Time Series Analysis: Solutions to exercise 1

January 17, 2023

1. (a) This process is strongly stationary, this can be deduced by the definition of the process. Indeed any joint distribution of finitely many points of the process is invariant to time shift because a shift in the white noise process W_t does not change the specifications of the process. This also implies implies weak stationarity.

   (b) yes, see a)

2. The cos function has period 2π. Letting $A' \in [0, 2\pi]$ be such that $A = k2\pi + A'$ with $k \in \mathcal{Z}$ gives $X_t = a\cos(A't + V)$,so that we will assume. $A \in [0, 2\pi]$. We have then

$$\begin{aligned} E(X_t) &= aE(\cos(At + V)) \\ &= a\int_{-\infty}^{\infty} \cos(x) f_{At+V}(x)dx \\ &= a\int_{-\infty}^{\infty} \cos(x) \int_{-\infty}^{\infty} f_V(x-s) f_{At}(s) ds dx \\ &= 0, \end{aligned}$$

   where we used Fubini's theorem and

$$\int_{-\infty}^{\infty} \cos(x) f_V(x-s)dx = \frac{1}{2\pi}\int_{s}^{s+2\pi} \cos(x)dx = 0.$$

   Moreover, for the autocovariance we have

$$\begin{aligned} \gamma(s,t) &= E(X_s X_t) \\ &= a^2 E(\cos(At+V)\cos(As+V)) \\ &= \frac{a^2}{2}\{E(\cos(A(s+t)+2V) + E(\cos(A(s-t))))\}, \end{aligned}$$

   where the first term is zero and the second one only depends on the difference $|s-t|$. So the process is weakly stationary.

3. (a)

$$\frac{1}{6}\sigma^2$$

   (b)

$$\gamma_{X,Y}(s,t) = \begin{cases} \frac{1}{3}\sigma^2 & \text{if } s = t \\ \frac{1}{6}\sigma^2 & \text{if } |s-t| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

4. (a) By definition and symmetry of the covariance $\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_t, X_{t+h}) = \gamma(-h)$ and then $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(-h)}{\gamma(0)} = \rho(-h)$.

   (b) As before, $\gamma_{XY}(h) = \text{Cov}(X_{t+h}, Y_t) = \text{Cov}(Y_t, X_{t+h}) = \gamma_{YX}(-h)$, but there is no reason for the cross-covariance to be symmetric around zero.

5. (a) False

   (b) True

   (c) True

   (d) False

6. (a) This time series cannot have constant variance and thus cannot be weakly stationary due to its exponential growth.

   (b) This time series is a moving average and is automatically stationary.

   (c) This time series is not weakly stationary since $\text{Cov}(X_1, X_2) = \sigma^2$, but $\text{Cov}(X_2, X_3) = 0$, so the covariance function is not a function of only the shift in time.

   (d) This time series is weakly stationary with auto-covariance function

$$\gamma(h) = \begin{cases} \sigma^2 & \text{if } h \text{ even} \\ 0 & \text{if } h \text{ odd} \end{cases}$$

   and the process has constant mean 0. By definition of the process it is also strongly stationary.

   (e) Since W_t has by symmetry the same distribution as W_t then the process X_t has the same distribution as W_t, so it is weakly and strongly stationary.

   (f) By construction this process is weekly stationary since it is uncorrelated and has mean zero. Nevertheless it is not strongly stationary since marginally the distribution are different by the hypotheses of the exercise.

7. (a) $(1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}$.

   (b) $(1 - B)^2 X_t = 2 + W_t - 2W_{t-1} + W_{t-2}$

   (c) $\binom{n}{2}$

   (d) $(1 + B)(1 - B) = 1 - B + B - B^2 = 1 - B^2$.

8. (a) $\frac{0.1\sigma^2}{1-0.9^2}$

   (b) $-\frac{0.1\sigma^2}{1-0.9^2}$.

9.

$$\begin{aligned}
Y_t - S_t &= Y_t - \frac{1}{2}(Y_t + Y_{t-1}) \\
&= \frac{1}{2}(2Y_t - (Y_t + Y_{t-1})) \\
&= \frac{1}{2}(Y_t - Y_{t-1}) \\
&= \frac{1}{2}(1-B)X_t
\end{aligned}$$

10. (a) We have $E((1-B)Y_t) = \beta_1$ and $\mathrm{Cov}(DY_t, DY_{t+h}) = 2\gamma(h) - \gamma(h-1) - \gamma(h+1)$, then $(1-B)Y_t$ is stationary.

