Assignment 4

2022-12-10

Question 1

a) We have $\mathrm{E}(X_t) = \mathrm{E}(Y - 2Z_{t-1} + Z_t)$ and since Y is independent of $\{Z_t\}$, it follows that

$$\mathrm{E}(X_t) = \mathrm{E}(Y - 2Z_{t-1} + Z_t) = \mathrm{E}(Y) - 2\,\mathrm{E}(Z_{t-1} + \mathrm{E}(Z_t)) = 1.$$

Furthermore,

$$\mathrm{E}(X_t X_{t+h}) = \mathrm{E}((Y - 2Z_{t-1} + Z_t)(Y - 2Z_{t+h-1} + Z_{t+h})) = \mathrm{E}(Y^2) = \mathrm{Var}(Y) - \mathrm{E}(Y)^2 = 0$$

because in the expansion, any term with Z_{t+k} will have 0 expectation. Thus, $\{X_t\}$ is stationary. We can see that $\{Y_t\}$ is not stationary because X_t depends on t and thus, $\mathrm{E}(Y_t)$ depends on t.

b) We have $\nabla_d(W_t + 1) = W_t + 1 - W_{t-d} - 1 = W_t - W_{t-d}$ so it follows that

$$\begin{aligned}
\nabla_d^2(W_t + 1) &= \nabla_d(W_t - W_{t-d}) \\
&= W_t - W_{t-d} - W_{t-d} + W_{t-2d} \\
&= W_t - 2W_{t-d} + W_{t-2d} \\
&= at^2 + bts_t + X_t - 2(a(t-d)^2 + b(t-d)s_t + X_{t-d}) + a(t-2d)^2 + b(t-2d)s_t + X_{t-2d} \\
&= a(t^2 - 2t^2 + 4dt - 2d^2 + t^2 - 4dt + 4d^2) + bs_t(t - 2t + 2d + t - 2d) + X_t - 2X_{t-d} + X_{t-2d} \\
&= ad^2 + bs_t 0 + X_t - 2X_{t-d} + X_{t-2d} = ad^2 + X_t - 2X_{t-d} + X_{t-2d}.
\end{aligned}$$

If $\{X_t\}$ is stationary then $Y_t = ad^2 + (X_t - X_{t-d}) - (X_{t-d} - X_{t-2d})$ is also stationary because $\{(X_t - X_{t-d})\}$ and $\{(X_{t-d} - X_{t-2d})\}$ are stationary.

Question 2

a) Note that $aY_{t-1} = 2aX_{t-1} - aX_{t-2}$ so

$$\begin{aligned}
aY_t &= 2aX_t - aX_{t-1} \\
&= 2a(0.2X_{t-1} + Z_t - Z_{t-1}) - a(0.2X_{t-2} + Z_{t-1} - Z_{t-2}) \\
&= 0.4aX_{t-1} - 0.2aX_{t-2} - aZ_{t-1} + aZ_{t-2} + 2aZ_t \\
&= \frac{1}{5}(2aX_{t-1} - aX_{t-2}) - aZ_{t-1} + aZ_{t-2} + 2aZ_t \\
&= \frac{1}{5}aY_{t-1} - aZ_{t-1} + aZ_{t-2} + 2aZ_t.
\end{aligned}$$

Let $a = 1/2$, we have

$$aY_t = \frac{1}{5}aY_{t-1} - aZ_{t-1} + aZ_{t-2} + Z_t$$

so $\{aY_t\}$ follows an ARMA(p, q) model for $a = 1/2$, $\alpha_1 = 1/5$, $\beta_1 = \beta_2 = a$, $p = 1$, and $q = 2$.

b) We have

$$\begin{aligned}
\operatorname{Cov}(X_t, Z_t) &= \mathrm{E}[(X_t - \mathrm{E}(X_t))(Z_t - \mathrm{E}(Z_t))] \\
&= \mathrm{E}[(X_t - \mathrm{E}(X_t))Z_t] \\
&= \mathrm{E}(X_t Z_t) - \mathrm{E}[Z_t \, \mathrm{E}(X_t)].
\end{aligned}$$

