1. Let Z_t be IID N(0,1), and define

$$X_t = \begin{cases} Z_t, & \text{if t is even} \\ (Z_{t-1}^2 - 1)/\sqrt{2}, & \text{if t is odd} \end{cases}$$

Based on the definition of white noise, show that X_t is WN(0,1). Hint: prove $E(X_t) = 0, \forall t$; $Var(X_t)$ is a constant for all t; X_{t+h} and X_t are uncorrelated for all t and any $h > 0$. You might need some fast facts:

- If $Z_t \sim N(0,1)$, then $Z_t^2 \sim \chi^2(1)$. You can use the mean and variance of $\chi^2(1)$ directly if needed;
- All the odd moments of a normal distribution is 0.
- If Z_t and Z_s are independent, so are Z_t and Z_s^2

$$E(X_t) = \begin{cases} E(Z_t) = 0 \\ E\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}\right) = \frac{E(Z_{t-1}^2) - 1}{\sqrt{2}} = \frac{1-1}{\sqrt{2}} = 0 \end{cases}$$

$$Var(X_t) = \begin{cases} Var(Z_t) = 1 \\ Var\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}\right) = \frac{Var(Z_{t-1}^2)}{2} = \frac{2 \cdot 1}{2} = 1 \end{cases}$$

(Var of chi-square)

- when t, t+h both even:

$$Cov(X_t, X_{t+h}) = Cov(Z_t, Z_{t+h}) = 0$$

- when t, t+h both odd:

$$Cov(X_t, X_{t+h}) = Cov\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}, \frac{Z_{t-1+h}^2 - 1}{\sqrt{2}}\right) = \frac{1}{2} Cov(Z_{t-1}^2, Z_{t-1+h}^2) = 0$$

- when t is even, t+h is odd

$$Cov(X_t, X_{t+h}) = Cov\left(Z_t, \frac{Z_{t-1+h}^2 - 1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} Cov(Z_t, Z_{t-1+h}^2) = 0$$

- when t is odd, t+h is even

$$Cov(X_t, X_{t+h}) = Cov\left(\frac{Z_{t-1}^2 - 1}{\sqrt{2}}, Z_{t+h}\right) = \frac{1}{\sqrt{2}} Cov(Z_{t-1}^2, Z_{t+h}) = 0$$

$$\Rightarrow X_t \sim WN(0,1)$$

2. Let $\{\epsilon_t\} \sim IID\ \mathrm{N}(0, \sigma^2)$, and let a, b and c be constants. For the following processes, (i) calculate the mean and autocovariance function; (ii) specify which, if any, of the processes are stationary.

(a) $X_t = a + b\epsilon_t + c\epsilon_{t-2}$

(b) $X_t = \epsilon_t \epsilon_{t-1}$

(a) $X_t = a + b\epsilon_t + c\epsilon_{t-2}$

$E(X_t) = E(a + b\epsilon_t + c\epsilon_{t-2}) = a$

$\mathrm{cov}(X_{t+h}, X_t) = \gamma(h)$

- when $h=0$:

$\mathrm{cov}(X_{t+h}, X_t) = \mathrm{Var}(X_t) = \mathrm{Var}(a + b\epsilon_t + c\epsilon_{t-2}) = (b^2 + c^2)\sigma^2$

- when $h=1$

$\mathrm{cov}(X_{t+h}, X_t) = \mathrm{cov}(a + b\epsilon_{t+1} + c\epsilon_{t-1}, a + b\epsilon_t + c\epsilon_{t-2})$

$= b^2 \mathrm{cov}(\epsilon_{t+1}, \epsilon_t) + c^2 \mathrm{cov}(\epsilon_{t-1}, \epsilon_{t-2})$

$\{\epsilon_t\} \overset{iid}{\sim} N(0, \sigma^2)$

$+ bc\ \mathrm{cov}(\epsilon_{t+1}, \epsilon_{t-2}) + bc \cdot \mathrm{cov}(\epsilon_{t-1}, \epsilon_t) = 0$

- when $h=2$

$\mathrm{cov}(X_{t+h}, X_t) = \mathrm{cov}(a + b\epsilon_{t+2} + c\epsilon_t, a + b\epsilon_t + c\epsilon_{t-2})$

