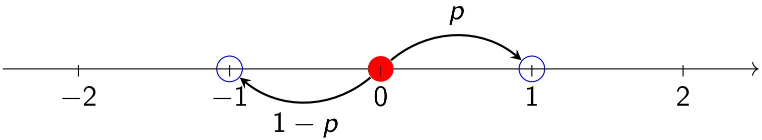
Mathematics and numerics for data assimilation and state estimation – Lecture 5

Summer semester 2020

Summary of lecture 4

- Random walks on $\mathbb{Z}^d$: described by distribution of X_0 and its iid steps $\{\Delta X_n\}$.



- For an RW $\{X_0, \Delta X_0, \Delta X_1, \ldots\}$, on $\mathbb{Z}^d$, a state $s \in \mathbb{Z}^d$ is recurrent if by setting $X_0 = s$, we obtain that

$$\mathbb{P}(X_n = s \quad \text{for infinitely many } n) = 1.$$

The last condition is equivalent to (Thm 9, Lecture 4),

$$\mathbb{P}(T < \infty) = 1 \quad \text{for} \quad T = \inf\{n \geq 1 \mid X_n = X_0\}.$$

- Convergence of random variables (Chebychev's inequality, weak law of large numbers, mean-square convergence).

Plan for this lecture

- The Markov property – memorylessness
- Markov chains
- Invariant distributions

Markov chains

We consider the dynamics of a discrete-time stochastic process $\{Z_n\}$ that takes values on a state-space $\mathbb{S}$ that is discrete; meaning it is either finite, e.g. $\mathbb{S} = \{1, 2, 3\}$, or countable, e.g. $\mathbb{S} = \mathbb{Z}^d$.

Definition 1 (Markov chain)

A sequence $\{Z_n\}_{n\geq 0}$ of $\mathbb{S}$-valued rv is a discrete-time (and discrete-space) Markov chain if

1. it is equipped with an initial distribution $\pi^0(z) := \mathbb{P}(Z_0 = z)$, and
2. satisfies the so-called Markov property ("memorylessness")

$$\mathbb{P}(Z_{n+1} = z_{n+1} \mid Z_n = z_n, \ldots, Z_0 = z_0) = \mathbb{P}(Z_{n+1} = z_{n+1} \mid Z_n = z_n) \quad (1)$$

holds for any $n \geq 0$ and $z_0, \ldots, z_n \in \mathbb{S}$ for which

$$\mathbb{P}(Z_n = z_n, \ldots, Z_0 = z_0) > 0.$$

Alternative statement of the Markov property

To avoid the provided $\mathbb{P}(Z_n = z_n, \ldots, Z_0 = z_0) > 0$, one may state the Markov property as follows:

$$\begin{aligned}&\mathbb{P}(Z_{n+1} = z_{n+1}, Z_n = z_n, \ldots, Z_0 = z_0) \\ &\qquad = \mathbb{P}(Z_{n+1} = z_{n+1} \mid Z_n = z_n)\, \mathbb{P}(Z_n = z_n, \ldots, Z_0 = z_0)\,. \end{aligned} \tag{2}$$

Note also that $\sum_{z \in \mathbb{S}} \pi^0(z) = 1$.

Example 2

Any random walk $\{Z_n\}$ on $\mathbb{S} = \mathbb{Z}^d$ is a Markov chain. Since $Z_{n+1} = Z_n + \Delta Z_n$ with $\{\Delta Z_n\}$ iid, it follows that

$$
\begin{aligned}
&\mathbb{P}(Z_{n+1} = z_{n+1} \mid Z_n = z_n, \ldots, Z_0 = z_0) \\
&= \mathbb{P}(Z_n + \Delta Z_n = z_{n+1} \mid Z_n = z_n, \ldots, Z_0 = z_0) \\
&= \\
&=
\end{aligned}
$$

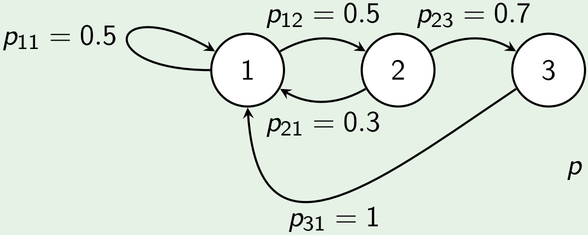
provided $\mathbb{P}(Z_n = z_n, \ldots, Z_0 = z_0) > 0$.