Assuming that $\{X_t\}$ is stationary,

$$\operatorname{Cov}(X_t, Z_t) = \mathrm{E}(X_t Z_t) - \mathrm{E}[Z_t \, \mathrm{E}(X_t)] = \mathrm{E}(X_t Z_t) - \mathrm{E}(Z_t)\, \mathrm{E}(X_t)$$

so

Question 3

a) The autocovariance function

$$\gamma_Z(h) = \begin{cases} \sigma^2 & \text{for } h = 0 \\ 0 & \text{for } h \neq 0 \end{cases}$$

then the spectral density

$$f_Z(\lambda) = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \gamma_Z(h) e^{-ih\lambda} = \frac{1}{2\pi} \sigma^2.$$

b) $\{X_t\}$ is an MA(q) time series by introducing new WN $Z_t^{'} = Z_{t+1}$ We have

$$\mathrm{E}(X_t) = \mathrm{E}\left(\frac{1}{2} Z_{t+1} + Z_t - \frac{1}{2} Z_{t-1}\right) = \frac{1}{2}\, \mathrm{E}(Z_{t+1}) + \mathrm{E}(Z_t) - \frac{1}{2}\, \mathrm{E}(Z_{t-1}) = \sigma^2$$

not dependent on t. Doing the same thing for $\mathrm{E}(X_t X_{t+h})$ to see that the expansion only contains Z_t terms so it is also independent of t. Thus, $\{X_t\}$ is stationary.

The autocovariance function

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{i=0}^{2-h} \beta_i \beta_{i+h} & h = 0, 1, \ldots, 2, \\ \gamma_X(-h) & h = -1, \ldots, -2, \\ 0, & \text{otherwise} \end{cases}$$

and the spectral density

$$
\begin{aligned}
f_X(\lambda) &= \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\gamma_X(h)e^{-ih\lambda} \\
&= \frac{1}{2\pi}\sum_{h=-2}^{2}\gamma_X(h)e^{-ih\lambda} \\
&= \frac{1}{2\pi}\left(\sum_{h=-2}^{-1}\gamma_X(h)e^{-ih\lambda}+\sum_{h=0}^{2}\gamma_X(h)e^{-ih\lambda}\right) \\
&= \frac{1}{2\pi}\left(\gamma_X(0)+2\sum_{h=1}^{2}\gamma_X(h)e^{-ih\lambda}\right) \\
&= \frac{1}{2\pi}\left(\sigma^2\sum_{i=0}^{2}\beta_i^2+2\sum_{h=1}^{2}\gamma_X(h)e^{-ih\lambda}\right) \\
&= \frac{1}{2\pi}\left(\frac{\sigma^2}{2}+2\gamma_X(1)e^{-i\lambda}+2\gamma_X(2)e^{-i2\lambda}\right) \\
&= \frac{1}{2\pi}\left(\frac{\sigma^2}{2}+2\sigma^2(\beta_0\beta_1+\beta_1\beta_2)e^{-i\lambda}+2\sigma^2(\beta_0\beta_2)e^{-i2\lambda}\right) \\
&= \frac{1}{2\pi}\left(\frac{\sigma^2}{2}+2\sigma^2(0)e^{-i\lambda}+-\frac{\sigma^2}{2}e^{-i2\lambda}\right) \\
&= \frac{1}{2\pi}\left(\frac{\sigma^2}{2}+-\frac{\sigma^2}{2}e^{-i2\lambda}\right)
\end{aligned}
$$

c) $\{X_t\}$ is a linear transformation of $\{Z_t\}$ because $X_t = \frac{1}{2}Z_t^{'}+Z_{t-1}^{'}-\frac{1}{2}Z_{t-2}^{'}$ which is a linear combination of $\{Z_t^{'}\}$ by letting $Z_t^{'}=Z_{t+1}$. The corresponding filter coefficients $\psi_0 = 1/2$, $\psi_1 = 1$, $\psi_2 = -1/2$, and $\psi_k = 0$ for $k \notin \{0,1,2\}$. The transfer function

