Bocconi University - 20236 Time Series Analysis.

Exercises with R

Sonia Petrone

Introduction to time series analysis with R.

You can find many **tutorials** on the web, such as these:

https://www.youtube.com/watch?v=EmZqlcKkJMM ,

https://www.youtube.com/watch?v=orjLGFmx6l4

on R:

https://www.youtube.com/watch?v=s3FozVfd7q4//

on R markdown: https://www.youtube.com/watch?v=DNS7i2m4sB0

on ggplot2: https://www.youtube.com/watch?v=49fADBfcDD4

and more..

Time series decomposition. Exponential smoothing.

These exercises are meant to introduce you to **R** and **Rmarkdown** (see **LAB1 on BBoard**) and time series analysis with **R**.

1. **Exercise 1** (*This exercise is solved in LAB 1, which is posted on BBoard*)

Consider the time series **co2**, monthly observations of atmospheric concentrations of co2 at Mauna Loa, from 1959 to 1997; this series is available in R.

```
? co2
?decompose
```

- Plot the data.
- The series clearly shows a trend and a seasonal behavior. Describe such components using classical time series decomposition methods (**R** function: **decompose**).

2. **Exercise 2**.

Consider another classical data set, **UKgas**, available in **R**. This time series provides the quarterly UK gas consumption from the first quarter of 1960 to the forth quarter of 1986, in millions of therms.

```
? UKgas
```

- First, plot the data.
- Describe the structural components of the UKgas time series: decompose the series, and plot the results.

Would you use an additive or multiplicative time series decomposition? Or, you could take the log of the series: why?

```
decUK = decompose(log(UKgas), type="additive")
```

- Comment briefly: are you satisfied with your results? Is there anything that you would like to improve, or any assumption you would like to relax?
- Compare the previous results with those obtained by using Holt & Winters exponential smoothing with trend and seasonality. Comment briefly.

HINT: we still see a time-varying seasonal component, that changes around 1973. On the contrary, classical time series decomposition assumes ***constant*** *seasonal factors. In fact, we would like to allow for* ***time varying*** *seasonal factors. Holt & Winters exponential smoothing is a first step in that direction. We will see more with Dynamic Linear Models (DLMs).*

3. **Exercise 3**.

The dataset **Nile**, available in **R**, provides the measurements of the annual flow of the river Nile at Ashwan, for the period 1871-1970.

```
? Nile
```

This is a very popular dataset, used as a classical example of a time series that shows a change point.

- Plot the data, together with the online one-step-ahead forecasts obtained by simple exponential smoothing. Use the R default choice of the smoothing parameter α (what value of α is chosen)?
- Then, compare the results obtained for different choices of the smoothing parameter α, namely: $\alpha = 0.1$, $\alpha = 0.9$, and the default choice of α. Comment on the effect of α on the forecasts.
- Compare the results above, using the MAPE as a simple measure of the predictive performance of the model. Comment briefly.

```
# hint:
mape = function(y, yhat){ mean(abs(y - yhat)/y) }
```

4. **Coronavirus data**

With the COVID-19 pandemic, we experienced the importance of reliable data and forecasts forsupporting decisions under uncertainty.

I provide on BBoard a dataset of daily data on the evolution of the COVID-19 epidemic in Italy, since February 20, 2020.

- Import the data in RStudio and provide meaningful plots.
- Make some first analysis of these data, with the tools of exploratory time series analysis leaned so far - e.g.: what is the trend of contagion? Can you make one step ahead forecasts? Can you formally express the uncertainty about your forecast? (the answe to the latter question is no).

Gaussian Processes with R

These exercises are aimed at giving you more familiarity with the notions of stochastic processes; mean function and autocovariance function of a stochastic process; and stationarity (see the lecture notes on

"Introduction to probabilistic models for time series analysis", posted on BBoard).

In classic time series analysis, a time series is a *discrete time* stochastic process (Y_t) where the inxed t refers to time. But more generally, it can be in continuous time; in fact, here we will work with a continuous time Gaussian process. Thus, this is also an exercise about Gaussian Processes with *R*. We use the package mvtnorm.

Simulating Gaussian processes

For this exercise you have to simulate trajectories from various types of Gaussian Processes with different mean function $\mu(t)$ and autocovariance functions $\gamma(t, t')$. Plotting the sampled trajectories will help you explore and understand the behavior of the process.

Constant mean function, independent

Simulate a Gaussian process with mean function $\mu(t) = 10$ and auto-covariance function

$$\begin{aligned} \gamma(t, t') &= 3 \qquad \text{if } t = t' \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

(*You will generate y_t for t in a grid in $(1, T)$; thus, you generate samples from the (appropriate) finite-dimensional Gaussian distribution.*)

- Is this process stationary? Why?
- Plot 5 trajectories until $t = 100$
- Compute the estimate of the mean function and overlap it to the plot above.
- What happens for varying σ^2?
- Now, consider a single trajectory (y_t), and pretend it is the observed sample. Estimate the autocorrelation function (ACF) and plot it (i.e., plot the correlogram) for lags $h = 1, \ldots, 10$. Compare the estimated ACF with the true one and comment briefly.

