Stochastic Simulations

Autumn Semester 2024

Prof. Fabio Nobile *Assistant: Matteo Raviola*

Lab 09 – 14 November 2024

Markov Chains

We recall some concepts on the theory of Markov chains on a discrete state space $\mathcal{X}$.

Irreducibility

Let P be be a transition matrix on $\mathcal{X}$. We say that a state $x_i \in \mathcal{X}$ communicates with a state $x_j \in \mathcal{X}$ if $\mathbb{P}(X_n = x_j, \text{ for some } n \mid X_0 = x_i) > 0$; equivalently, if $\exists n \geq 0$ such that $P_{i,j}^{(n)} > 0$. A Markov chain is *irreducible* if every state x_j communicates with every other state x_i, i.e.,

$$\forall i, j, \quad \exists n \geq 0 \text{ such that } P_{i,j}^{(n)} > 0.$$

Recurrence

A state $x_i \in \mathcal{X}$ is *recurrent* if $\mathbb{P}(X_n = x_i \text{ infinitely often }) = 1$, that is, x_i is visited infinitely-often with probability 1. A Markov chain $\{X_n\}$ is recurrent if everry state is recurrent. It is known that every irreducible recurrent Markov chain $\{X_n\}$ on a discrete state space has a (not necessarily finite) invariant distribution π that is unique up to a multiplicative constant. However, if the state space is *finite*, every irreducible Markov chain $\{X_n\}$ is recurrent and has a unique invariant probability distribution.

Aperiodicity

The *period* of a state x_i is the largest integer d satisfying the following property: $P_{i,i}^{(n)} = 0$, whenever n is not divisible by d. The period of x_i is given by $d(i)$. We say that if $d(i) > 1$, then the state x_i is *periodic*. We say that the state x_i is *aperiodic* otherwise. If a Markov chain $\{X_n\}$ is irreducible and has an aperiodic state, then all states are aperiodic, in which case we say that $\{X_n\}$ is aperiodic. In particular, an irreducible Markov chain $\{X_n\}$ is aperiodic if there exists a state $x_j \in \mathcal{X}$ such that $P_{jj} > 0$. It is known that an irreducible Markov chain $\{X_n\}$ on a *finite* state space $\mathcal{X}$ converges to π, i.e., $\pi_j = \lim_{n\to\infty} \mathbb{P}(X_n = x_j)$, $x_j \in \mathcal{X}$, if and only if $\{X_n\}$ is aperiodic.

Exercise 1

Consider the random walk $\{X_n \in \mathbb{Z}, n \in \mathbb{N}_0\}$ with $X_0 \sim \lambda$ on the lattice $\mathcal{X} := \{i\colon i \in \mathbb{Z}, |i| \le 2N^2\}$, whose transition probabilities are given by

$$\mathbb{P}\big(X_{n+1} = i \pm 1 \big| X_n = i\big) = \alpha\left(1 \mp \frac{i}{2N^2}\right) , \quad |i| \le 2N^2 ,$$
$$\mathbb{P}\big(X_{n+1} = i \big| X_n = i\big) = 1 - 2\alpha ,$$

for some $\alpha \in]0, \frac{1}{2}]$ and $N \in \mathbb{N}$.

1. Implement an algorithm that simulates the Markov Chain $\{X_n \in \mathbb{Z}, n \in \mathbb{N}_0\}$. Use your implementation to address the following points for different values of $N \ge 1$:

   (a) Assess numerically that the Markov chain converges to an invariant distribution by simulating multiple (independent) chains, each starting in 0 (i.e. $\lambda = \delta_0$). That is, monitor the following quantities (rather, suitable Monte Carlo approximations) as functions of the Markov chain length n.

      i. $\mathbb{E}_\lambda\big({X_n}^p\big)^{1/p}$ for $p \in \{1, 2, 4\}$,

      ii. $M_{X_n}(t) := \mathbb{E}_\lambda\big(e^{tX_n}\big)$ for $t \in [-1, 1]$.

      Speculate on the invariant distribution.

   (b) For $N = 10$, compute the eigenvalues and eigenvectors of the transition matrix P. Use the obtained results to deduce the invariant distribution π. **Hint:** Use `np.linalg.eig(P)`.

