STAT 753: Stochastic Models and Simulations
Jaleesa Houle
University of Nevada, Reno - Spring 2024

Homework 7

Theory

1. For the simple random walk starting from 3, with probabilities 0.6 and 0.4 going up or down, find the probability:

   a. It will hit 10 before -1.

   Solution

   For a random walk with probability p of moving up $(+1)$ and probability q of moving down (-1), the general probability of hitting value n+1 (i.e, q_{n+1}) before hitting value n-1 (i.e, q_{n-1}) is

$$Q_x = pQ_{x+1} + qQ_{x-1} \tag{1}$$

   For cases where $p \neq q$, a solution to this equation can be found by guessing $Q_x = C^x$. Equation 1 then becomes

$$C^x = pC^{x+1} + qC^{x-1} \tag{2}$$

   where after solving, the solution is of the form

$$Q_x = k\left(\frac{q}{p}\right)^x + d. \tag{3}$$

   .

   To find k and d in Equation 3, we can note two "boundary conditions" of sorts: $Q_a = 1$ is the probability that (starting from a) you hit a before b, and $Q_b = 0$ is the probability that (starting from a) you hit b before a.

   For this problem, we let $a = 10$ and $b = -1$. This yields the system of equations and corresponding solutions for k and d:

$$\begin{aligned} Q_a &= 1 = Q_{10} \\ &= k\left(\frac{.4}{.6}\right)^{10} + d \\[2ex] Q_b &= 0 = Q_{-1} \\ &= k\left(\frac{.4}{.6}\right)^{-1} + d \end{aligned} \tag{4}$$

$$=> = \begin{Bmatrix} k = -0.6745 \\ d = 1.01169 \end{Bmatrix}.$$

Thus, $Q_x = -0.6745\left(\frac{2}{3}\right)^x + 1.01169$. The probability of hitting 10 before -1 if starting from 3 is then $Q_3 = -0.6745\left(\frac{2}{3}\right)^3 + 1.01169 \approx 0.81183$.

b. It reaches 5 in 10 steps, and always stays strictly above zero.

Solution

The number of paths that reach 5 in 10 steps and stay strictly above 0 can be found by using the reflection principle. First, we can determine the number of paths from (0, 3) to (10, 5) using the system of equations

$$\begin{aligned} a + b &= 10 \\ a - b &= 2 \\ &=> \begin{Bmatrix} a = 6 \\ b = 4 \end{Bmatrix}. \end{aligned} \tag{5}$$

Therefore, the number of ways to go from (0, 3) to (10, 5) is $\binom{10}{6} = 210$. Now, reflecting a path over the line $y = 0$, the reflected line will end at (10, -5). The number of paths from (0, 3) to (10, -5) can be determined the same way as above:

$$\begin{aligned} a + b &= 10 \\ a - b &= -8 \\ &=> \begin{Bmatrix} a = 1 \\ b = 9 \end{Bmatrix}. \end{aligned} \tag{6}$$

The number of ways to go from (0, 3) to (10, -5) is then $\binom{10}{1} = 10$, and the total number of paths that end at 5 and stay strictly above 0 is $210 - 10 = 200$. The corresponding probability is $200p^aq^b = 200(0.6)^6(0.4)^4 = 0.2388$.

2. For the AR(1) starting from 2: $X_n = 0.9X_{n-1}+2+2.5Z_n$ where Z_n are IID standard normal, find the stationary distribution.

Solution

The general AR(1) model is defined as $Y_n = aY_{n-1} + c + Z_n$. For The given AR(1) model, there is an additional constant in front of Z_n, which we will denote as d. The equation then becomes $Y_n = aY_{n-1} + c + dZ_n$. By writing out the equations for Y_1, Y_2, and Y_3, we can then find the generalized equation for Y_n.