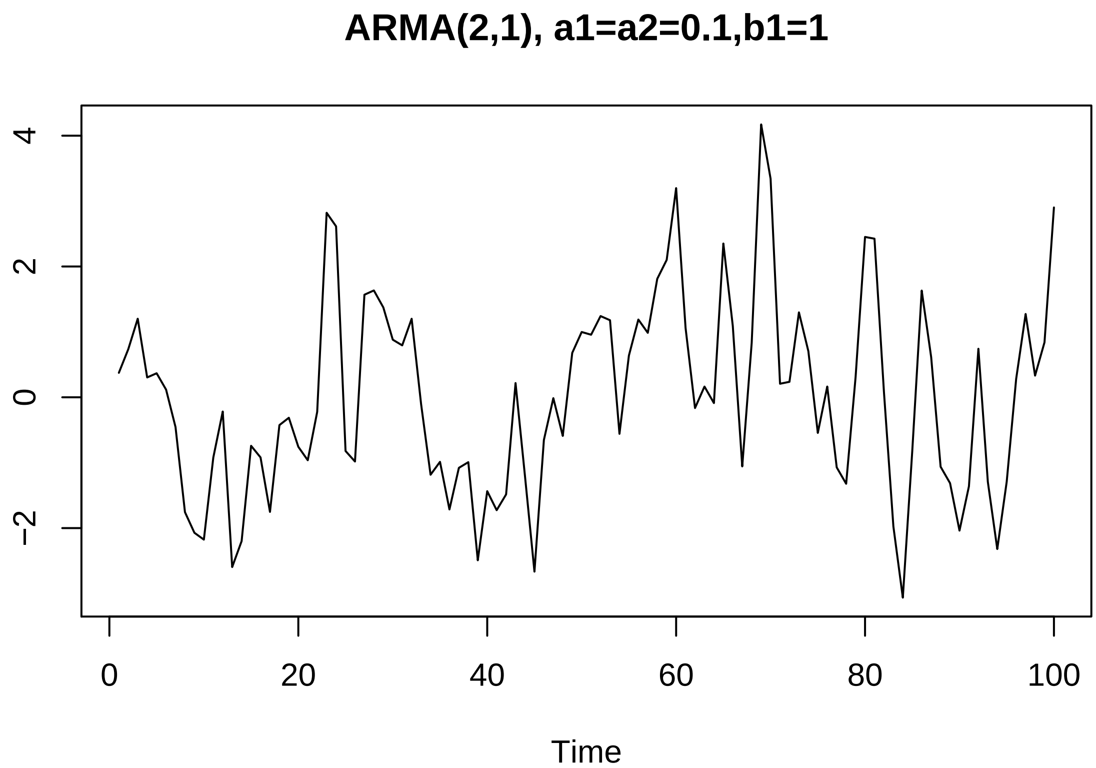
$$\psi(\lambda) = \sum_j \psi_j e^{-ij\lambda} = \frac{1}{2}+e^{-i\lambda}-\frac{1}{2}e^{-2\lambda i}.$$

Question 4

a)

Let $n = 100$, $p = 2$, and $q = 1$ with $\alpha_1 = \alpha_2 = 0.1$, $\beta_1 = 1$ then the plot of the time series is

```
arma1=arima.sim(100,model=list(ar=c(0.1,0.1),ma=1))
plot(arma1,ylab="",main="ARMA(2,1), a1=a2=0.1,b1=1")
```



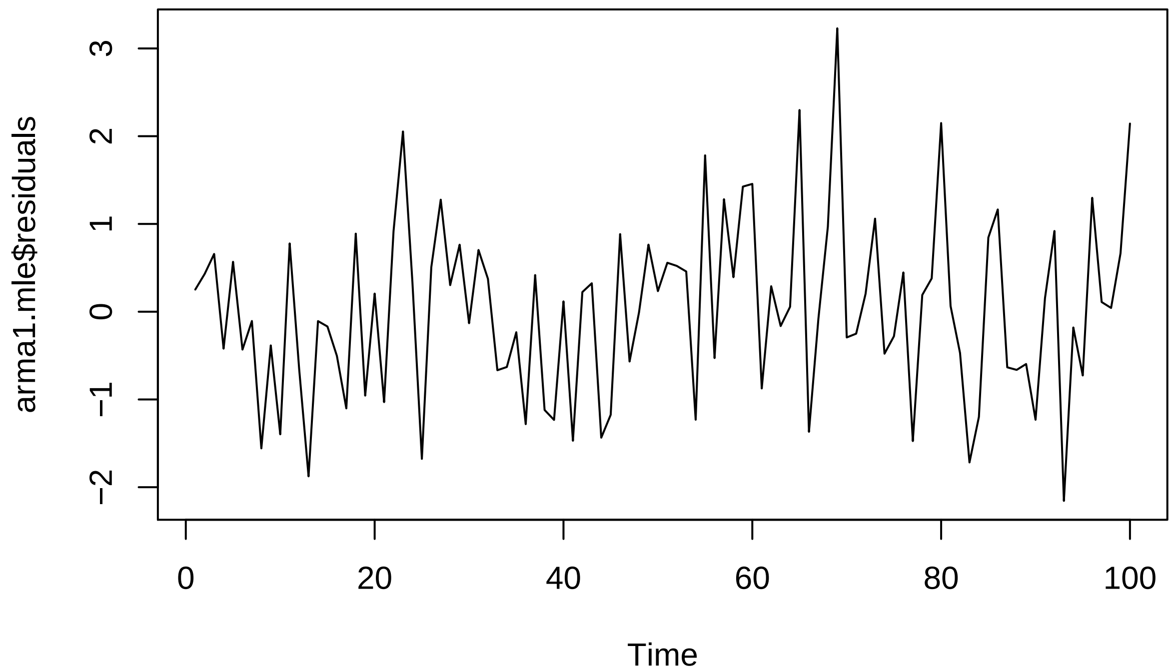
We then find the estimates of the model's parameters.

```
arma1.mle = arima(arma1,order=c(2,0,1),method="ML",include.mean=F);
arma1.mle
```

```
##
## Call:
## arima(x = arma1, order = c(2, 0, 1), include.mean = F, method = "ML")
##
## Coefficients:
##          ar1     ar2     ma1
##       0.1297  0.0719  0.9397
## s.e.  0.1468  0.1451  0.1230
##
## sigma^2 estimated as 1.011:  log likelihood = -143.59,  aic = 295.17
```