Constant mean function, squared-exponential kernel.

Simulate a Gaussian process with constant mean function $\mu(t) = 10$ and autocovariance function

$$\gamma(t, t') = \sigma^2 \exp\left(-\phi\,(t - t')^2\right)$$

This is called squared-exponential; or a squared-exponential kernel. This form of the autocovariance function expresses the idea that r.v.'s Y_t and $Y_{t'}$ that are closer in time (the distance $(t - t')^2$) is small) are more strongly correlated. The parameter ϕ (also often expressed as $1/2\ell$) plays the role of a *decay* parameter. Simulating trajectories for different values of the decay parameter and/or the variance helps to get more aware of their effect.

- Plot 5 trajectories until $t = 100$ for $\sigma^2 = 3$ and your choice of ϕ.
- Do that for different choices of σ^2 and ϕ: the point you want to illustrate is the effect of ϕ and σ^2. Include the mean function in your plot.
- Now, consider a single trajectory and pretend it is the observed sample. Would it make sense to estimate the autocorrelation function (ACF) as in the exercise above? Why?
- How could you use, instead, replicates of the sample (here, simulated trajectories; use $N = 1000$ trajectories) to estimate the ACF?

Time-varying mean function, squared-exponential kernel.

Repeat the first point of the exercise above. This time however use the following mean function:

$$\mu(t) = \begin{cases} t & 0 \leq t \leq 6\pi \\ 6\pi - 10 \cdot \sin(t/6) & 6\pi < t \leq 100 \end{cases}$$

Constant mean function, Matern kernel.

Simulate a Gaussian process with mean function $\mu(t) = 10$ and autocovariance function:

$$k(t, t') = \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{2\nu}{\ell} |t' - t| \right)^{\nu} K_{\nu} \left(\frac{2\nu}{\ell} |t' - t| \right),$$

where $\Gamma(\nu)$ is the Gamma function and K_ν is the Modified Bessel function (`besselk`).

Repeat the first step of the exercise above.

ARIMA models

This exercise is taken from the book *Forecasting: Principles and Practice* (3rd Edition), by Rob J Hyndman and George Athanasopoulos \ `https://otexts.com/fpp2/` \ (Exercise 7, Chapter 8, Sect 8.11)

You would use **R**packages for time series forecasting associated with the book (install and load them).

Consider **wmurders**, the number of women murdered each year (per 100,000 standard population) in the United States.

1. By studying appropriate graphs of the series in R, find an appropriate ARIMA(p, d, q) model for these data. Should you include a constant in the model? Explain.
2. Fit the model using R and examine the residuals. Is the model satisfactory?
3. Forecast three times ahead. Create a plot of the series with forecasts and prediction intervals for the next three periods shown.
4. Does auto.arima() give the same model you have chosen? If not, which model do you think is better? (and, what would chatGPT suggest, and WHY?)

Markov chains and inference for Markov chains

These are analytic exercises. In **R**, you might simulate larger data sets and repeat the exercises on inference for Markov chains.

Basic notions on Markov chains.

1. Consider a time series $(Y_t)_{t\geq 0}$ with $Y_t \in \{1, 2, \ldots, K\}$.

   - When is $(Y_t)_{t\geq 0}$ a Markov chain?

   - Provide an example.

2. Consider a homogeneous Markov chain $(Y_t)_{t\geq 0}$, with a finite state-space $\{1, 2, \ldots, K\}$.

   - What is the initial distribution? And what is the transition matrix? What would change for non-homogeneous Markov chains?

   - Suppose that $K = 3$ and write the expression of $P(Y_2 = 1 \mid Y_1 = 2, Y_0 = 2)$. - Then write the expression of
   $$P(Y_0 = 2, Y_1 = 2, Y_3 = 1, Y_4 = 2, Y_5 = 2, Y_6 = 2, Y_7 = 1, Y_8 = 3).$$

3. Using the DAG (directed acyclic graph) representation of the dependence structure of a Markov chain, show that Y_t is conditionally independent of $(Y_1, \ldots, Y_{t-3})$, given Y_{t-2}.

4. (*from past written proof*)

   Consider a Markov chain $(Y_t)_{t\geq 0}$ with state space $\mathcal{Y} = \{1, 2, 3\}$, initial value $Y_0 = 1$ and transition matrix

$t-1 \setminus t$	1	2	3
1	.6	.4	0
2	0	.7	.3
3	.1	.1	.8

   What is the probability that $Y_2 = 2$? And what is the probability that $Y_1 = 1$, given that $Y_2 = 2$ (and $Y_0 = 1$)?

5. (*past written proof*)

   Suppose that $(Y_t)_{t\geq 0}$ is a Markov chain with transition matrix

t-1 \t	1	2	3
1	0.6	0.4	0
2	0.3	0.6	0.1
3	0	0.2	0.8

   What is the state-space of this Markov chain?

