands only that the total flow into each state j equals the total flow out. Reversible chains are a strict subset of stationary chains.

Example 16 (Random walk on an undirected graph)**.** Let $G = (V, E)$ be a finite connected undirected graph with no isolated vertices, and consider the random walk where from vertex i the chain moves to a uniformly chosen neighbor. Then $P_{ij} = 1/\deg(i)$ if $\{i, j\} \in E$ and 0 otherwise. The distribution $\pi_i = \deg(i)/(2|E|)$ satisfies detailed balance: $\pi_i P_{ij} = (\deg(i)/2|E|) \cdot (1/\deg(i)) = 1/2|E|$, which is symmetric in i and j. The chain is reversible. Reversibility is closely tied to undirected structure: the same chain on a directed graph (where edges $i \to j$ and $j \to i$ may have different weights, or only one direction may exist) is generally not reversible.

Definition 17 (Mixing time)**.** For an irreducible aperiodic chain with stationary distribution π and any $\epsilon > 0$, the *mixing time* is

$$\tau_{\text{mix}}(\epsilon) = \min\left\{t \geq 0 : \max_{\mu_0} \left\|\mu_0 \mathbf{P}^t - \pi\right\|_{\text{TV}} \leq \epsilon\right\}.$$

The mixing time records how many steps the chain needs to bring its distribution within ϵ in total-variation distance of π, in the worst case over starting distributions. It is the natural quantitative complement to the qualitative convergence theorem: the convergence theorem says we get there, the mixing time says when.

Theorem 18 (Mixing time bound for reversible chains)**.** *For a reversible, irreducible, aperiodic chain with stationary π on a finite state space,*

$$\tau_{\text{mix}}(\epsilon) \leq \frac{1}{1 - |\lambda_2|} \log \frac{1}{\epsilon \pi_{\min}},$$

where λ_2 is the second-largest eigenvalue in modulus of $\mathbf{P}$ and $\pi_{\min} = \min_{i \in S} \pi_i$.

The bound makes the structural role of the spectral gap explicit: the smaller the gap $1 - |\lambda_2|$, the longer the chain takes to mix. Reversibility is what licenses the spectral-theorem-based argument behind the bound: a reversible chain's transition matrix is similar to a symmetric matrix (via the diagonal change-of-basis $\text{diag}(\sqrt{\pi_i})$), so the spectral theorem applies and the convergence rate is governed by the gap directly.

The political-economy reading. Mixing-time arguments are most commonly seen in formal political-science settings as bounds on how quickly a poll or survey can detect long-run distributional

facts about an opinion-dynamics or voter-affiliation model. If the underlying dynamics have a small spectral gap, the chain mixes slowly and a large sample is needed to overcome the chain's autocorrelation; if the gap is large, the chain mixes quickly and short samples suffice. The mixing-time machinery also underlies MCMC (forward-referenced in the §4 footnote): sampling efficiency is governed by the constructed chain's mixing time, and the practitioner's job is to construct chains whose spectral gaps can be bounded.

6 Markov decision processes

Most political-economy dynamic-decision problems have stochastic transitions. A voter's belief about a candidate's competence evolves as new electoral signals arrive; an authoritarian regime's legitimacy stock evolves under random mobilization shocks; a long-lived government's optimal-taxation problem responds to stochastic shocks to the macroeconomy; a campaign's vote share fluctuates under stochastic turnout, advertising response, and news shocks. In each setting the agent's action at state s does not pin down tomorrow's state but rather the *distribution* over tomorrow's states. The dynamic-optimization handout (#23) developed the deterministic version of dynamic programming with $f(s, a)$ as the next-period state; the natural generalization for stochastic dynamics is the *Markov decision process* (MDP), a Markov chain with decisions, in which the agent's action plus the current state determines a probability distribution over next-period states and the agent maximizes expected discounted total reward.