Example 3 (Three-state chain)

Consider a Markov chain $\{Z_n\}$ on $\mathbb{S} = \{1, 2, 3\}$. For any $n \geq 0$, let

$$p_{ij} := \mathbb{P}\left(Z_{n+1} = i \mid Z_n = j\right)$$

with dynamics described by the below transition graph



$$p = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0 & 0.7 \\ 1 & 0 & 0 \end{pmatrix}$$

Simplifying notation and terminology

- For $0 \leq k \leq n$ and points $z_k, \ldots, z_n \in \mathbb{S}$ let

$$z_{k:n} := (z_k, \ldots, z_n), \quad (\text{so} \quad z_{n:n} = z_n).$$

- Similarly, for the Markov chain, let

$$Z_{k:n} := (Z_k, \ldots, Z_n).$$

In the new notation, the Markov property (1) becomes

$$\mathbb{P}\left(Z_{n+1} = z_{n+1} \mid Z_{0:n} = z_{0:n}\right) = \mathbb{P}\left(Z_{n+1} = z_{n+1} \mid Z_n = z_n\right)$$

whenever $\mathbb{P}(Z_{0:n} = z_{0:n}) > 0$.

Product decomposition of joint Markov-chain distributions

Definition 4

A transition function is a mapping $p : \mathbb{S} \times \mathbb{S} \to [0,1]$ satisfying the constraint

$$\sum_{z \in \mathbb{S}} p(c,z) = 1 \quad \text{for any} \quad c \in \mathbb{S}. \tag{3}$$

The $n+1$-st **transition function** of a Markov chain $\{Z_n\}$ is for all $z, c \in \mathbb{S}$ defined by

$$p_{n+1}(c,z) := \begin{cases} \mathbb{P}(Z_{n+1} = z \mid Z_n = c) & \text{if } \mathbb{P}(Z_n = c) > 0 \\ \mathbb{1}_{\{c\}}(z) & \text{otherwise.} \end{cases}$$

Note: for all $c \in \mathbb{S}$ s.t. $\mathbb{P}(Z_n = c) > 0$, the definition of $p_{n+1}(c, \cdot)$ is unique, but for zero-probability outcomes c, whatever definition satisfying (3) is valid.

Verification of constraint?

Application of the transition function

By the Markov property, we obtain

$$\begin{aligned}\mathbb{P}\left(Z_{0:n} = z_{0:n}\right) &= \mathbb{P}(Z_{0:n-1} = z_{0:n-1})\mathbb{P}\left(Z_n = z_n \mid Z_{n-1} = z_{n-1}\right) \\ &= \mathbb{P}(Z_{0:n-1} = z_{0:n-1})p_n(z_{n-1}, z_n)\end{aligned}$$

where two cases must be taken into account:

1. if $\mathbb{P}\left(Z_{n-1} = z_{n-1}\right) > 0$ then this follows from definition, and
2. if $\mathbb{P}\left(Z_{n-1} = z_{n-1}\right) = 0$ then the euqality still holds as it becomes $0 = 0$.

By recursive application,

$$\mathbb{P}\left(Z_{0:n} = z_{0:n}\right) = \tag{4}$$

Definition 5 (Time-homogeneity)

A Markov chain is time-homogeneous if there exists a transition function p that is independent of time n, such that

$$\mathbb{P}(Z_{n+1} = z \mid Z_n = c) = p(c, z)$$

whenever $\mathbb{P}(Z_n = c) > 0$.

We say that $\{Z_n\}$ is *Markov*(π^0, p).

See for instance, the three-state chain Example , where $p(i, j) = p_{ij}$, and π^0 remains to be specified.

Transition probabilities for time-homogeneous Markov chains

For the rest of this lecture, we consider a chain $\{Z_n\}$ that is *Markov*(π^0, p).