   (c) Assess the validity of the ergodic theorem. That is, verify that

   $$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} f(X_n) = \mathbb{E}_\pi(f) , \quad \mathbb{P}_\lambda\text{–a.s.},$$

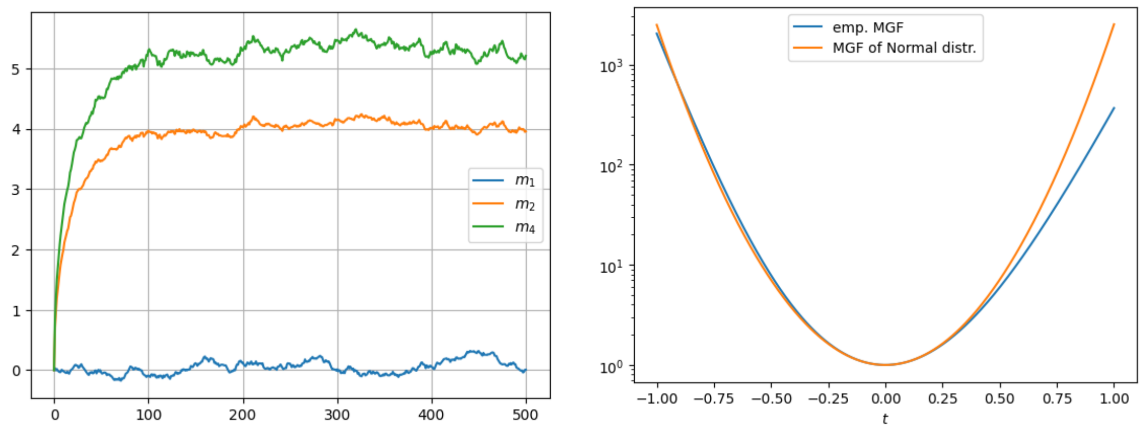
   for any $f : \mathcal{X} \to \mathbb{R}$, with $\sum_n |f(X_n)|\pi_n < \infty$. Specifically, investigate this identity for the moments used in Point 1(a)i and monitor the rate of convergence as a function of n.

2. Consider the rescaled Markov chain $Y_n := \frac{1}{N}X_n$ with state space $\mathcal{Y} := \{x_i \equiv \frac{i}{N}\colon i \in \mathbb{Z}, |i| \le 2N^2\}$. Show by means of numerical simulations that the invariant distribution $\nu \equiv \nu_N$ of $\{Y_n \in \mathbb{Z}, n \in \mathbb{N}_0\}$ is an accurate approximation to the standard normal measure. Moreover, illustrate that the approximation quality improves as N increases.

Solution

1. A possible implementation is shown below:

```
def latticeRW(n=100,X0=0,N=2,aa=0.5):
    X=np.zeros(n+1)
    X[0]=X0
    t=np.arange(0,n+1)
    for i in range(n):
        xcurr=X[i]
```



(a) Moments $m_p \equiv m_p(n) = \sqrt[p]{\mathbb{E}({X_n}^p)}$ as a function of n.

(b) Empirical moment generating function compared to the one of a Normal distribution.

Figure 1: Convergence plots

```
        pup=(1-xcurr/2/N**2)*aa*(np.abs(xcurr)<=2*N**2)
        plow=(1+xcurr/2/N**2)*aa*(np.abs(xcurr)<=2*N**2)
        r=np.random.random(1)
        if (r<=plow):                        #decreses state
            X[i+1]=X[i]-1
        elif( (plow<r)*(r<=plow+pup)): # increses state
            X[i+1]=X[i]+1
        else:
            X[i+1]=X[i]
    return X,t
```

(a) The numerical results shown here are obtained for $\alpha = 1/4$, $N = 4$, and the expectations are approximated by Monte Carlo generating $R = 10^3$ independent chains. When inspecting the evolution of the resulting moment approximations as a function of n, Fig. 1a, we notice an exponential convergence to a stationary state. Moreover, the transition probabilities of the chain leads us to suspect that the invariant distribution could be close to a Gaussian. This suspicion is further strengthened when comparing the approximated moment generating function at $n = 500$ with the moment generating function of a normal distribution, as is shown in Figure 1b.