(b) By using the relation $(t-1)^r = \sum_{i=0}^{r} \binom{r}{i} t^{r-1}(-1)^i$, we have

$$(1-B)Y_t \sum_{r=0}^{k-1} \beta_r t^r - \beta_k \sum_{i=1}^{k} \binom{k}{i} t^{k-i}(-1)^i - \sum_{s=0}^{k-1} \beta_s \sum_{i=1}^{s} \binom{s}{i} t^{s-i}(-1)^i + \varepsilon_t - \varepsilon_{t-1}.$$

Hence $E((1-B)Y_t)$ is a polynomial of degree at most $k-1$ in t. Consider then $Y - t = p(t) + \varepsilon_t$, where $p(t) = \sum_{r=0}^{k} \beta_r t^r$. We will show by induction that $(1-B)^j Y_t$ has a polynomial trend of degree at most $k-j$. We just have shown that $(1-B)Y_t$ has a polynomial trend of degree at most $k-1$. So we assume that $D^j Y_t$ has a polynomial trend of degree at most $k-j$ and show that $(1-B)^{j+1}Y_t$ has a polynomial trend of degree at most $k-j-1$.

We have $(1-B)^{j+1}Y_t = (1-B)(1-B)^j Y_t$, by the induction hypothesis, $(1-B)^j Y_t$ has a polynomial trend of degree at most $k-j$. As taking the first difference reduces the degree of the trend by at least 1, regardless of the former degree, $(1-B)^{j+1}Y_t$ has a polynomial trend of degree at most $k-j-1$ and the induction step is established. We obtain that $(1-B)^k Y_t$ has a polynomial trend of degree at most $k-k=0$, thus $E((1-B)^k Y_t)$ is constant. Moreover, since

$$D^k Y_t = \sum_{j=0}^{k} \binom{k}{j} (-1)^j Y_{t-j},$$

we have

$$\mathrm{Cov}((1-B)^k Y_t, (1-B)^k Y_{t+h}) = \sum_{i,j=0}^{k} \binom{k}{i}\binom{k}{j} (-1)^{i+j} \gamma(h+j-i),$$

and the right hand side is not a function t but only a function of the time-lag h.

(c) We have

$$(1-B)^k Y_t = \sum_{j=0}^{k} (-1)^j \binom{k}{j} [p(t-j) + \varepsilon_{t-j}] = c_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_k \varepsilon_{t-k},$$

where $c_0 = \sum_{j=0}^{k} (-1)^j \binom{k}{j} p(t-j)$ and $\theta_j = (-1)^j \binom{k}{j}$ for $j = 1, \ldots, k$. So $(1-B)^k$ is an $MA(k)$ process.

(d) See solution of (b).

11. (a) With $\theta = (\theta_{12}, \theta_{21})$ we have

$$\begin{aligned} L(\theta) &= P(Y_1 = y_1, \ldots, Y_n = y_n) \\ &= P(Y_1 = y_1) \prod_{t=2}^{n} P(Y_t | Y_{t-1} = y_{t-1}) \\ &= P(Y_1 = y_1) \theta_{12}^{n_{12}} (1 - \theta_{12})^{n_{11}} \theta_{21}^{n_{21}} (1 - \theta_{21})^{n_{22}}. \end{aligned}$$

If we ignore the first term $\log P(Y_1 = y_1)$, we have

$$\log L(\theta) = \ell(\theta) = n_{12} \log \theta_{12} + n_{11} \log(1 - \theta_{12}) + n_{21} \log \theta_{21} + n_{22} \log(1 - \theta_{21}).$$

(b) Maximizing the likelihood yields $\hat{\theta}_{12} = n_{12}/(n_{12} + n_{11})$ and $\hat{\theta}_{21} = n_{21}/(n_{21} + n_{22})$. The asymptotic variances are the diagonal entries of the inverse Fisher information matrix (this part is outside the course) and this yields

$$\mathrm{Var}(\hat{\theta}_{12}) = \frac{n_{11} n_{12}}{(n_{12} + n_{11})^2}.$$

$$\mathrm{Var}(\hat{\theta}_{21}) = \frac{n_{22} n_{21}}{(n_{21} + n_{22})^2}.$$

If the model depends on only one parameter $\theta' = \theta_{12} = \theta_{21}$, then the likelihood becomes

$$\ell(\theta') = (n_{12} + n_{21}) \log(\theta') + (n_{11} + n_{22}) \log(1 - \theta').$$

Hence we get $\hat{\theta}' = (n_{12} + n_{21})/(n_{12} + n_{11} + n_{21} + n_{22})$.