- Applying (4) in the time-homogeneous setting yields

$$\mathbb{P}\left(Z_{0:n} = z_{0:n}\right) = \pi^0(z_0) \prod_{i=0}^{n-1} p(z_i, z_{i+1}) \tag{5}$$

- As an extension of the initial state distribution, we introduce for n-th state distribution

$$\pi^n(z_n) := \mathbb{P}(Z_n = z_n)$$

- **Observation:** By marginalization,

$$\pi^n(z_n) = \sum_{z_{0:n-1} \in \mathbb{S}^n} \mathbb{P}\left(Z_{0:n} = z_{0:n}\right)$$

Theorem 6

Let $\{Z_n\}$ *be Markov*(π^0, p) *and for any* $k \geq 2$*, define*

$$p^{*k}(z_0, z_k) = \sum_{z_{1:k-1} \in \mathbb{S}^k} p(z_0, z_1) p(z_1, z_2) \ldots p(z_{k-1}, z_k).$$

Then p^{*k} *is a transition function for* $\{Z_{kn}\}_n$ *and, in particular,*

$$p^{*k}(z_0, z_k) = \mathbb{P}\left(Z_k = z_k \mid Z_0 = z_0\right)$$

whenever $\mathbb{P}(Z_0 = z_0) > 0$.

Verification:

Transition functions and n-th state distributions

Note that

$$\pi^1(z_1) = \sum_{z_0 \in \mathbb{S}} \pi^0(z_0) p(z_0, z_1)$$

and,

$$\begin{aligned}
\pi^n(z_n) &= \sum_{z_{n-1} \in \mathbb{S}} \pi^{n-1}(z_{n-1}) p(z_{n-1}, z_n) \\
&= \dots \\
&= \sum_{z_0 \in \mathbb{S}} \pi^0(z_0) p^{*n}(z_0, z_n).
\end{aligned}$$

For finite state-spaces this can be associated to vector-matrix products.

Corollary 7

Let $\{Z_n\}$ be Markov(π^0, p) on a finite state-space $\mathbb{S} = \{1, 2, \ldots, d\}$ and introduce the notation

$$\pi_i^n := \pi^n(i), \quad p_{ij} := p(i,j) \quad \text{and} \quad p_{ij}^k := p^{*k}(i,j).$$

Then

$$p^k = pp^{k-1} = p^{k-1}p \quad k \geq 2$$

and, with π^n representing a row-vector in $\mathbb{R}^d$,

$$\pi^n = \pi^{n-1}p = \pi^0 p^n \quad n \geq 1.$$

Theorem 8 (Transition probabilities for time-homogeneous Markov chains)

Let $\{Z_n\}$ be Markov(π^0, p). Then for any $m > n \geq 0$ and $z_n, \ldots, z_m \in \mathbb{S}$, it holds that

$$\mathbb{P}\left(Z_{n:m} = z_{n:m}\right) = \pi^n(z_n) \prod_{i=n}^{m-1} p(z_i, z_{i+1})$$

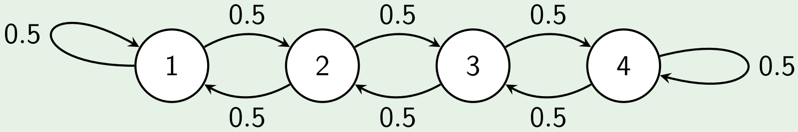
Example 9

Let $\mathbb{S} = \{1, 2, 3, 4\}$ and

$$\pi^0 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} / 4$$

and

$$p = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$



Then $\pi^n = \pi^0$ for all $n \geq 0$.

Invariant distributions

Definition 10

Let π be a probability distribution on $\mathbb{S}$. We call π an **invariant/stationary/equilibrium** distribution for the transition function p if it holds that

$$\pi(z) = \sum_{c \in \mathbb{S}} \pi(c) p(c, z) \quad \forall z \in \mathbb{S},$$

or, in matrix notation, if

$$\pi = \pi p.$$

Note that for $\{Z_n\}$ that is $Markov(\pi^0, p)$ where π^0 is invariant, it holds that

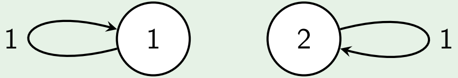
$$Z_0 \stackrel{D}{=} Z_1 \stackrel{D}{=} Z_2 \stackrel{D}{=} \dots$$

How many invariant distributions?