Definition 19 (MDP)**.** A *Markov decision process* consists of:

- a state space S (finite or countable),
- an action correspondence $A : S \rightrightarrows A^*$ specifying the feasible actions $A(s) \subseteq A^*$ at each state,
- a transition kernel $P(s' \mid s, a)$, a probability distribution over S for each (s, a),
- a per-period reward function $r : S \times A^* \to \mathbb{R}$,
- a discount factor $\beta \in [0, 1)$.

A *stationary policy* is a function $\pi : S \to A^*$ with $\pi(s) \in A(s)$ for every s. Under policy π, the state evolves as a time-homogeneous Markov chain with transition probabilities $P^\pi(s' \mid s) = P(s' \mid s, \pi(s))$.

The value of a policy from a starting state is the expected discounted reward:

$$V^\pi(s) = \mathbb{E}^\pi \left[\sum_{t=0}^{\infty} \beta^t r(s_t, \pi(s_t)) \,\middle|\, s_0 = s \right],$$

where $\mathbb{E}^\pi$ is the expectation under the chain induced by π. The agent's task is to choose a policy to maximize the value:

$$V^*(s) = \sup_\pi V^\pi(s).$$

Under bounded rewards and the technical hypotheses of §3 of the dynamic-optimization handout (essentially, $A(s)$ compact, r and P continuous in a for each s), the supremum is attained by a stationary policy.

Theorem 20 (Stochastic Bellman equation)**.** *The optimal value function V^* satisfies*

$$V^*(s) = \max_{a \in A(s)} \left[r(s,a) + \beta \sum_{s' \in S} P(s' \mid s,a) V^*(s') \right],$$

and a stationary policy π^* is optimal if and only if for every s,

$$\pi^*(s) \in \arg\max_{a \in A(s)} \left[r(s,a) + \beta \sum_{s'} P(s' \mid s,a) V^*(s') \right].$$

The structural difference from the deterministic Bellman equation of handout #23 is exactly one substitution: the deterministic continuation value $V^*(f(s,a))$ becomes the conditional-expected continuation value $\sum_{s'} P(s' \mid s,a) V^*(s') = \mathbb{E}[V^*(s_{t+1}) \mid s_t = s, a_t = a]$. The recursive structure, the FOC characterization of optimal actions, and the entire downstream apparatus carry through.

The Bellman operator in the stochastic setting is

$$(TV)(s) = \max_{a \in A(s)} \left[r(s,a) + \beta \sum_{s'} P(s' \mid s,a) V(s') \right],$$

acting on bounded functions $V : S \to \mathbb{R}$ with the sup norm $\|V\|_\infty$.

Theorem 21 (Stochastic Bellman operator is a contraction)**.** *Suppose r is bounded on $S \times A^*$. Then T is a contraction with modulus β on the space of bounded functions on S with the sup norm.*

Proof. For any bounded V_1, V_2 and any s, let $a_1^* \in \arg\max_a [r(s,a) + \beta \sum_{s'} P(s' \mid s,a) V_1(s')]$. Then

$$\begin{aligned}
(TV_1)(s) - (TV_2)(s) &\le r(s,a_1^*) + \beta \sum_{s'} P(s' \mid s,a_1^*) V_1(s') - \left[r(s,a_1^*) + \beta \sum_{s'} P(s' \mid s,a_1^*) V_2(s') \right] \\
&= \beta \sum_{s'} P(s' \mid s,a_1^*) (V_1(s') - V_2(s')) \\
&\le \beta \|V_1 - V_2\|_\infty \sum_{s'} P(s' \mid s,a_1^*) \\
&= \beta \|V_1 - V_2\|_\infty .
\end{aligned}$$

By symmetry $|(TV_1)(s) - (TV_2)(s)| \le \beta \|V_1 - V_2\|_\infty$ for every s, hence $\|TV_1 - TV_2\|_\infty \le \beta \|V_1 - V_2\|_\infty$. □

By Banach's fixed-point theorem (handout #23), T has a unique fixed point V^* on the space of bounded functions, and $V_{n+1} = TV_n$ converges to V^* from any starting V_0 at the geometric rate β. The slogan from #23 carries: discounting turns the Bellman operator into a contraction, regardless of whether transitions are deterministic or stochastic.