The estimated parameters are $\hat{\alpha_1} = 0.16$, $\hat{\alpha_2} = 0.088$, and $\hat{\beta}_1 = 0.947$ which are pretty different from the true ones. The residuals plot is

```
plot(arma1.mle$residuals)
```



There is no visible pattern to the residual plot which suggests that the estimated coefficients are appropriate. We use the Portmanteau test to evaluate the quality of the fit

```
Box.test(resid(arma1.mle),type="Box-Pierce")$p.value
```

```
## [1] 0.9444947
```

The returned p-value is 0.792 which suggests an ok fit.

b) We fit an AR(2) model.

```
ar(x=arma1, order.max = 2, method=c("yw"))
```

```
##
## Call:
## ar(x = arma1, order.max = 2, method = c("yw"))
##
## Coefficients:
##       1        2
##  0.7465  -0.2877
##
## Order selected 2  sigma^2 estimated as  1.37
```

The Yule-Walker estimates of the parameters are $\hat{\alpha_1} = 0.86$ and $\hat{\alpha_2} = -0.38$ which are very different from that of (a). The Portmanteau test returns a p-value of 0.42 which suggests that this is a good fit but since the estimated parameters from this method are more dissimilar with the true values compared to the ML-estimated ones so the quality of these estimates are not as good as that of a).

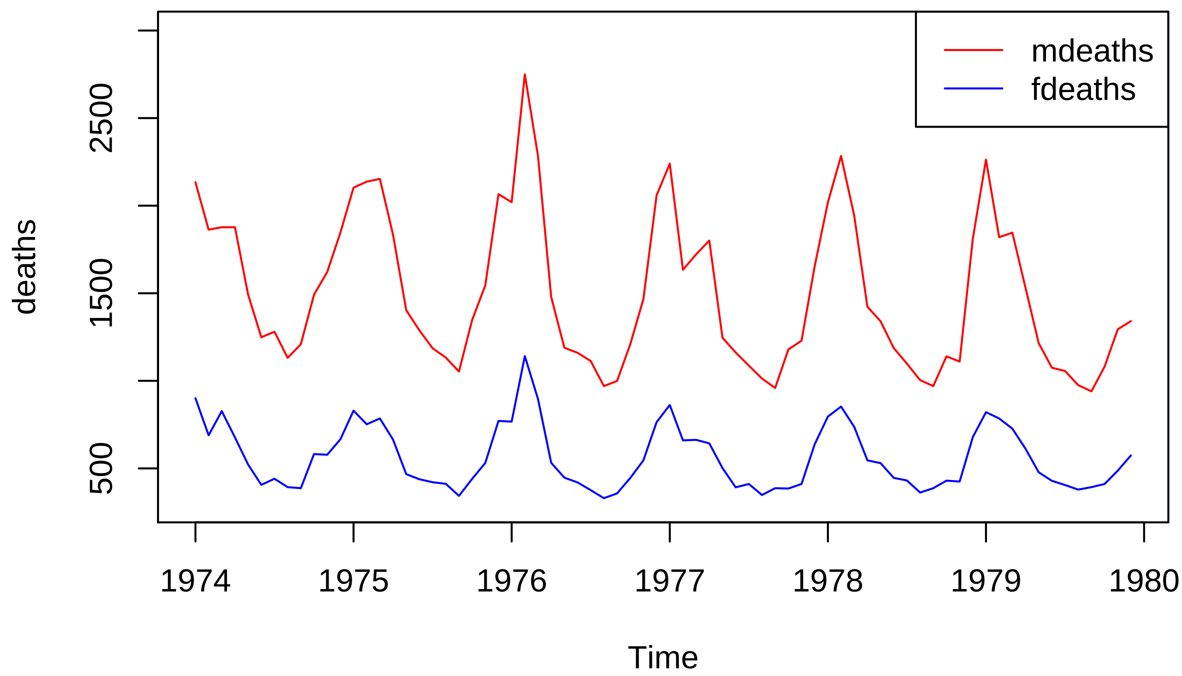
```r
Box.test(ar(x=arma1, order.max = 2)$resid,type="Box-Pierce")$p.value
```

```
## [1] 0.1316694
```