(c) We can use the likelihood ration statistic W to compare the model with only θ' as parameter with the model parametrized by $\theta = (\theta_1, \theta_2)$.

$$W = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}')).$$

If the simpler model is correct, then the asymptotic distribution of W will be χ_1^2 and we can compare the observed value of W against this distribution (which is identical to the distribution of the square of a standard normal distributed random variable).

(d) We have $n_{11} = 5, n_{12} = 9, n_{21} = 9, n_{22} = 2$. This yields $\hat{\theta}_{12} = 0.64, \hat{\theta}_{21} = 0.82$. The value of the likelihood ratio statistic W is 5.62, which corresponds to a p-value of about 0.0178, as $P(Z > 5.62) = 0.0178$ if Z has a χ_1^2 distribution. We can hence reject the null hypothesis that the simpler model with just one parameter is adequate at level 0.05.

12. (a)

$$\begin{aligned}
\mathrm{Var}(\overline{Y}) &= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathrm{Var}(Y_i) + 2\sum_{i<j}\mathrm{Cov}(Y_i, Y_j)\right\} \\
&= \frac{1}{n^2}\left\{n\gamma(0) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathrm{Cov}(Y_i, Y_j)\right\} \\
&= \frac{1}{n^2}\left\{n\gamma(0) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma(i-j)\right\}
\end{aligned}$$

If we define $h = j - i$, then

$$\begin{aligned}
\mathrm{Var}(\overline{Y}) &= \frac{1}{n^2}\left\{n\gamma(0) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\gamma(h)\right\} \\
&= \frac{1}{n^2}\left\{n\gamma(0) + 2\sum_{h=1}^{n-1}(n-h)\gamma(h)\right\} \\
&= \frac{\gamma(0)}{n}\left\{1 + \frac{2}{n}\sum_{h=1}^{n-1}(n-h)\rho(h)\right\}.
\end{aligned}$$

(b)

$$\sum_{h=1}^{n-1}(n-h)\alpha^h = n\sum_{h=1}^{n-1}\alpha^h - \sum_{h=1}^{n-1}h\alpha^h = n\alpha\frac{1-\alpha^n}{1-\alpha} - \sum_{h=1}^{n-1}h\alpha^h,$$

where we use the relation $\sum_{i=0}^{n-1} r^i = (1-r^n)/(1-r)$ for $|r| < 1$ and the hint giving

$$\sum_{h=1}^{n-1}h\alpha^h = \alpha\sum_{h=1}^{n-1}h\alpha^{h-1} \to \frac{\alpha}{(1-\alpha)^2}.$$

Hence

$$\begin{aligned}
\mathrm{Var}(\overline{Y}) &= \frac{1}{n}\frac{\sigma^2}{1-\alpha^2}\left[1 + \frac{2}{n}\left\{n\alpha\frac{1-\alpha^n}{1-\alpha} - \sum_{h=1}^{n-1}h\alpha^h\right\}\right] \\
&= \frac{1}{n}\frac{\sigma^2}{1-\alpha^2}\left[1 + \frac{2}{\alpha}\frac{1-\alpha^n}{1-\alpha}\right] + O(n^{-2}) \\
&\approx \frac{1}{n}\frac{\sigma^2}{1-\alpha^2}\left[1 + \frac{2\alpha}{1-\alpha}\right] = \frac{\sigma^2}{n(1-\alpha)}
\end{aligned}$$

When $\alpha \to 1$, the variance explodes for fixed values of n. When $\alpha \to -1$, the variance becomes proportional to $(4n)^{-1}\sigma^2$. Finally, in the case of $\alpha = 0$, the variance is simply $n^{-1}\sigma^2$.