$$
\begin{aligned}
Y_1 &= aY_0 + c + dZ_1 \\
Y_2 &= a(aY_0 + c + dZ_1) + c + dZ_2 \\
&= a^2 + ac + c + d(aZ_1 + Z_2) \\
Y_3 &= a(a^2 + ac + c + adZ_1 + dZ_2) + c + dZ_3 \\
&= a^3 + a^2c + ac + c + d(a^2Z_1 + aZ_2 + Z_3) \\
&... \\
Y_n &= a^nY_0 + (a^{n-1} + ... + a + 1)c + d(a^{n-1}Z_1 + ... + aZ_{n-1} + Z_n).
\end{aligned} \tag{7}
$$

The corresponding expected value for Y_n (with $Z_n \sim N(0, 1)$ IID) can then be computed as

$$
\begin{aligned}
E[Y_n] &= E[a^nY_0 + (a^{n-1} + ... + a + 1)c + d(a^{n-1}Z_1 + ... + aZ_{n-1} + Z_n)] \\
&= E[a^nY_0 + (a^{n-1} + ... + a + 1)c] \\
&= a^nY_0 + \frac{1 - a^n}{1 - a}c \\
&... \\
\lim_{x\to\infty} E[Y_n] &= \frac{c}{1 - a}.
\end{aligned} \tag{8}
$$

Similarly, the variance can be shown to be

$$
\begin{aligned}
var[Y_n] &= var[a^nY_0 + (a^{n-1} + ... + a + 1)c + d(a^{n-1}Z_1 + ... + aZ_{n-1} + Z_n)] \\
&= var[d(a^{n-1}Z_1 + ... + aZ_{n-1} + Z_n)] \\
&= d^2(a^{2n-2} + ... + a^2 + 1)\sigma^2 \\
&= d^2\frac{1 - a^{2n}}{1 - a}\sigma^2 \\
&... \\
\lim_{x\to\infty} var[Y_n] &= \frac{d^2\sigma^2}{1 - a^2}.
\end{aligned} \tag{9}
$$

Thus, the stationary distribution for an AR(1) with $X_n = 0.9X_{n-1}+2+2.5Z_n$ where Z_n are IID standard normal will have a mean of $c/(1 - a) = 2/(1 - 0.9) = 20$ and a variance of $d^2\sigma^2/(1 - a^2) = 2.5^2/(1 - 0.9^2) \approx 32.895$.

Coding

1. Simulate simple random walk from Theory 1 and find empirical probabilities of A and B.

Solution

Part A

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

N = 1000
p=0.6
q=0.4
#Sn = np.empty((N, 0)).tolist()
Sn =[]
Z= [3] #S_0 - start position
np.random.seed(1)
for i in range(N):
    A = [3]
    while (A[-1] != 10) and (A[-1] != -1):
        A.append(A[-1] + np.random.choice(a=[-1,1],p=[q,p]))

    Sn.append(A)

count = 0
for i in range(N):
    if Sn[i][-1] == 10:
        count+=1

print('Empirical probability of hitting 10 before -1:', count_10/N)
>> Empirical probability of hitting 10 before -1: 0.812
```

It can be seen that the empirical probability of hitting 10 before -1 after running 1000 simulations was 0.812. The theoretical probability found in Part A was approximately 0.81183, which is in strong agreement with the empirical result.

Part B.

```
T = 10
N = 10000
p=0.6
q=0.4
Sn =[]
np.random.seed(1)
for i in range(N):
```

```
    Z = np.random.choice(a=[-1,1],p=[q,p], size=T+1)
    Z[0]= 3 #S_0 - start position
    Sn.append(np.cumsum(Z))

count = 0
for i in range(N):
    if (Sn[i][-1] == 5) and (np.min(Sn[i])>0):
        count+=1

print('Probability of path ending at 5 and staying strictly above
    0:',count/N)
>> Probability of path ending at 5 and staying strictly above 0: 0.2377
```

The theoretical probability of the random walk starting at 3 and ending at 5 while staying strictly above 0 was found to be 0.2388. It is clear that the empirical probability, 0.2377, is in strong agreement with the theoretical solution.

2. Simulate 2000 steps of AR(1) from Theory 2 and find the empirical stationary distribution (plot the histogram and find the empirical mean and variance) after a burn-in period of 100 steps. Plot the autocorrelation function.

Solution