For a finite state-space $\mathbb{S} = \{1, 2, \ldots, d\}$ there is either 1 or infinitely many invariant distributions.

Example 11

$\mathbb{S} = \{1, 2\}$ and $p_{ij} = \mathbb{1}_{\{i\}}(j)$.



Invariant distributions

$$\pi =$$

Theorem 12 (FJK 2.2.33)

Consider $\mathbb{S} = \{1, 2, \ldots, d\}$ and a transition function p. If there exists an $m \geq 1$ such that p^m is strictly positive, then there exists a unique invariant distribution $\pi = (\pi_1, \ldots, \pi_d)$ and

$$\lim_{n\to\infty} \pi_j^n = \pi_j \quad \forall j \in \mathbb{S}$$

and

$$\lim_{n\to\infty} p_{ij}^n = \pi_j \quad \forall i, j \in \mathbb{S}.$$

Meaning **any** *initial distribution π^0 converges to the invariant distribution.*

Observation: If $\lim_{n\to\infty} p_{ij}^n = \pi_j$, then

$$\lim_{n\to\infty} p_{ij}^n = \lim_{n\to\infty} p_{ij}^{n+1} = \ldots$$

Matrix-eigenvalue interpretation of invariant distributions

- π invariant distribution implies that $(\pi, 1)$ is an eigenpair of p since

$$\pi p = \pi 1$$

- Since every row of p sums to 1, $(p - I)[1, 1, \ldots, 1]^T = 0$ meaning 1 is an eigenvalue of p.
- Need to verify that corresponding row-eigenvector π is non-negative (at least one such is (FJK 2.2.39)).
- If (π, λ) is unique eigenpair of p with $\pi \geq 0$ and $\lambda = 1$, then the invariant distribution is unique.
- Otherwise, convex combinations invariant distributions will also be invariant.

Example 13

Let $\mathbb{S} = \{1, 2\}$ and

$$p = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

Eigenvalues

$$\lambda_1 = 1, \quad \lambda_2 = 1/4,$$

with ℓ^1-normalized right-eigenvectors

$$\pi_1 = [1,\ 2]/3, \quad \pi_2 = [1,\ -1]/2.$$

And,

$$\lim_{n\to\infty} p^n = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

Relation to irreducibility

What are sufficient conditions to ensure that for some $m \geq 1$, p^m is strictly positive?

Definition 14

Consider a transition matrix p associated to Markov chains on $\mathbb{S} = \{1, 2, \ldots, d\}$. p is said to be

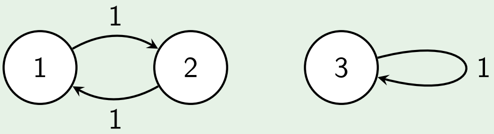
- **irreducible** if for any $i, j \in \mathbb{S}$ there exists an $m \geq 1$ such that $p^m_{ij} > 0$, and
- the i-th state is said to be **aperiodic** if $p^n_{ii} > 0$ for any sufficiently large n.

Lemma 15 (1.8.2, Norris, Markov Chains)

If p is irreducible and has an aperiodic state, then p^m is strictly positive for some $m \geq 1$.

Example 16

$$p = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

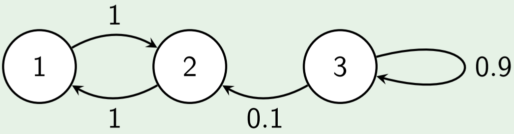


Irreducible?

Aperiodic states?

Example 17

$$p = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$



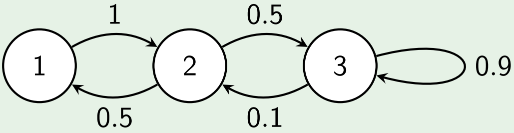
Irreducible?

Aperiodic states?

Example 18

Reducible chain $\mathbb{S} = \{1, 2, 3\}$

$$p = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$



Irreducible?

Aperiodic states?

Recall the result:

Lemma 19 (1.8.2, Norris, Markov Chains)

If p is irreducible and has an aperiodic state, then p^m is strictly positive for some $m \geq 1$.

