1. Answer the following questions.

   (a) (4 points) A birth-and-death Markov chain is a process with countably infinite state space and two types of transitions: births from i to $i+1$ and deaths from i to $i-1$. The transition probabilities can be defined as follows:

$$p_{ij} = \begin{cases} q_i, & \text{if } j = i-1 \\ p_i, & \text{if } j = i+1 \\ 1 - p_i - q_i, & \text{if } j = i \\ 0, & \text{otherwise} \end{cases}$$

   for $0 \leq p_i, q_i$ and $p_i + q_i \leq 1$.

   Determine the transition matrix. Furthermore, give some real-world applications of the Death-and-Birth chains.

   (b) (3 points) Describe the steps Box and Jenkins (1970) suggested should be involved in constructing an ARMA model.

(c) (2 points) You want to fit an ARMA model to a given data. Suppose you first wish to restrict attention to pure models (AR and MA). How would you proceed with identifying lag orders?

(d) (2 points) Describe the Chapman–Kolmogorov relationship in plain terms.

(e) (2 points) Suppose $X_0, X_1, \cdots$ be a Markov chain with transition matrix $\mathbf{P}$ and initial distribution $\boldsymbol{\alpha}$. State the distribution of X_n.

2. Consider the following model:

$$(1 - 0.3B + 1.1B^2)y_t = (1 - 0.5B)w_t.$$

(a) (2 points) Characterize this model as a model in the ARMA(p, q) family, that is, identify p and q.

(b) (2 points) Determine whether this model corresponds to a stationary process. Explain your answer in plain language without providing any proof.

(c) (2 points) Write this model without using backshift notation.

(d) (2 points) Are there any redundant parameters? Justify.

(e) (2 points) Is this model causal? Show this.

(f) (2 points) Is this model invertible? Show this.

3. (5 points) Let the fourth order autoregressive process

$$x_t = \phi_4 x_{t-4} + w_t, \quad 0 < \phi_4 < 1, \quad \text{where } \{w_t\} \overset{iid}{\sim} (0, \sigma_w^2).$$

Determine the autocorrelation function. Comment on the significance of this particular model.

4. (5 points) The number of failures N_t, which occur in a computer network over the time interval $[0, t)$, can be described by a homogeneous Poisson process $\{N_t, t \geq 0\}$. On average, there is a failure after every 4 hours, i.e. the intensity of the process is equal to $\lambda = 0.25$ per hour. What is the probability of at most 1 failure in $[0, 8)$, at least 2 failures in $[8, 16)$, and at most 1 failure in $[16, 24\ )$ (time unit: hour)?

5. Consider a machine that is capable of producing three types of parts. The state of the machine at time period n is denoted by a random variable X_n that takes values in $S = \{0, 1, 2, 3\}$, where 0 means the machine is idle and $i = 1$, 2 or 3 means the machine produces a type i in the time period. Suppose the machine's production schedule is Markovian in the sense that the next type of part it produces, or a possible idle period, depends only on its current state, and the probabilities of these changes do not depend on time. Suppose X_n is a Markov chain with the transition matrix

$$P = \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0.5 \\ 0.3 & 0.4 & 0 & 0.3 \\ 0.4 & 0 & 0.4 & 0.2 \\ 0.3 & 0 & 0.2 & 0.5 \end{bmatrix}$$

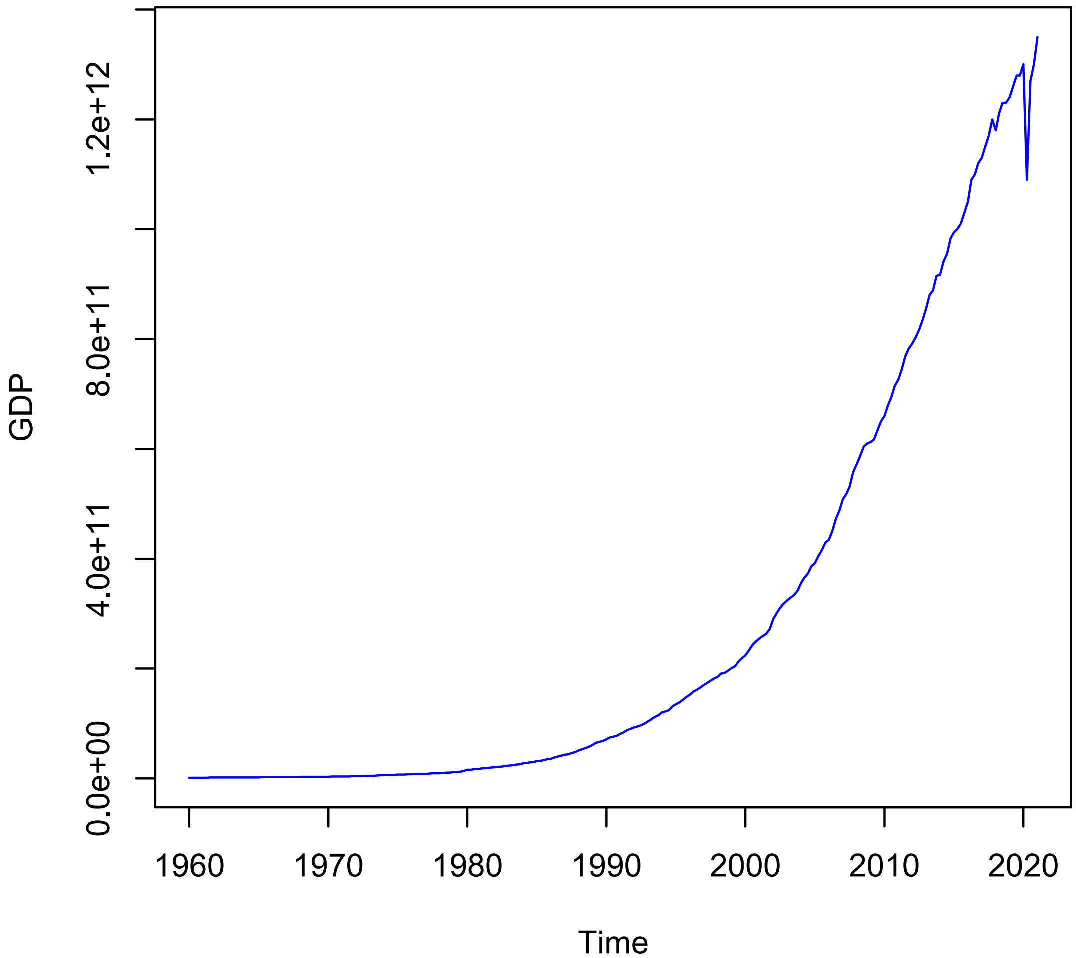
(a) (5 points) Find the stationary distribution, $(\pi_1, \pi_2, \pi_3, \pi_4)$.