13. (a) Let us assume that the process Y_t is weakly stationary, and let $E(Y_t) = \mu$ and $\text{Cov}(Y_t, Y_{t+h}) = \gamma(h)$. Then $E(Y_t^2) = \mu^2 + \gamma(0)$ and

$$V(h) = \frac{1}{2}E(Y_t^2) - E(Y_t Y_{t+h}) + \frac{1}{2}E(Y_t^2) = \gamma(0) + \mu^2 - (\gamma(h) + \mu^2) = \gamma(0)(1 - \rho(h).$$

If we have a white noise process, $V(h) = \gamma(0)1\{h \neq 0\}$, and if we have $\rho(h) = \exp(-\lambda|h|)$, then $V(h) = \gamma(0)(1 - \exp(-\lambda|h|)$.

(b) v_{ij} is an unbiased estimator of $V(h)$, where $h = |t_i - t_j|$. SMoothing is useful for estimating $V(h)$ for the values of h where no pair (t_i, t_j) with $|t_i - t_j| = h$ exists.

(c) $\overline{v}_h$ is an unbiased estimator of $V(h)$ where $h = |t_i - t_j|$. Moreover, by the Cuchy-Schwarz inequality we obtain that $\text{Var}(\overline{v}_h) \leq \text{Var}(v_{ij})$.

14. Since $Y_1, \ldots, Y_n$ satisfy the Markov property,

$$\begin{aligned}
P(Y_1 = y_1, \ldots, Y_n = y_n) &= P(Y_1 = y_1) \prod_{r=2}^{n} P(Y_1 = y_1, \ldots, Y_r = y_r | Y_1 = y_1, \ldots Y_{r-1} = y_{r-1}) \\
&= P(Y_1 = y_1) \prod_{r=2}^{n} P(Y_r = y_r | Y_1 = y_1, \ldots Y_{r-1} = y_{r-1}) \\
&= P(Y_1 = y_1) \prod_{r=2}^{n} P(Y_r = y_r | Y_{r-1} = y_{r-1})
\end{aligned}$$

15.

$$\begin{aligned}
P(Y_t \leq y_t | Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1) &= P(\mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} + mu) + \varepsilon_t \leq y_t | Y_{t-1} = y_{t-1}, \ldots) \\
&= P(\varepsilon_t \leq y_t - \mu - \phi_1(y_{t-1} - \mu) - \phi_2(y_{t-2} + mu)) \\
&= P(Y_t \leq y_t | Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2})
\end{aligned}$$

So the process is a Markov process of order 2 and

$$Y_t | Y_{t-1}, \ldots, Y_1 \quad = \quad Y_t | Y_{t-1}, Y_{t-2} \quad \sim \quad quad\mathcal{N}(mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} + mu), \sigma^2).$$

This implies that if we denote by $f(y_1, y_2; \mu, \phi_1, \phi_2, \sigma^2)$ the joint likelihood of the initial observations y_1 and y_2, then the likelihood function for the parameters $(\mu, \phi_1, \phi_2, \sigma^2)$ takes the form

$$L(\mu, \phi_1, \phi_2, \sigma^2) = (y_1, y_2; \mu, \phi_1, \phi_2, \sigma^2) \prod_{i=3}^{n} f(y_t | y_{t-1}, y_{t-2}; \mu, \phi_1, \phi_2, \sigma^2),$$

where

$$f(y_t | y_{t-1}, y_{t-2}; \mu, \phi_1, \phi_2, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \mu - \phi_1(y_{t-1} - \mu) - \phi_2(y_{t-2} - \mu))^2\right\}.$$

16. (a) The roots are respectively

$$\begin{aligned}
z_1 = 0.8708287, z_2 = -2.8708287, \\
z_1 = 3 + i, z_2 = 3 - i, \\
z_1 = 1.268790, z_2 = -1.125932.
\end{aligned}$$

(b) The first polynomial has two real roots, one root in the unit disk making the process not stationary. The second polynomial has complex roots which lie outside of the unit disk, in this case the process is stationary and we do observe an oscillating behaviour. The last polynomial has two real roots outside the unit disk, the associated time series is then stationary but no oscillating behaviour happens because no root is complex.

17. The first sufficient (and necessary) condition is that all eigenvalues of the matrix Φ should lie inside the unit disk to prevent an exploding behaviour (the reasoning and proof is similar as in the theorem of the Lecture Notes for univariate process). The sufficient condition for causality is that the vector $(W_t^{(1)}, W_t^{(2)})$ should be independent of the previous process values X_{t-1} which is the case in our setting.