Proof: Assume the index $i \in \mathbb{S}$ is aperiodic, i.e., $p_{ii}^n > 0$ for all $n \geq N$. For indices $j, k \in \mathbb{S}$, let us show that there exist an m_{jk} such that

$$p_{jk}^{\bar{m}} > 0 \quad \forall \bar{m} \geq m_{jk}.$$

Since p is irreducible, there exists $n_{ji}, n_{ik} \geq 1$ such that

$$p_{ji}^{n_{ji}} > 0 \quad \text{and} \quad p_{ik}^{n_{ik}} > 0.$$

Consequently, for any $\bar{m} \geq n_{ji} + n_{ik} + N$

$$p_{jk}^{\bar{m}} = p_{jk}^{n_{ji}+n_{ik}+\bar{m}-(n_{ji}+n_{ik})} \geq p_{ji}^{n_{ji}} p_{ik}^{n_{ik}} p_{ii}^{\bar{m}-(n_{ji}+n_{ik})} > 0.$$

Recurrence and construction of invariant distributions

Definition 20

Consider an **irreducible** transition function p associated to a state-space $\mathbb{S}$. Then we say that p is recurrent if it for any state $i \in \mathbb{S}$ and Markov chain $\{Z_n^i\} \sim Markov(\mathbb{1}_{\{i\}}, p)$ holds that

$$\mathbb{P}(Z_n^i = i \quad \text{for infinitely many } n) = 1, \tag{6}$$

which for the hitting time $T_i := \inf\{n \geq 1 \mid Z_n^i = i\}$ is equivalent to

$$\mathbb{P}(T_i < \infty) = 1.$$

Lemma 21

If p is **irreducible** *and the state-space is finite, then p is recurrent.*

Proof: Let us write $\mathbb{S} = \{1, 2, \ldots, d\}$. Since $\mathbb{S}$ is finite, there must be at least one pair of states $i, j \in \mathbb{S}$ satisfying

$$\mathbb{P}(Z_n^i = j \quad \text{for infinitely many } n) > 0 \tag{7}$$

since otherwise we reach the contradiction

$$\begin{aligned} 0 &= bP(Z_n^i \notin \mathbb{S} \quad \text{for infinitely many } n) \\ &\geq 1 - \sum_{j \in \mathbb{S}} bP(Z_n^i = j \quad \text{for infinitely many } n) = 1. \end{aligned}$$

And

$$\begin{aligned} &\mathbb{P}(Z_n^j = j \quad \text{for infinitely many } n) \\ &= \mathbb{P}(Z_n^i = j \quad \text{for infinitely many } n \cap \{Z_n^i = j \quad \text{for some } n\}) \\ &= \mathbb{P}(Z_n^i = j \quad \text{for infinitely many } n) > 0. \end{aligned} \tag{8}$$

Theorem 9, Lecture 4 extends to the current setting, so by defining

$$N^j := \sum_{n \in \mathbb{N}} \mathbb{1}_{Z_n^j = j} \quad \text{(total visits at state } j\text{)},$$

we obtain for $\lambda_j := \mathbb{P}(T^j < \infty)$ that

$$\mathbb{P}(N^j = k) = \begin{cases} (1-\lambda_j)\lambda_j^{k-1} & \text{if } \lambda_j < 1 \\ \mathbb{1}_{k=\infty} & \text{if } \lambda_j = 1 \end{cases}$$

Consequently,

$$0 < \mathbb{P}(Z_n^j = j \quad \text{for infinitely many } n) = \mathbb{P}(N^j = \infty) = \begin{cases} 0 & \text{if } \lambda_j < 1 \\ 1 & \text{if } \lambda_j = 1. \end{cases}$$

Conclusion: λ_j must equal 1 and j is a recurrent state.