(b) If the Markov chain converges, then we expect exponential convergence $e(n) = |\lambda_2|^n$, where λ_2 is the eigenvalue with largest absolute value smaller than one. This is also visible in the numerical results of the previous point. Estimates of this value for $\alpha = 1/4$ are shown below:

```
N = 1: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.75^n
N = 2: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.9375^n
N = 4: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.984375^n
N = 8: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.996094^n
N = 10: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.9975^n
N = 20: expected rate of convergence (in TV norm): C*|lambda_2|^n = C*0.999375^n
```

(c) As a consequence of the CLT for Markov chains, we observe the canonical $1/\sqrt{n}$ convergence rate. We observe the trend in Fig. 2

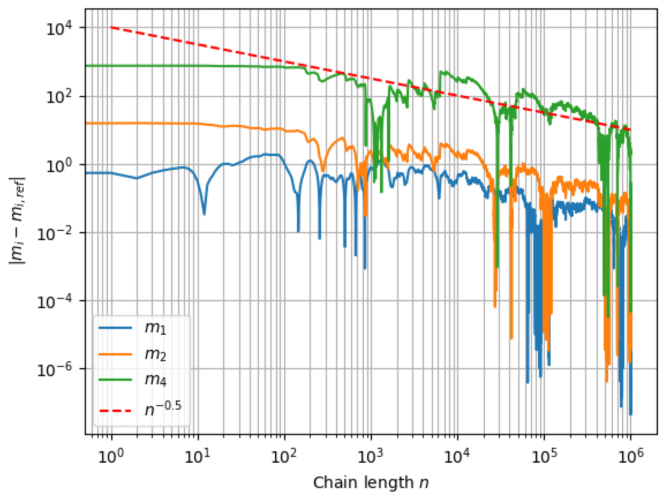


Figure 2: Convergence of ergodic estimator to reference value

Python code for lattice simulation:

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Notice that this takes quite a bit of time to run.
"""
import numpy as np
import matplotlib.pyplot as plt
def latticeRW(n=100,X0=0,N=4,aa=0.25):
        X=np.zeros(n+1)
        X[0]=X0
        t=np.arange(0,n+1)
        for i in range(n):
                xcurr=X[i]
                pup=(1-xcurr/2/N**2)*aa*(np.abs(xcurr)<=2*N**2)
                plow=(1+xcurr/2/N**2)*aa*(np.abs(xcurr)<=2*N**2)
                r=np.random.random(1)
                if (r<=plow):                   #decreses state
                        X[i+1]=X[i]-1
                elif( (plow<r)*(r<=plow+pup)): # increses state
                        X[i+1]=X[i]+1
                else:
                        X[i+1]=X[i]
        return X,t
```

```python
#
#   Defines some parameters
#
N=4
x0=0
aa=1/4
N_runs=int(1E3) #this might be quite a large number
n=500
x1=np.zeros((N_runs,n+1))
x2=np.zeros((N_runs,n+1))
x4=np.zeros((N_runs,n+1))
M=np.zeros((n+1,n+1))
tau=np.linspace(-1,1,n+1)

# runs the chain many times and obtains the quantities needed
for i in range(N_runs):
    x1[i,:]=latticeRW(n,x0,N,aa)[0]
    x2[i,:]=x1[i,:]**2.0
    x4[i,:]=x1[i,:]**4.0

#computes moments
e1=np.mean(x1,0)
e2=np.mean(x2,0)**0.5
e4=np.mean(x4,0)**0.25
Mx=np.mean(M,0)

plt.plot(e1);
plt.plot(e2);
plt.plot(e4);
plt.grid(True,which='both')
plt.legend([r'$m_1$',r'$m_2$',r'$m_4$'])
plt.savefig('../figures/fig1_a.png',format='png',bbox_inches='tight')
plt.show()

# computes the MGF
tauvec=np.linspace(-1,1,n+1)
ntau=n+1
MGF=np.zeros((n+1,ntau))
for r in range(N_runs):
    for i in range(ntau):
        MGF[:,i]=MGF[:,i]+np.exp(tauvec[i]*x1[r,:])
MGF=MGF/N_runs
mu=e1[-1]
si2=e2[-1]**2.0-mu**2
plt.semilogy(tauvec,MGF[-1,:])
```

```
plt.semilogy(tauvec,np.exp(tauvec*mu+si2/2*tauvec**2.));
plt.legend(['emp. MGF','MGF of Normal distr.']);
plt.xlabel(r'$t$');

plt.savefig('../figures/fig1_b.png',format='png',bbox_inches='tight')
plt.grid(True,which='both')