Question 5

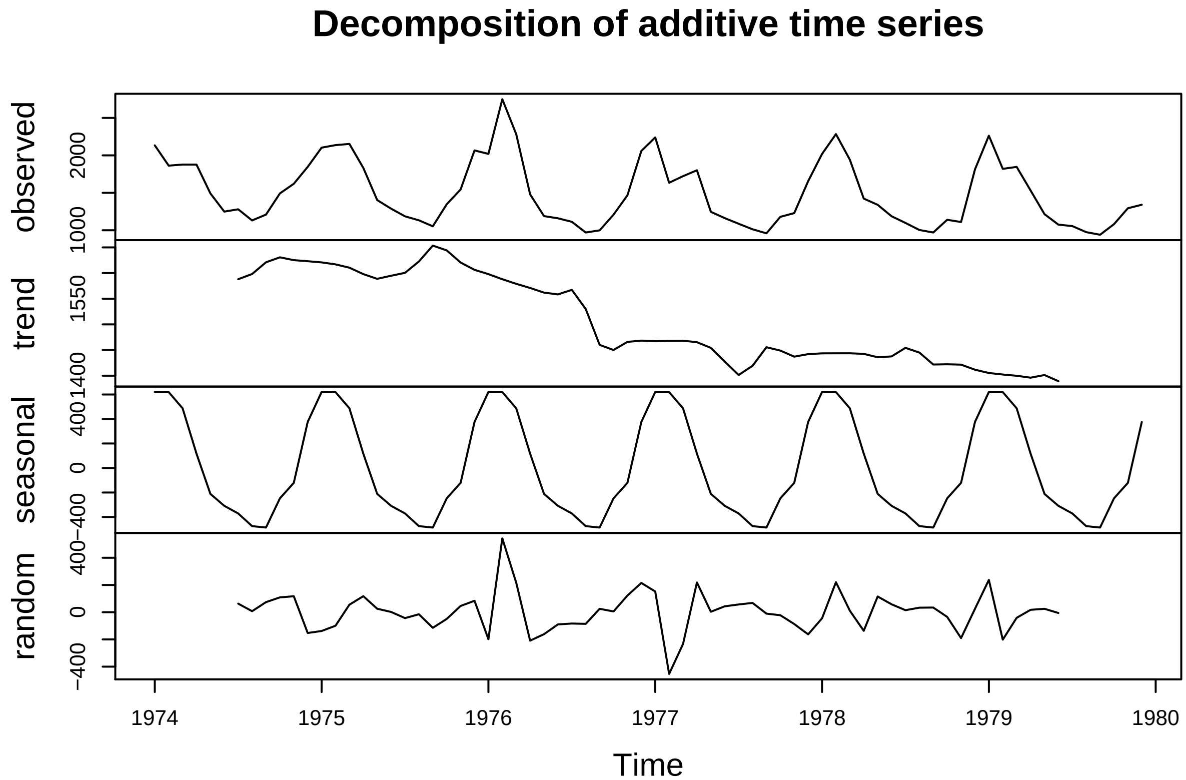
a) The plot of *mdeaths* and *fdeaths* is

```r
plot.ts(mdeaths, type='l', col="red", ylim=c(300, 3000), ylab="deaths")
lines(fdeaths, type="l", col="blue")
legend("topright", legend=c("mdeaths", "fdeaths"),
       col=c("red", "blue"),
       lty=1)
```