It remains to verify that $N^k = \infty$ a.s. for all $k \in \mathbb{S} \setminus \{j\}$. Observe first that

$$\mathbb{P}(N^k = \infty) = 1 \iff \mathbb{P}(N^k = \infty) > 0$$
$$\iff \mathbb{E}\left[N^k \right] = \infty \iff \sum_{n \in \mathbb{N}} p_{kk}^n = \infty$$

where the last $\iff$ follows from

$$\mathbb{E}\left[N^k \right] = \sum_{n \in \mathbb{N}} \mathbb{E}\left[\mathbb{1}_{Z_n^k = k} \right] = \sum_{n \in \mathbb{N}} \mathbb{P}(\mathbb{1}_{Z_n^k = k}) = \sum_{n \in \mathbb{N}} p_{kk}^n.$$

Since $\mathbb{P}(N^j = \infty) = 1$, we know that $\sum_{n \in \mathbb{N}} p_{jj}^n = \infty$. And by the irreducibility of p, there exist $m_1, m_2 \geq 1$ such that $p_{kj}^{m_1} p_{jk}^{m_2} > 0$. So for any $n \geq m_1 + m_2$,

$$p_{kk}^n \geq p_{kj}^{m_1} p_{jj}^{n-(m_1+m_2)} p_{jk}^{m_2}$$

and

$$\sum_{n \in \mathbb{N}} p_{kk}^n \geq p_{kj}^{m_1} p_{jk}^{m_2} \sum_{n \in \mathbb{N}} p_{jj}^n = \infty.$$

Q.E.D.

Construction of invariant measures

For an irreducible transition function p associated to $\mathbb{S} = \{1, 2, \ldots, d\}$, we fix a state $k \in \mathbb{S}$, the chain $\{Z_n^k\} \sim Markov(\mathbb{1}_{\{k\}}, p)$ and introduce

$$\gamma_j^k := \mathbb{E}\left[\sum_{n=0}^{T^k - 1} \mathbb{1}_{Z_n^k = j}\right] \quad \text{for} \quad j \in \mathbb{S}.$$

(the expected number of visits spent at state j in between vists to k).

Theorem 22 (Theorem 1.7.5, Norris, Markov Chains)

For every $k \in \mathbb{S}$,

$$\gamma^k = \gamma^k p,$$

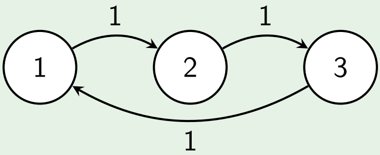
which makes

$$\pi := \frac{\gamma^k}{\sum_{j \in \mathbb{S}} \gamma_j^k}$$

is an invariant distribution.

Example 23

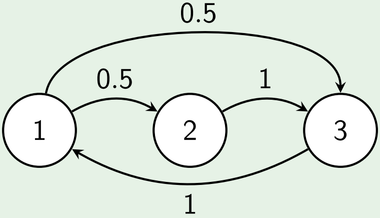
$$p = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



Irreducible but periodic chain. $p_{ii}^n > 0$ only for $n = 3, 6, 9, \ldots$. So Lemma 19 does not apply.
But $\gamma^1 = \gamma^2 = \gamma^3 = [1,\ 1,\ 1]$, giving rise to $\pi = \gamma^1/3$.

Example 24

$$p = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



Irreducible chain with aperiodic state 3. So Lemma 19 does apply.
But theorem 22 also:

$$\gamma^1 = [1, 0.5, 1], \quad \gamma^2 = [2, 1, 2], \quad \gamma^3 = [1, 0.5, 1]$$

Simulation of a time-homogeneous Markov chain

For $\{Z_n\} \sim Markov(\pi^0, p)$ on $\mathbb{S} = \{1, 2, \ldots, d\}$ the main challenges for simulation are to draw the inital state and the transitions:

1. Draw $Z_0 \sim \pi^0$
2. $\ldots$
3. given $Z_n = i$, draw $Z_{n+1} \sim [p_{i1}, p_{i,2}, \ldots, p_{id}]$

Same challenge for every step: draw a sample/new state from a distribution $f = [f_1, \ldots, f_d]$.

Sampling method:

1. construct a vector

$$\bar{f} = \begin{pmatrix} f_1 \\ f_1 + f_2 \\ \vdots \\ \sum_{j=1}^{d-1} f_j \\ 1 \end{pmatrix}$$

2. Draw a uniformly distributed rv $U \sim U[0,1]$ and determine new state by:

$$j(U) := \min\{k \in \{1, 2, \ldots, d\} \mid \bar{f}_k > U\}.$$

Exercise: verify that $\mathbb{P}(j(U) = \ell) = f_\ell$.

Next time

Filtering of discrete time and space Markov Chains