plt.show()
```

Python code for matrix construction:

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-


#Constructs matrix and outputs 2nd largest eig

import numpy as np
import matplotlib.pyplot as plt


def build_matrix(a,N):
    Nt=4*N**2+1
    P=np.zeros((Nt,Nt))

    for i in range(Nt):
        xcurr=i-2*N**2
        P[i,i]=1-2*a
        if (i<Nt-1):
            P[i,i+1]=(1-xcurr/2/N**2)*a*(np.abs(xcurr)<=2*N**2)
        if (i>0):
            P[i,i-1]=(1+xcurr/2/N**2)*a*(np.abs(xcurr)<=2*N**2)
    # obtains second largest eig
    e,v=np.linalg.eig(P)
    e.sort()
    return P,e[-2]

a=1/4
N=[1,2,4,8,10,20]
for n in N:
    _,e=build_matrix(a,n)
    e=np.real(e) # to remove the 0j imaginary part
    print('N = '+str(n) + ': expected rate of convergence (in TV norm): C*|lambda_2|^n =
```

Exercise 2

A random walk on the integers $I = \{0, 1, 2\dots\}$ can be constructed in the following way. For $0 < p < 1/2$, let $Y_0, Y_1, \dots$ be i.i.d random variables with $P(Y_i = 1) = p$ and $P(Y_i = -1) = 1 - p$. Define two random walks as (1) $X_n = \max\{X_{n-1} + Y_n, 0\}$ and (2) $Z_n = |Z_{n-1} + Y_n|$.

1. Show that (X_n) and (Z_n) are Markov chains.
2. Show that an invariant probability measure of the chains (X_n) and (Z_n) is given by

$$\hat{\pi} = \left[1, \left(\frac{p}{1-p}\right), \dots, \left(\frac{p}{1-p}\right)^k, \dots\right] a_0, \quad k \geq 0$$
$$\bar{\pi} = \left[1, \frac{1}{1-p}, \frac{p}{(1-p)^2}, \dots, \frac{p^{k-1}}{(1-p)^k}, \dots\right] b_0, \quad k \geq 1,$$

   respectively. Find p, a_0, b_0 such that the expressions above are probability distributions.
3. Let $p = 1/8$. Assess numerically the convergence of both Markov chains to their invariant distribution by simulating multiple (independent) chains of length $n = 100$, each staring in 0 (i.e. $\lambda = \delta_0$). That is, plot the empirical distribution of X_n, Z_n vs $\hat{\pi}$ and $\bar{\pi}$, respectively. Repeat your experiments for $m = n + 1$. Explain your results.
4. Discuss the periodicity of both chains.

Solution

1. This follows from the definition of Markov chain, since X_{n+1} only depends on X_n. Thus,

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

2. Recall that π is an invariant distribution of a Markov chain with transition matrix P if

$$\pi P = \pi.$$

   Thus, we will verify this for our case. Consider the two Markov transition matrices P and Q given by

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & \dots \\ 1-p & 0 & p & 0 & 0 & \ddots \\ 0 & 1-p & 0 & p & \ddots & \ddots \\ 0 & 0 & 1-p & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 1-p & 0 & p & 0 & 0 & \ddots \\ 0 & 1-p & 0 & p & \ddots & \ddots \\ 0 & 0 & 1-p & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1}$$

   where P corresponds to the transition matrix of the chain generated by $X_{n+1} = \max\{0, X_n + Y\}$ and Q corresponds to that of $Z_{n+1} = |Z_n + Y|$. We will show that to P and Q there correspond invariant distributions

$$\hat{\pi} = \left[1, \left(\frac{p}{1-p}\right), \ldots, \left(\frac{p}{1-p}\right)^k, \ldots\right] a_0, \quad k \geq 0,$$
$$\bar{\pi} = \left[1, \frac{1}{1-p}, \frac{p}{(1-p)^2}, \ldots, \frac{p^{k-1}}{(1-p)^k}, \ldots, \right] b_0, \quad k \geq 1.$$

respectively. We begin with P. For simplicity, denote the $i-$th column of P by P_i. Notice that for P_1, we have $\hat{\pi}P_1 = [(1-p) + (1-p)p/(1-p)]a_0 = a_0 = \hat{\pi}_1$. Moreover, for the $k-$th column we have

$$\begin{aligned} \hat{\pi}P_k &= \hat{\pi}_{k-1}P_{k-1,k} + \hat{\pi}_{k+1}P_{k+1,k} = p\left(\frac{p}{1-p}\right)^{k-1} a_0 + (1-p)\left(\frac{p}{1-p}\right)^{k+1} a_0 \\ &= \left(p + \frac{p^2}{1-p}\right)\left(\frac{p}{1-p}\right)^{k-1} a_0 = \left(\frac{p-p^2+p^2}{1-p}\right)\left(\frac{p}{1-p}\right)^{k-1} a_0 = \left(\frac{p}{1-p}\right)^k a_0 = \hat{\pi}_k, \end{aligned} \tag{2}$$

and as such, the chain with transition matrix P is invariant with respect to $\hat{\pi}$.