```
N = 2000
n = 100
X_n = [2]
a = 0.9
c = 2
b = 2.5
np.random.seed(1)
Z_n = np.random.normal(0,1,N)
f = lambda X, Z: 0.9*X + c + 2.5*Z
for i in range(N):
    X_n.append(f(X_n[i], Z_n[i]))

print('Mean of Xn after 100 step burn-in period:', np.mean(X_n[100:]))
print('Variance of Xn after 100 step burn-in period:',np.var(X_n[100:],
    ddof=1))

>> Mean of Xn after 100 step burn-in period: 20.800046374254684
>> Variance of Xn after 100 step burn-in period: 30.40941231437156

# plotting histogram of AR(1) after burn-in period
plt.title('Distribution of Xn (after 100 step burn-in period)')
sns.histplot(X_n[100:], color='blue', alpha=0.6, bins=60,
    stat='probability')

# plotting the AR model
```

```
fig = plt.subplots(figsize=(8,6))
t = np.arange(N+1)
plt.plot(t[:100], X_n[:100], color='grey', alpha=0.9, label='"burn-in"
↪   period')
plt.plot(t[100:], X_n[100:], color='blue', label='AR(1) after burn-in
↪   period', alpha=0.6)
plt.axhline(np.mean(X_n), color='k', label='sample mean')
plt.fill_between(x=t, y1=np.mean(X_n)- np.std(X_n), y2= np.mean(X_n)+
↪   np.std(X_n), alpha=0.1, color='blue', label='+/- 1 standard deviation')
plt.ylabel('Value')
plt.xlabel('Time step')
plt.title('AR(1) Model')
plt.legend(loc='upper right')
plt.xlim(-5,2010)
plt.ylim(0,50)