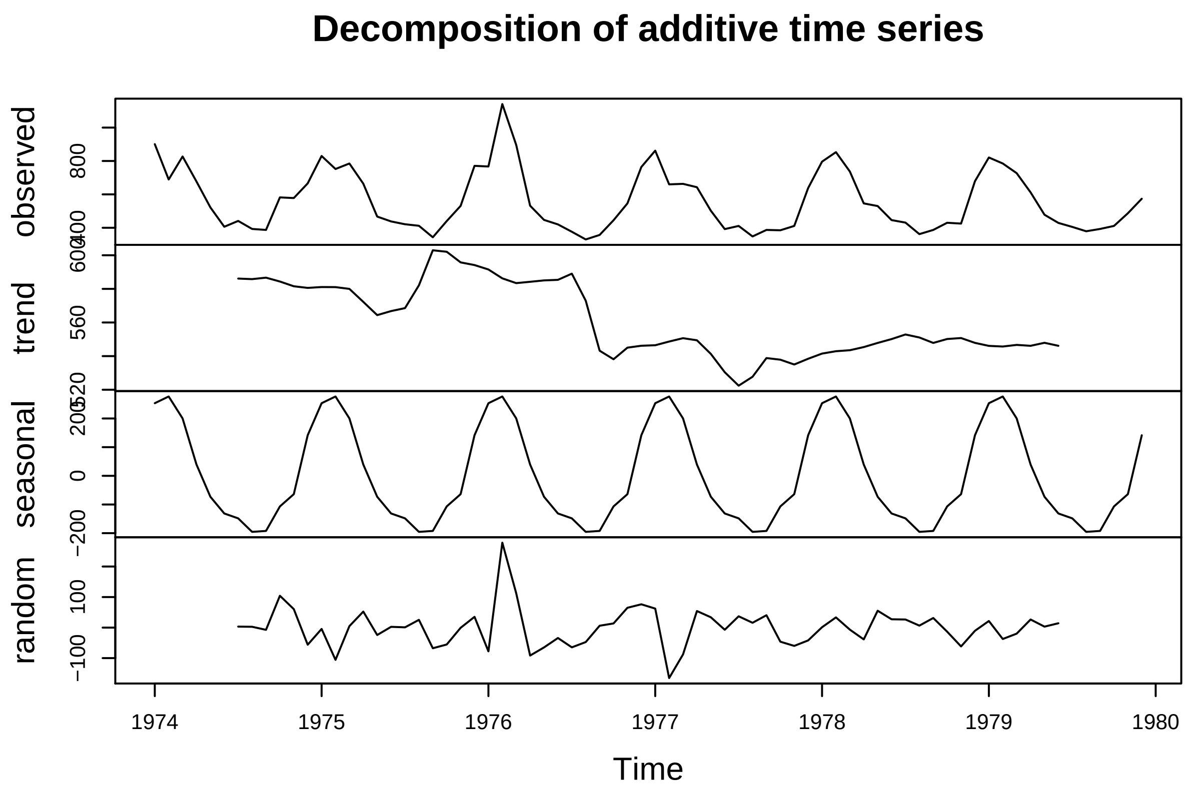
We can also decompose the time series into its components. The decomposition for *mdeaths* is:

```r
decompm=decompose(mdeaths);
plot(decompm)
```



And the decomposition of $fdeaths$ is

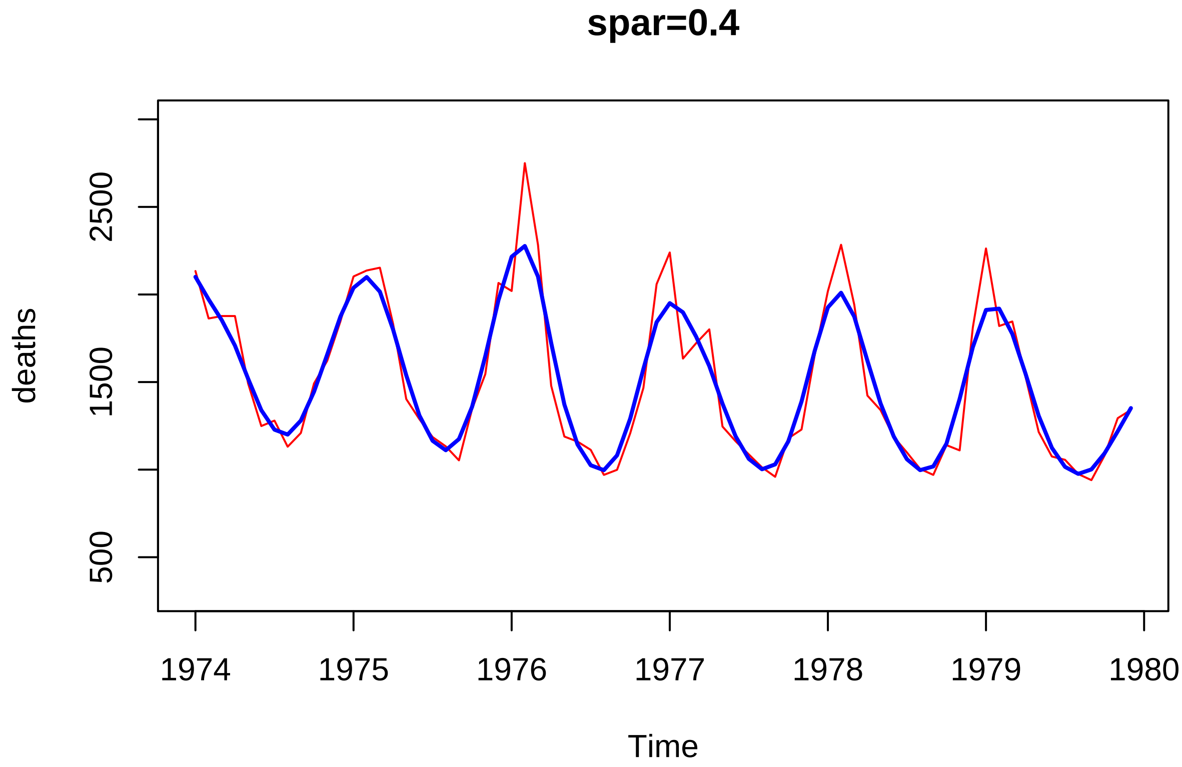
```
decompf=decompose(fdeaths);
plot(decompf)
```