$= bc\ \mathrm{cov}(\epsilon_t, \epsilon_t) = bc\sigma^2$

- when $h > 2$

$\mathrm{cov}(X_{t+h}, X_t) = 0$

$\Rightarrow$ stationary!! since $E(X_t)$ & autocovariance are independent of t

(b) $X_t = \epsilon_t \epsilon_{t-1}$

$E(X_t) = E(\epsilon_t \epsilon_{t-1}) = 0$

when $h=0$

$$Cov(X_{t+h}, X_t) = Var(X_t) = E(\epsilon_t^2 \epsilon_{t-1}^2) - (\overset{0}{\overset{\parallel}{E(\epsilon_t \epsilon_{t-1})}})^2$$

$$= E(\epsilon_t^2) E(\epsilon_{t-1}^2) = \sigma^2 \cdot \sigma^2 = \sigma^4$$

$\Rightarrow$ Stationary, when $h=0$

3. In the lecture notes, we have proved that the Moving Average process with order 1, or $MA(1)$, is stationary. If the Moving Average process with order q, or $MA(q)$ is defined as

$$X_t = Z_t + \theta_1 Z_{t-1} + ... + \theta_q Z_{t-q}, \{Z_t\} \sim WN(0, \sigma^2)$$

Prove that $MA(2)$ is also stationary, by finding the mean, variance, autocovariance and autocorrelation functions.

$MA(2)$

$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$

$E(X_t) = E(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}) = 0$

$Var(X_t) = (1+\theta_1^2+\theta_2^2)\sigma^2$

- when $h=0$

$cov(X_{t+h}, X_t) = Var(X_t) = (1+\theta_1^2+\theta_2^2)\sigma^2$

- when $h=1$

$cov(X_{t+1}, X_t) = Cov(Z_{t+1}+\theta_1 Z_t + \theta_2 Z_{t-1}, Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2})$

$= cov(Z_{t+1}, Z_t) + \theta_1 cov(Z_{t+1}, Z_{t-1}) + \theta_2 cov(Z_{t+1}, Z_{t-2}) +$

$\theta_1 cov(Z_t, Z_t) + \theta_1^2 cov(Z_t, Z_{t-1}) + \theta_1\theta_2 cov(Z_t, Z_{t-2}) +$

$\theta_2 cov(Z_{t-1}, Z_t) + \theta_1\theta_2 cov(Z_{t-1}, Z_{t-1}) + \theta_2^2 cov(Z_{t-1}, Z_{t-2})$

$= \theta_1\sigma^2 + \theta_1\theta_2\sigma^2 = (\theta_1+\theta_1\theta_2)\sigma^2$

- when $h=2$ (like above)

$cov(X_{t+2}, X_t) = \theta_2\sigma^2$

- when $h \geq 3$

$Cov(X_{t+3}, X_t) = 0$

$\Rightarrow$ stationary. $E(X_t)$, $Var(X_t)$, $cov(X_{t+h}, X_t)$ independent of t

Autocorrelation

$$corr(X_{t+h}, X_t) = \frac{Cov(X_{t+h}, X_t)}{Var(X_t)} = \frac{\gamma(h)}{\gamma(0)} = \rho(h)$$

$$\rho(h) = \begin{cases} 1 & \text{when } h=0 \\ \dfrac{\theta_1 + \theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{when } h=1 \\ \dfrac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{when } h=2 \\ 0 & \text{when } h\geq 3 \end{cases}$$

4. For the time series given below,

$$x_0 = 4, x_1 = 6.1, x_2 = 5.2, x_3 = 6.5, x_4 = 8.9, x_5 = 8, x_6 = 8.2, x_7 = 11.4, x_8 = 10$$

use the classical decomposition method to calculate by hand, the estimates of trend component m_t, seasonal component s_t and random noise ϵ_t, using a seasonal lag $h = 3$ and an additive model. (Hint: for the observations that can not be estimated by this method, leave it blank; choose the centered MA filter with order =3)

$x_t - m_t$ $\qquad$ $x_T - m_t - S_t$

obs t	x_t	m_t	d_t	s_t	ϵ_t
0	4			-0,685	
1	6,1	5.1	1	1,21	-0,21
2	5,2	5.93	-0,73	-0,55	-0,18
3	6,5	6,87	-0,37	-0.685	0,315
4	8.9	7.8	1.1	1,21	-0,11
5	8	8,37	-0,37	-0,55	0,18
6	8,2	9,2	-1	-0,685	-0,315
7	11.4	9.87	1,53	1,21	0,32
8	10			-0,55	

5. For each of the three time series plots (a), (b) and (c), pick an most matching autocorrelation plot from (d), (e) and (f).