As for Q, we have that for Q_1, $\bar{\pi}Q_1 = (1-p)/(1-p)b_0 = b_0$, for Q_2, we have $\bar{\pi}Q_2 = b_0 + p/(1-p)b_0 = \bar{\pi}_2$. Moreover, for Q_k, $k > 2$, we have

$$\begin{aligned} \bar{\pi}Q_k &= \bar{\pi}_{k-1}p + \bar{\pi}_{k+1}(1-p) \\ &= p\frac{p^{k-2}}{(1-p)^{k-1}}b_0 + (1-p)\frac{p^k}{(1-p)^{k+1}}b_0 \\ &= \frac{p^{k-1}}{(1-p)^{k-1}}b_0 + \frac{p^k}{(1-p)^k}b_0 \\ &= \frac{p^{k-1}}{(1-p)^k}b_0 = \bar{\pi}_k \end{aligned}$$

which implies that $\bar{\pi}$ is an invariant distribution for the Markov chain with transition matrix Q.

Notice that for both chains to be probability distributions, the sum of the components of $\hat{\pi}$ and $\bar{\pi}$ must add to 1. On the case of $\bar{\pi}$, a_0 and p must be chosen such that

$$a_0 \sum_{k=0}^{\infty} \left(\frac{p}{1-p}\right)^k = 1.$$

Notice that the previous sum is finite if $p/(1-p) < 1$, which occurs for any $p < 1/2$. In turn, it can be shown that (by computing the value of the previous geometric series) in this case, the normalization constant is

$$a_0 = \frac{1-2p}{1-p}.$$

A similar approach can be applied to Q to obtain

$$b_0 = \frac{1-2p}{2-2p}. \tag{3}$$

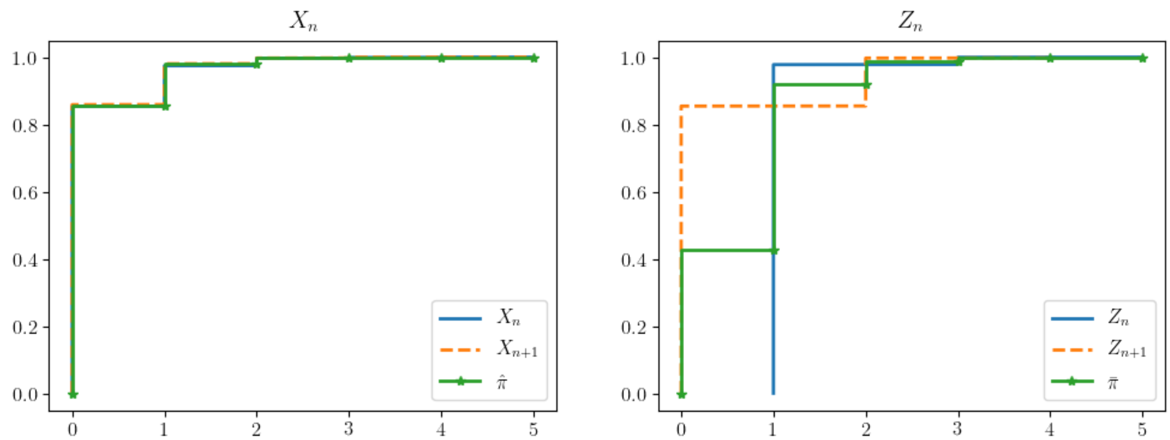


Figure 3: Empirical CDF of X_n (left) and Z_n (right) estimated from 10000 independent runs. As we can see, the process X_n converges to the true distribution, whilst Z_n does not. this is due to the periodicity of Z_n.

3/4. Python code is attached. We repeat our experiments $N = 10000$ times starting at $X_0 = Z_0 = 0$. Figure 3 shows the empirical CDF of X_n (left) and Z_n (right) for $n = 100$ and $n = 101$. As we can see, the chain with transition matrix P converges to its invariant distribution, while the one with transition matrix Q does not. This is due to the chain with transition matrix Q not being aperiodic; notice that if the chain is at state 0, the chain will move to state 1 with probability 1 and only returns to 0 after an even number of steps.