It can be seen that there exists seasonality in the two datasets since the number of deaths of both genders spike every year at the beginning of the year. One obvious difference between the two datasets is that the *mdeaths* is far higher than *fdeaths* (male oppression). Furthermore, for both datasets, there seems to be a faint downward trend of the number of deaths but this trend is stronger in *mdeaths* than *fdeaths*.

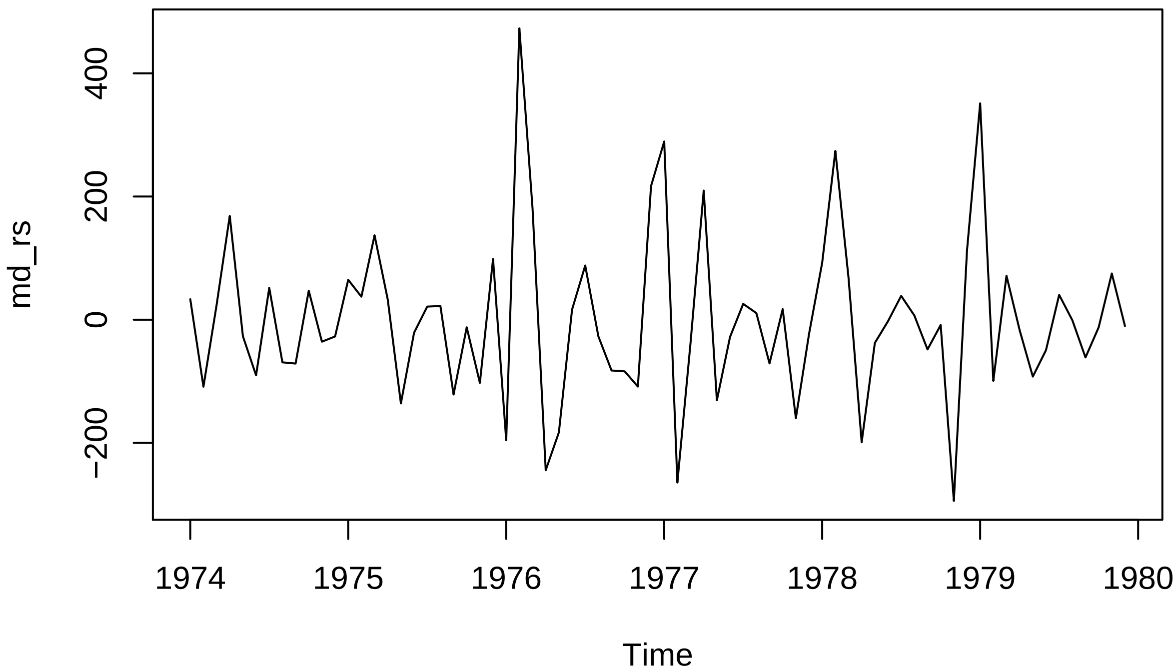
b) For this part, we only consider *mdeaths*, the plot of the time series *mdeaths* with its smoothing spline is

```
plot.ts(mdeaths, type='l', col="red", ylim=c(300, 3000),
        ylab="deaths", main="spar=0.4")
sm = smooth.spline(x=mdeaths, spar=0.4)
lines(sm, col="blue", lwd=2)
```