sm.graphics.tsa.plot_acf(X_n, lags=100)
plt.ylabel('Correlation')
plt.xlabel('Time lags')
plt.title('Autocorrelation function for AR(1)')
plt.show()
```

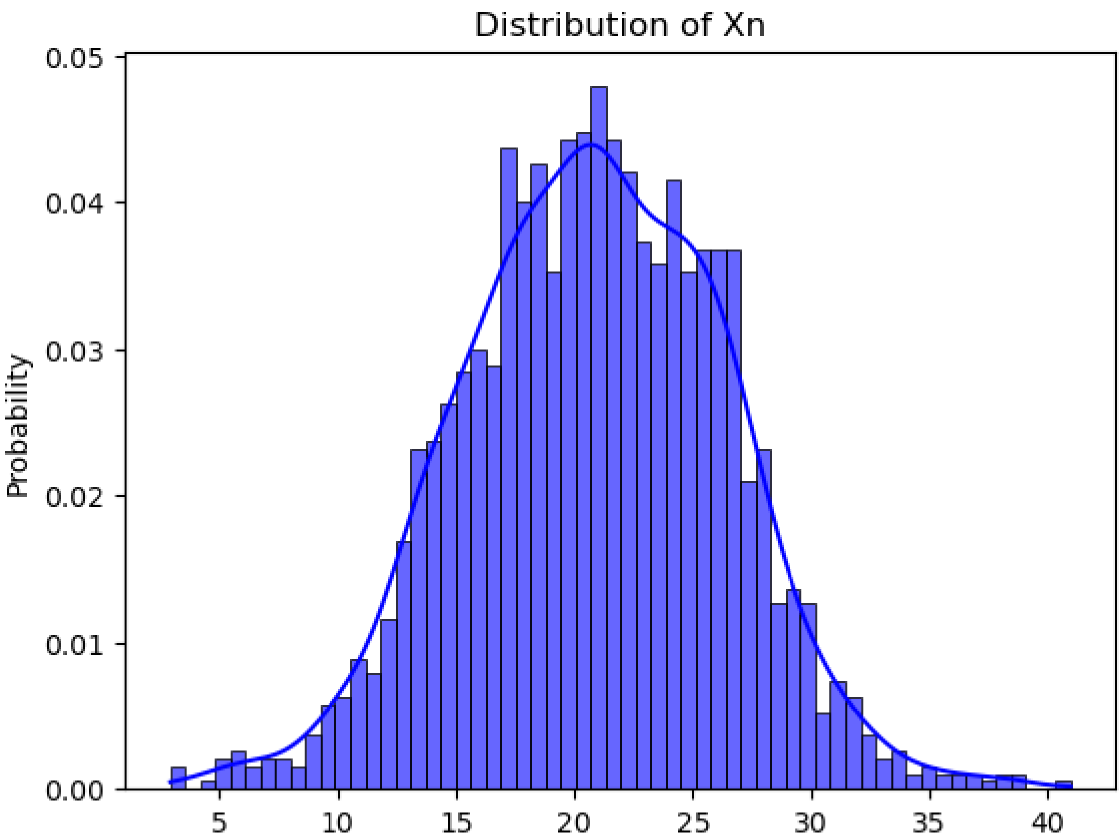

Figure 1: Histogram of AR(1) model after "burn-in" period. The distribution has a mean of 20.8 and variance of 30.41.

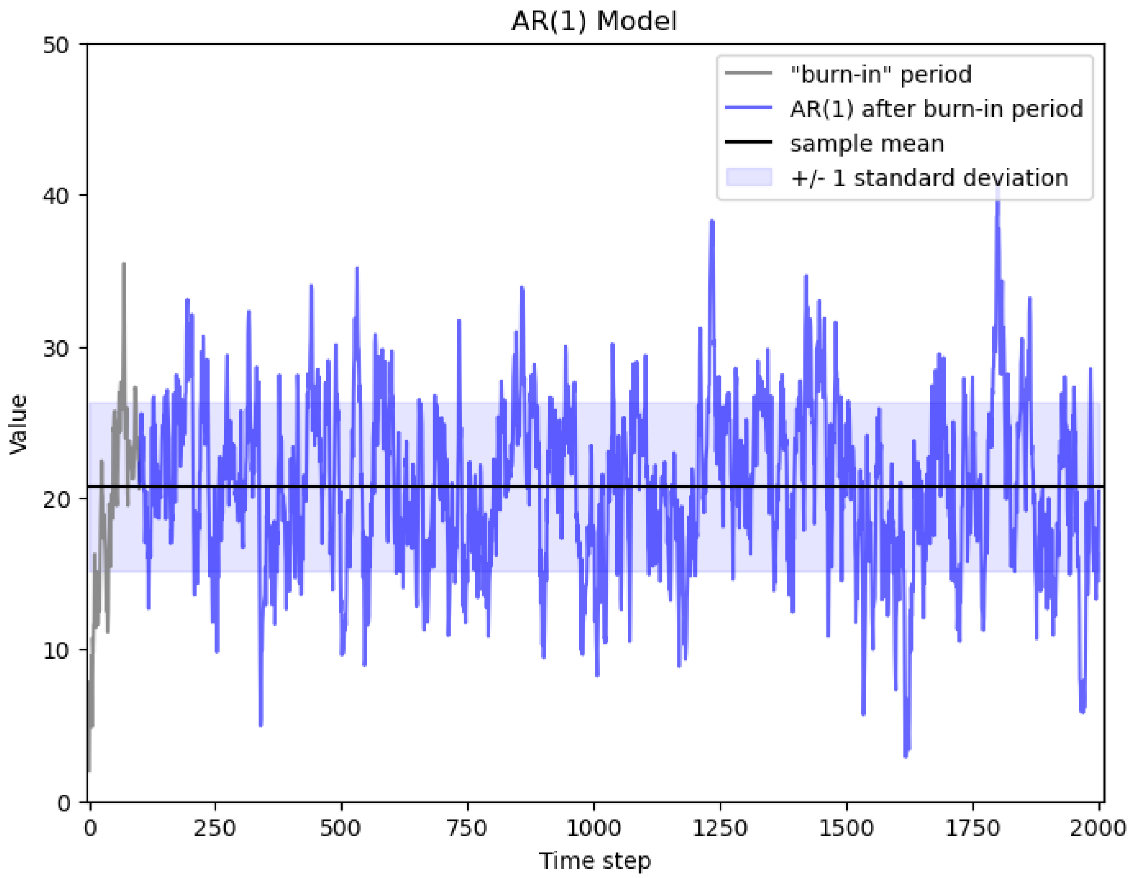

Figure 2: AR(1) model, with "burn-in" period in grey and the remaining time series in blue.