A second standard algorithm is *policy iteration*: starting from any policy π_0, compute its value V^{π_0} (by solving the linear system $V^{\pi_0}(s) = r(s, \pi_0(s)) + \beta \sum_{s'} P(s' \mid s, \pi_0(s)) V^{\pi_0}(s')$), then update π_1 to be the maximizer of the right-hand side of the Bellman equation given V^{π_0}, and iterate. Policy iteration converges in finitely many steps on a finite state and action space, since the policy improves monotonically and there are only finitely many policies.

Example 22 (Authoritarian-regime survival as MDP)**.** Continuing the regime-survival example from the dynamic-optimization handout's Exercise 10. The state $s_t \in \{0, 1, 2, \ldots, S_{\max}\}$ is the regime's legitimacy stock; the action $a_t \in [0, 1]$ is the level of repression. The regime survives to the next period with probability $p(s, a) = s(1 - q(a))/S_{\max}$, where q is increasing in a (more repression suppresses mobilization but signals weakness). The legitimacy stock evolves stochastically: with probability $p(s, a)$ the regime survives and the new stock is $\min(s+1, S_{\max})$ (legitimacy accumulates with successful periods), with probability $1 - p(s, a)$ the regime collapses (an absorbing state with reward 0). The per-period reward conditional on survival is $r(s, a) = s - \kappa a$. The Bellman equation is

$$V^*(s) = \max_{a \in [0,1]} \left[s - \kappa a + \beta\, p(s, a)\, V^*(\min(s + 1, S_{\max})) \right],$$

with $V^*(\text{collapsed}) = 0$. The optimal repression policy $\pi^*(s)$ trades off the immediate cost κ against the survival-probability gain $\beta p(s, a)V^*(s+1)$; the stochastic-Bellman framework is what makes "the regime balances cost and survival" into a precise comparative-statics statement. Comparative-static questions like "does a more patient regime repress more or less?" (effects of changes in β) are the kinds of questions the framework is designed to answer.

The single-agent MDP framework generalizes naturally to multi-agent strategic settings in which each agent solves an MDP given the others' policies, and the resulting equilibrium concept is one of the most-used in dynamic political economy.[3]

7 What's next

This handout opens a single-handout cluster on Markov chains and Markov decision processes, slotted at #24 between dynamic optimization (#23) and the applications clusters that follow. Three strands extend it.

General-state-space Markov chains and MCMC. The convergence theorem extends from finite to countable state spaces under positive recurrence, and to general (uncountable) state spaces under the Doeblin or Harris-recurrence conditions. The general-state-space theory is the structural backbone of Markov-chain Monte Carlo, the dominant computational technology in modern Bayesian political methodology — the structural correctness of MCMC at this level of generality is exactly the convergence theorem of §4 applied to a constructed reversible chain whose stationary distribution is the target. Meyn and Tweedie (2009) is the canonical reference; Robert and Casella (2004) treats MCMC algorithmically.

Continuous-time Markov chains. An alternative to discretizing time is to work in continuous time, with the chain specified by a generator matrix $\mathbf{Q}$ and the dynamics governed by the Kolmogorov

[3] *Stochastic dynamic games* generalize MDPs to multiple decision-makers each solving an MDP given the others' policies; the equilibrium concept is *Markov-perfect equilibrium* (MPE), in which every agent's policy is a best response to the others' policies and depends only on the payoff-relevant state (not on the calendar time or the entire history). Maskin and Tirole (2001) is the canonical reference on MPE in a political-economy-relevant setting (capital accumulation under strategic interaction), and the concept appears throughout dynamic political economy: dynamic legislative bargaining (Baron and Ferejohn 1989 as the static benchmark), dynamic political agency, dynamic models of authoritarian survival under strategic citizen mobilization, and electoral competition with state variables (e.g., incumbency advantage) all use Markov-perfect equilibrium. The forthcoming game-theory cluster will develop the equilibrium concept; for present purposes the key observation is that MPE inherits the contraction-mapping structure of the single-agent MDP only conditionally on the other agents' policies, so existence and uniqueness of equilibrium are subtler than in single-agent MDPs and typically require additional structural conditions (e.g., supermodularity, strategic complementarities).