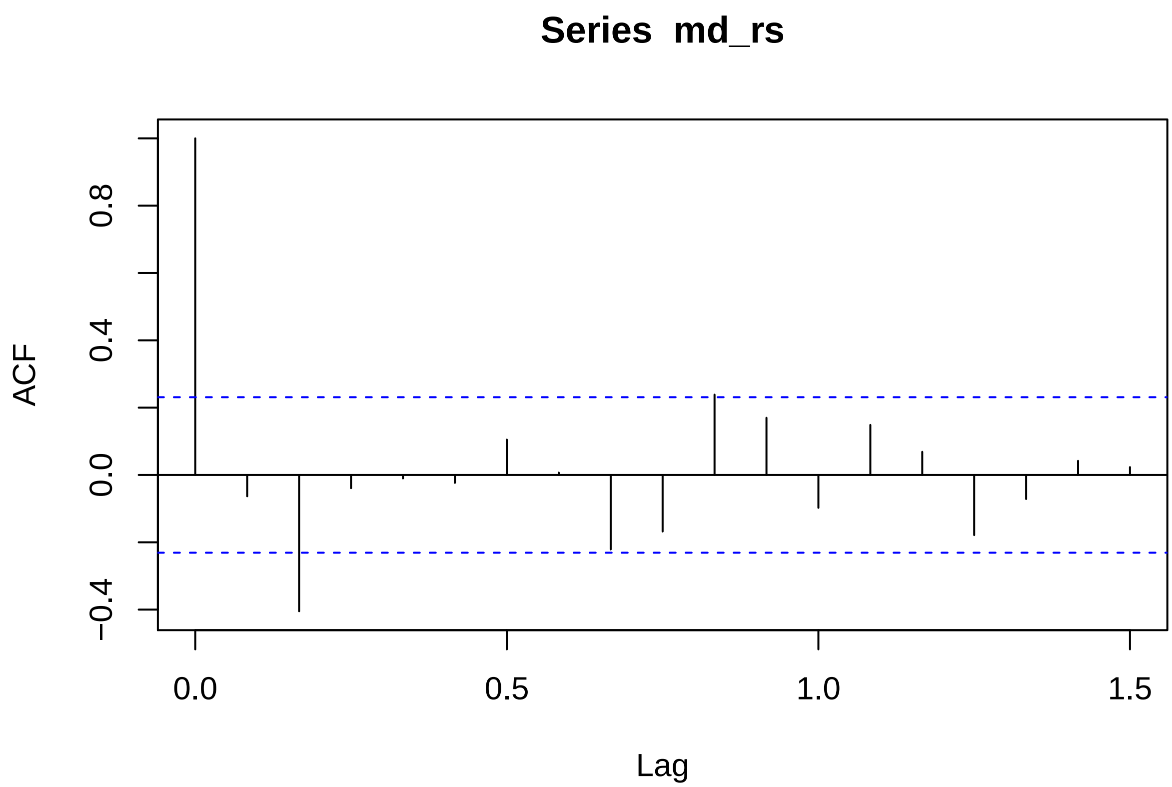
We then subtract the spline from the original time series to obtain a time series with no trend or seasonality.

```
md_rs = mdeaths - sm$y;
plot.ts(md_rs)
```



We then plot an estimate of the autocorrelation function on the de-trended and de-seasonalized time series.

```
acf(md_rs)
```



Since the ACF is mostly small, this confirms the fact that we are dealing with a de-trended and de-seasonalized time series. For $p \in \{0, \ldots, 5\}$ and $q \in \{0, \ldots, 5\}$, we fit different ARMA(p, q) models to the de-trended and de-seasonalized time series and obtain the best one. Different combinations of p and q and their corresponding ARMA(p, q) model's AIC are listed below.

```
for (p in c(1:5)) {
  for (q in c(1:5)) {

    print(c(p, q, arima(md_rs,order=c(p,0,q),method="ML",include.mean=F)$aic))
  }
}
```

```
## [1]   1.0000   1.0000 889.9842
## [1]   1.0000   2.0000 883.3127
## [1]   1.0000   3.0000 885.3632
## [1]   1.0000   4.0000 873.0483
## [1]   1.0000   5.0000 874.8177
## [1]   2.000   1.000 882.351
## [1]   2.0000   2.0000 883.9786
## [1]   2.0000   3.0000 871.3379
## [1]   2.0000   4.0000 856.2432
## [1]   2.0000   5.0000 877.0378
## [1]   3.0000   1.0000 884.3122
## [1]   3.0000   2.0000 885.6045
## [1]   3.0000   3.0000 872.1019
## [1]   3.0000   4.0000 856.4304
## [1]   3.0000   5.0000 860.2427
```

```
## [1]   4.0000   1.0000 882.9752
## [1]   4.0000   2.0000 866.0423
```

```
## Warning in log(s2): NaNs produced
```

```
## [1]   4.0000   3.0000 848.0102
## [1]   4.0000   4.0000 874.8433
## [1]   4.0000   5.0000 842.2854
## [1]   5.0000   1.0000 884.3229
## [1]   5.0000   2.0000 866.2353
## [1]   5.0000   3.0000 869.9696
## [1]   5.0000   4.0000 888.0419
```

```
## Warning in arima(md_rs, order = c(p, 0, q), method = "ML", include.mean = F):
## possible convergence problem: optim gave code = 1
```

```
## [1]   5.0000   5.0000 875.6489
```

It can be seen that the (p, q) value with the minimum AIC score is $(4, 5)$ with an AIC score of 842.

```
mdeaths_c = mdeaths[1:60];
year = index(mdeaths)[1:60];
yearsq=year^2;
```

d)