(b) (4 points) Determine if the chain is reversible using the stationary distribution in (a). In other words, is the chain in (a) reversible? Prove it if you say yes. In case the chain is not reversible, show it as well. In addition, check all the pairs of states.

6. Based on the current Gross Domestic Product (GDP) price in South Africa from January 1960 to January 2021, the following output was generated.

```
  DATE         GDP
1   1/1/1960 1326466713
2   4/1/1960 1340643343
3   7/1/1960 1358577031
4 10/1/1960 1395671712
5   1/1/1961 1403260968
6   4/1/1961 1398952875
```

(a) (2 points) Share your thoughts on the time series plot of the current Gross Domestic Product (GDP) price in South Africa.



(b) (3 points) Would you please describe the output below? Would you please suggest possible methods to make your data stationary if it is not?

```
Augmented Dickey-Fuller Test

data:  GDP
Dickey-Fuller = -1.0061, Lag order = 6, p-value = 0.936
alternative hypothesis: stationary
```

(c) (5 points) For the GDP data, we fitted three different ARIMA models: $\text{ARIMA}(1,1,1)(1,1,1)_4$, $\text{ARIMA}(1,1,1)(2,1,2)_4$ and $\text{ARIMA}(1,1,0)(0,1,1)_4$. Interpret the results of the model of your choice by comparing the outputs of three different models.

Model 1: ARIMA$(1,1,1)(1,1,1)_4$

```
$degrees_of_freedom
[1] 236

$ttable
       Estimate     SE  t.value p.value
ar1     -0.6022 0.2806  -2.1460  0.0329
ma1      0.4448 0.3073   1.4476  0.1490
sar1     0.0762 0.1047   0.7277  0.4675
sma1    -0.8251 0.0537 -15.3772  0.0000

$AIC
[1] -4.638407

$AICc
[1] -4.637715

$BIC
[1] -4.596789

sigma^2 estimated as 0.0004994:  log likelihood = 568.35,
```

Model 2: ARIMA$(1, 1, 1)(2, 1, 2)_4$

```
$degrees_of_freedom
[1] 234

$ttable
        Estimate     SE t.value p.value
ar1   -0.6024 0.2782 -2.1652  0.0314
ma1    0.4457 0.3046  1.4631  0.1448
sar1  -0.6010 0.3210 -1.8724  0.0624
sar2   0.0968 0.1009  0.9598  0.3381
sma1  -0.1467 0.3107 -0.4721  0.6373
sma2  -0.5705 0.2515 -2.2688  0.0242

$AIC
[1] -4.613625

$AICc
[1] -4.612161

$BIC
[1] -4.51336

sigma^2 estimated as 0.0005086:  log likelihood = 567.56
```

Model 3: ARIMA$(1, 1, 0)(0, 1, 1)_4$

```
$degrees_of_freedom
[1] 238

$ttable
        Estimate     SE  t.value p.value
ar1   -0.1685 0.0643  -2.6189  0.0094
sma1  -0.7911 0.0520 -15.2276  0.0000

$AIC
[1] -4.632101

$AICc
[1] -4.631895

$BIC
[1] -4.58913

sigma^2 estimated as 0.000516:  log likelihood = 565.8
```

7. (3 points) Define a nonhomogeneous Poisson process. Give the distribution, mean, and variance functions for the probability that k events will occur in time interval $(s, s+t]$.

8. (3 points) **Chapman–Kolmogorov relationship:** for all $0 < r < n$, then show that

$$P_{ij}^{n} = \sum_{k} P_{ik}^{r} P_{kj}^{(n-r)}.$$