forward / backward equations $\frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q} = \mathbf{Q}\mathbf{P}(t)$. Applications include some demographic-political models (random arrival of voters at the polls), regime-change dynamics with random arrival times, and dynamic-game models with continuous-time strategic interaction. Norris (1998) is the standard reference.

Markov-perfect equilibria. The forthcoming game-theory cluster will develop MPE as the strategic generalization of MDPs. Each player solves an MDP given the other players' policies, and the equilibrium policies must be mutual best responses. The contraction-mapping argument of §6 carries over conditional on the others' policies; existence and uniqueness in multi-agent settings typically need additional structure (supermodularity, monotone best-response dynamics).

For graduate-level treatments at this handout's level of abstraction: Norris (1998) is the canonical undergraduate-to-graduate reference for the discrete-time finite-state-space machinery of §2–§5; Levin, Peres, and Wilmer (2017) is the standard reference on mixing times and the modern view of the spectral-gap-and-mixing connection; Stokey, Lucas, and Prescott (1989, Ch. 8–12) treats MDPs with the dynamic-economic-modeling angle; Bertsekas (2017) treats MDPs algorithmically with the engineering / operations-research framing.

8 Exercises

Exercise 23. *Two-state chain by hand.* Let $\mathbf{P} = \left(\begin{smallmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{smallmatrix}\right)$. (a) Compute $\mathbf{P}^2$ and $\mathbf{P}^3$ by direct multiplication. (b) Verify Chapman–Kolmogorov by computing $P_{12}^{(2)}$ via direct calculation and via $\sum_k P_{1k}^{(1)} P_{k2}^{(1)}$. (c) Find the eigenvalues of $\mathbf{P}$ and read off λ_2.

Exercise 24. *Classification.* Consider a four-state chain with $S = \{1, 2, 3, 4\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Identify the communicating classes. For each class: is it closed? Recurrent or transient? What is the period? Which states are absorbing, if any?

Exercise 25. *Stationary distribution by linear algebra.* Let $\mathbf{P} = \frac{1}{4}\left(\begin{smallmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{smallmatrix}\right)$ on $S = \{1, 2, 3\}$. (a) Verify that the chain is irreducible and aperiodic. (b) Find π by solving $\pi\mathbf{P} = \pi$ subject to $\sum_i \pi_i = 1$. (c) Find the eigenvalues of $\mathbf{P}$ and identify λ_2.

Exercise 26. *Detailed balance verification.* For the chain in Exercise 25, verify that the stationary distribution satisfies detailed balance. Why is this expected from the symmetry of $\mathbf{P}$?

Exercise 27. *Random walk on the path graph.* Consider the random walk on the path graph $1 - 2 - 3 - 4 - 5$, where from vertex i the chain moves uniformly to a neighbor (the endpoints 1 and 5 have one neighbor each; the interior vertices have two). (a) Write down $\mathbf{P}$. (b) Find the stationary distribution π via the formula $\pi_i = \deg(i)/2|E|$ from Example 16. (c) Verify directly that π satisfies detailed balance.

Exercise 28. *Generational party-affiliation chain.* Continuing Example 3: let $\rho = 0.7$ (a parent's party persists with probability 0.7, defects to the other party with probability 0.3). (a) Find the stationary distribution. (b) Compute λ_2 and the spectral gap. (c) Estimate $\tau_{\text{mix}}(0.05)$ using the spectral-gap-based bound. (d) Discuss in two sentences how the answer in (a) would change if persistence were asymmetric ($P_{DD} \neq P_{RR}$). What stays the same?