(a) → (f)

(b) → (d)

(c) → (e)

6. The random walk $\{S_t, t = 0, 1, ..., n\}$ is obtained by cumulatively summing iid random variables. Define $S_0 = 0$,

$$S_t = X_1 + X_2 + ... + X_t, t = 1, 2, ..., n$$

and $\{X_t\}$ is an IID binary process with

$$P(X_t = 1) = 0.7, P(X_t = -1) = 0.3$$

(a) Write Python code to simulate this process with $n = 1000$. Do not provide the data in the homework answer. Include a screenshot of the codes for simulation in the pdf.

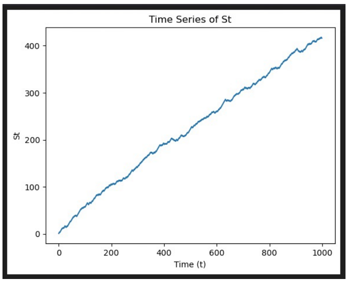
(b) Use the simulated data to draw a time series line plot, attach the plot in the pdf file, and describe the trend you have observed.

(c) Draw two plots of the smoothed trend using MA filter with k=30 and k=365 and attach the plots in the pdf file. What have you observed about the smoothness and trend in the two plots?

(a)

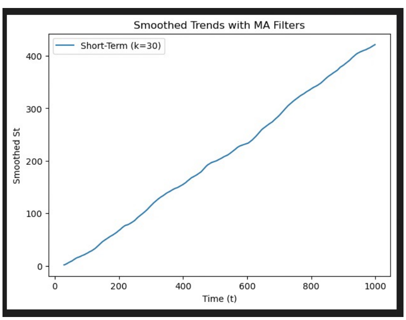
```
n = 1000
p = 0.7  # Probability of Xt = 1
q = 0.3  # Probability of Xt = -1

Xt = np.random.choice([1, -1], size=n, p=[p, q])
St = np.cumsum(Xt)
df = pd.DataFrame({'t': range(n), 'Xt': Xt, 'St': St})
```

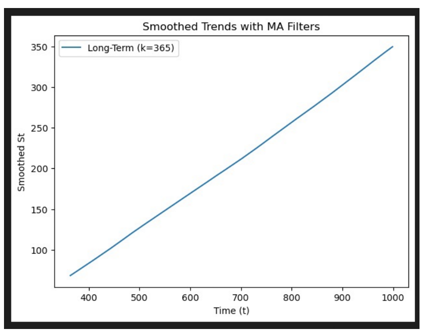
(b)



Since p(Xt=1)=0.7, Therefore, the time series line will shows a long-term trend of upward line. though there's some upward & downward walk

(c)





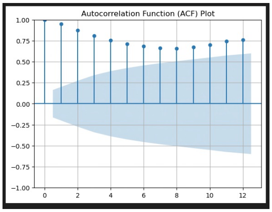
K=365 is smoother than k=30

7. The dataset airline-passengers.csv recorded the total number of airline passengers over time. The units are a count of the number of airline passengers in thousands. There are 144 monthly observations from 1949 to 1960.

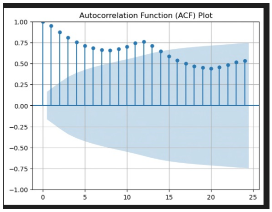
(a) Draw the time series plot and ACF plot. Attach the plots in the pdf. Try your best using the time series plot to explain why the general trend of the sample autocorrelation in ACF plot is decreasing to negative and then increase to around 0 when lag h is increasing, and why there are "jumps" around lags $h = 12, 24, 36, \ldots$.

(b) Again, based on the time series plot and the discussion above, choose the "model" and "period" for classical decomposition and create the decomposition plot. Include your choice of "model" and "freq" and the plot in the pdf.

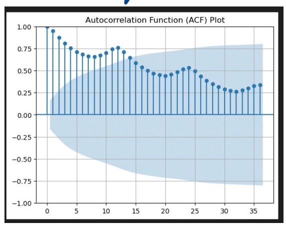
(a) lag=12

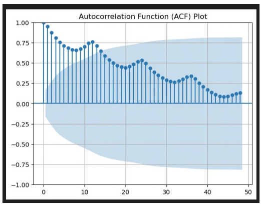


lag=24

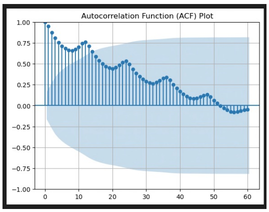


lag=36





lag=48



lag=60

I think that is because the airline trend is basically repeated yearly. so when we try to increase the lag, ACF plot will decreases. that is also why there are 'jumps' around h=12, 24...

(b) model: multiplicative freq: 12