For this model, the empirical mean and variance were 20.8 and 30.41, respectively. The theoretical mean and variance were 20 and 32.895, respectively, giving an error of approximately 4% and 7.55%.

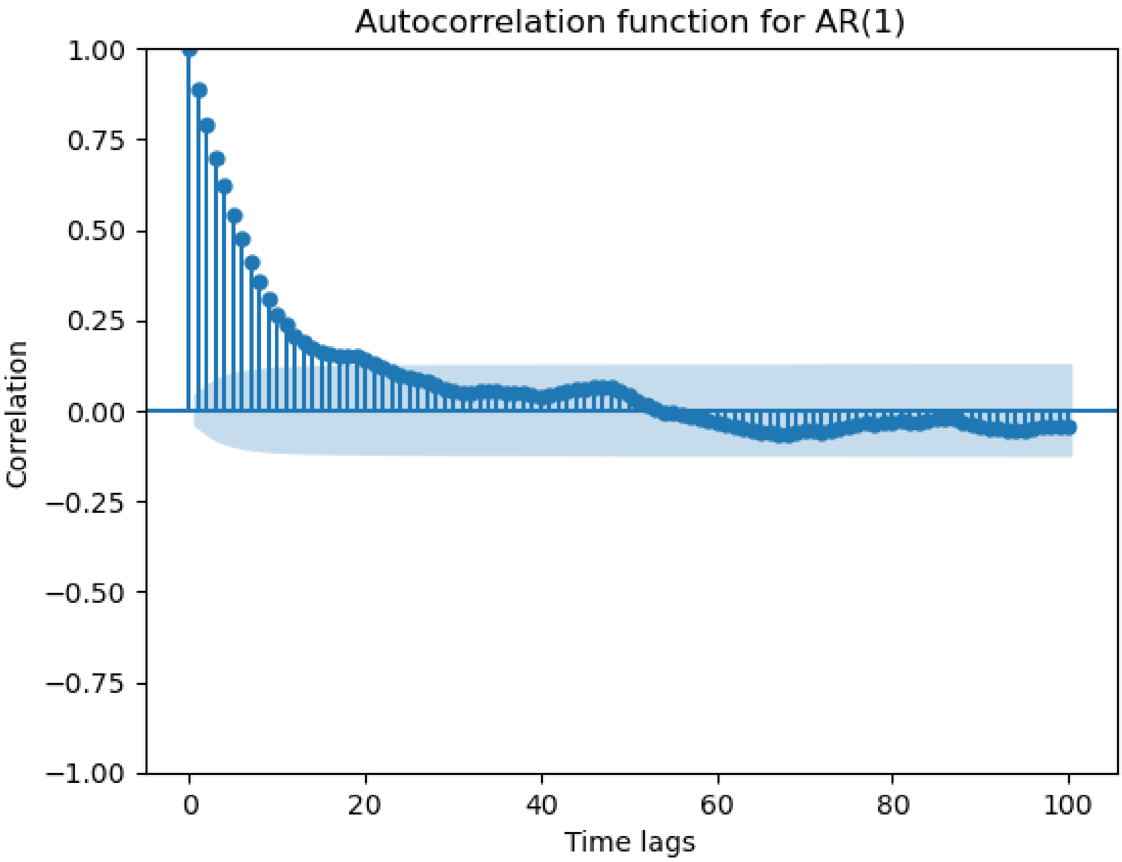

Figure 3: Autocorrelation function for AR(1) model with 100 time lags.

From Figure 3, we can see that the autocorrelation function decreases exponentially with rate a^n until around 20 lags, at which point the process becomes white noise.