Exercise 29. *Opinion dynamics with absorbing extremes.* Let $S = \{FL, ML, MR, FR\}$ (far-left, moderate-left, moderate-right, far-right), with FL and FR absorbing. From ML the voter moves to FL with probability α, stays at ML with probability $1 - \alpha - \gamma$, moves to MR with probability γ. Symmetrically from MR. (a) Identify the communicating classes and the recurrent and transient states. (b) Set up (without solving) the linear system for $h_i = \mathbb{P}(\text{absorbed at } FL \mid X_0 = i)$ for $i \in \{ML, MR\}$. (c) Discuss in two sentences what the model implies about long-run polarization as a function of the parameter α, the per-period probability of radicalization at the moderate states.

Exercise 30. *Small MDP by value iteration.* Consider an MDP with $S = \{1, 2\}$, $A = \{a, b\}$, $\beta = 0.9$, and rewards / transitions

	$r(s,a)$	$r(s,b)$	$P(s'=1 \mid s,a)$	$P(s'=1 \mid s,b)$
$s=1$	1	0	0.6	0.9
$s=2$	0	2	0.3	0.1

(with $P(s' = 2 \mid \cdot) = 1 - P(s' = 1 \mid \cdot)$). Starting from $V_0 \equiv 0$, perform two iterations of value iteration $V_{n+1} = TV_n$ and identify the optimal action at each state under V_2.

Exercise 31. *Authoritarian-regime survival, continued.* Continuing Example 22 with a small parameterization: let $S = \{0, 1, 2\}$, $A = \{0, 1/2, 1\}$, $\kappa = 0.5$, $\beta = 0.8$, $q(a) = 1 - a$ (so survival probability $p(s, a) = sa/2$). State 0 is collapsed (absorbing, $V^*(0) = 0$). (a) Write the Bellman equation as a system of two equations in $V^*(1), V^*(2)$. (b) Solve numerically (or by guess-and-iterate) for V^* and π^*. (c) Discuss in two or three sentences what happens to the optimal repression at $s = 1$ as β rises from 0.8 to 0.95 — and which interpretation of β (patience vs. expected longevity) gives the more politically interesting comparative static.

Exercise 32. *When the Markov property fails.* A standard model of voter learning treats the voter's belief $s_t \in [0, 1]$ as the state and updates it by Bayes' rule each period. Argue that this is Markovian: the belief at $t + 1$ depends on the belief at t and the period-t signal, with no further dependence on the past. Now suppose we drop the assumption that the voter's prior is correct and let her also update her trust in her own model. Argue informally that the original "state = belief" model is no longer Markovian, and identify the structural fix (enrich the state space). The exercise illustrates the modeling-theoretic point: the Markov property is a constraint on the choice of state space, and a non-Markovian dynamics on a coarse state space can usually be made Markovian by enriching the state.

References

Baron, David P. and John A. Ferejohn (1989). "Bargaining in Legislatures". In: *American Political Science Review* 83.4, pp. 1181–1206.

Bertsekas, Dimitri P. (2017). *Dynamic Programming and Optimal Control*. 4th ed. Belmont, MA: Athena Scientific.

Levin, David A., Yuval Peres, and Elizabeth L. Wilmer (2017). *Markov Chains and Mixing Times*. 2nd ed. Providence, RI: American Mathematical Society.

Maskin, Eric and Jean Tirole (2001). "Markov Perfect Equilibrium I: Observable Actions". In: *Journal of Economic Theory* 100.2, pp. 191–219.

Meyn, Sean and Richard L. Tweedie (2009). *Markov Chains and Stochastic Stability*. 2nd ed. Cambridge: Cambridge University Press.

Norris, J. R. (1998). *Markov Chains*. Cambridge: Cambridge University Press.

Robert, Christian P. and George Casella (2004). *Monte Carlo Statistical Methods*. 2nd ed. New York: Springer.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.