Python code

```python
import numpy as np
import matplotlib.pyplot as plt
#-------------------------------------
from matplotlib import rc
rc('font',**{'family':'serif','serif':['Computer Modern Roman'],
     'size' : '12'})
rc('text', usetex=True)
rc('lines', linewidth=2)
plt.rcParams['axes.facecolor']='w'
import matplotlib
matplotlib.rcParams['text.latex.preamble'] = [
    r'\usepackage{amsmath}',
    r'\usepackage{amssymb}']
#-----------------------------------------------
def chain1(n):
    xn=np.zeros(n)
    for i in range(n-1):
        u=np.random.random(1)
        if u <p:
            a=1
        else:
            a=-1
        xn[i+1]= np.max((xn[i] + a,0))
    return xn

def chain2(n):
    zn=np.zeros(n)
    for i in range(n-1):
        u=np.random.random(1)
        if u <p:
            a=1
        else:
            a=-1
        zn[i+1]= np.abs(zn[i] + a)
    return zn

#computes empirical cdf
def ecdf(data):
    """ Compute ECDF """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)

#computes pi hat
```

```python
def true_cdf_1(K,p):
    a0=(1-2*p)/(1-p) #this is the normalization constant
    pi_hat=(p/(1-p))**np.arange(K)*a0
    return pi_hat

#computes pi bar
def true_cdf_2(K,p):
    r=p/(1-p)
    b0=1/(1+(1/p)*(r/(1-r))) #this is the normalization constant
    print(b0)
    print(1/(1+1/(1-2*p)))
    pi_hat=(1/p)*(p/(1-p))**np.arange(1,K)*b0
    return np.concatenate([[b0],pi_hat])

n=100
p=1/8

N=10000
m1=np.zeros(N)
m2=np.zeros(N)
m3=np.zeros(N)
m4=np.zeros(N)
for i in range(N):
    m1[i]=chain1(n)[-1]
    m2[i]=chain1(n+1)[-1]
    m3[i]=chain2(n)[-1]
    m4[i]=chain2(n+1)[-1]

# computes empirical distribution
x1,e1=ecdf(m1)
x2,e2=ecdf(m2)
x3,e3=ecdf(m3)
x4,e4=ecdf(m4)

# computes the CDF for Xn
K1=np.max(x1) #highest value obtained
pi_hat=true_cdf_1(K1,p) #obtaines values of pi hat
cdf_pi_hat=np.cumsum(pi_hat) #computes cdf
cdf_pi_hat=np.concatenate([ [0],cdf_pi_hat]) # adds 0 for plotting

# Computes the CDF for Zn
K2=np.max([np.max(x3),np.max(x4)]) #highest value obtained
pi_bar=true_cdf_2(K2,p) #obtaines values of pi bar
cdf_pi_bar=np.cumsum(pi_bar) #computes cdf
cdf_pi_bar=np.concatenate([ [0],cdf_pi_bar]) # adds 0 for plotting

#fixes parameter for plotting
```

```python
fig, axes = plt.subplots(nrows = 1, ncols = 2, figsize = (12,4))

# plots cdf of Xn
axes[0].plot(x1,e1);
axes[0].plot(x2,e2,'--');
axes[0].step(np.arange(K1+1),cdf_pi_hat,'-*')
axes[0].set_title(r'$X_n$')
axes[0].legend([r'$X_{n}$',r'$X_{n+1}$',r'$\hat{\pi}$'])

# plots cdf of Zn
axes[1].plot(x3,e3);
axes[1].plot(x4,e4,'--');
axes[1].step(np.arange(K2+1),cdf_pi_bar,'-*')
axes[1].set_title(r'$Z_n$')
axes[1].legend([r'$Z_{n}$',r'$Z_{n+1}$',r'$\bar{\pi}$'])
plt.savefig('../figures/convergence_distribution.png',bbox_inches='tight')
```

Exercise 3

Given the transition matrix

$$P = \begin{pmatrix} 0.0 & 0.4 & 0.6 & 0.0 & 0.0 \\ 0.65 & 0.0 & 0.35 & 0.0 & 0.0 \\ 0.32 & 0.68 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.12 & 0.88 \\ 0.0 & 0.0 & 0.0 & 0.56 & 0.44 \end{pmatrix}, \tag{4}$$

and examine whether the corresponding chain is irreducible and aperiodic.

Solution

Notice that from the block structure of the matrix P, the chain is not irreducible since states 4 and 5 do not communicate with states $1, 2, 3$. Aperiodicity can be checked by definition.