   What is the marginal probability distribution of Y_2, given the starting value $Y_0 = 1$?

6. Consider a Markov chain $(Y_t)_{t\geq 0}$ with state space $\mathcal{Y} = \{1, 2, 3\}$, initial value $Y_0 = 1$ and transition matrix

$t-1 \setminus t$	1	2	3
1	.6	.4	0
2	0	.7	.3
3	.1	.2	.7

   Is this Markov chain irriducible? Motivate your answer.

7. *(past written proof).*

   Let $(Y_t)_{t\geq 0}$ be a Markov chain with finite state-space $\mathcal{Y}$. Discuss the limit behavior of the probability distribution $P(Y_n = \cdot \mid Y_0 = j)$ of Y_n given the initial state $Y_0 = j$.

   What can you say about the limit behavior of the marginal distribution of Y_n?

Inference for Markov chains.

1. (*past written proof*)

   In a survey, a random sample of $n = 100$ new graduates is quarterly monitored, along one year after graduation, recording, for each individual, her job status, coded as $Y = 1$: full time job; $Y = 2$: temporary job; $Y = 3$: unemployed. The graduation day, $t = 0$, is the same for all individuals in the survey.

   For individual i, the data are regarded as a sample from the categorical time series $(Y_{i,t})_{t\geq 0}$, where $Y_{i,t}$ denotes his job status at time t. Across individuals, the $(Y_{i,t})$, $i = 1, \ldots, n$, are modeled as independent and identically distributed *homogeneous* Markov chains.

   (a) Suppose the observed matrices of transition counts are

$t=0 \setminus t=1$	1	2	3	
1	10	0	5	15
2	5	20	5	30
3	10	15	30	55

$t=1 \setminus t=2$	1	2	3	
1	20	5	0	25
2	10	20	5	35
3	10	10	20	40

$t=2 \setminus t=3$	1	2	3	
1	35	0	5	40
2	15	10	10	35
3	0	10	15	25

$t=3 \setminus t=4$	1	2	3	
1	40	5	5	50
2	10	5	5	20
3	5	20	5	30

   (a) Compute the maximum likelihood estimate (MLE) of the transition probability $p_{3,1} = P(Y_t = 1 \mid Y_{t-1} = 3)$ (i.e., of a transition from "unemployed" to "full time job").

   (b) Provide the asymptotic confidence interval of level 0.95 for $p_{3,1}$ (*The* 0.9 *quantile of a standard Gaussian distribution is* 1.28*, the* 0.95 *quantile is* 1.65*, the* 0.975 *quantile is* 1.96).

   (c) Let us now relax the assumption that the Markov chain is homogeneous. Consider the transition matrix $\mathbf{P}_t \equiv [p_{i,j}(t)]$, where $p_{i,j}(t)$ is the probability of a transition from state i at time $t-1$, to state j at time t. Compute the MLE of $p_{3,1}(t)$, for $t = 1, 2$.

2. (*past written proof*)

   A social survey includes periodic interviews on a panel of $n = 100$ individuals; each individual is asked to express his/her opinion on a political group (1 ="in favor"; 2= "negative"; 3= "uncertain"). The interviews are taken monthly, from March ($t = 0$) until July ($t = 4$). The observed transition counts are reported in the following table.

$t-1 \setminus t$	1	2	3	
1	60	20	40	120
2	20	80	20	120
3	20	120	20	160

   Let us model the individual series $(Y_{i,t})$, $i = 1, \ldots, 100$ as independent and identically distributed homogeneous Markov chains, with common transition matrix $\mathbf{P} = [p_{i,j}]$. The statistical task is to estimate $\mathbf{P}$ based on the data.

(a) For individual 1, the data are $(1, 1, 3, 3, 2)$. Write the expression of the joint probability $P(Y_{1,1} = 1, Y_{1,2} = 3, Y_{1,3} = 3, Y_{1,4} = 2 \mid Y_{1,0} = 1; \mathbf{P})$.

(b) Write the expression of the likelihood of the $p_{i,j}$'s (consider the initial values $y_{i,0}$ as fixed). Then obtain the expression of the maximum likelihood estimate of $p_{i,j}$.

(c) What is the estimated probability that an individual who is in favour in May will turn negative in June? Provide the MLE, together with the asymptotic confidence interval of level 90% (*The* 0.9 *quantile of a standard Gaussian distribution is* 1.28*, the* 0.95 *quantile is* 1.65*, the* 0.975 *quantile is* 1.96).

Hidden Markov Models with R

Part (a) of this exercise is about Hidden Markov Models (HMMs) with **R**. We use the **R** package *depmixS4*. **Look at LAB 2, posted on BBoard**.

Part (b) suggests an extension to a covariate-dependent transition matrix in the HMM. This extension has not been covered in class, but is implemented in the package depmixS4; you might want to go into it.

Exercise.

The dataset provided in the file data_assHMM.cvs (posted on BBoard) provides monthly data including 10 years Italian government bond's interest rate, inflation represented by the Harmonised Index of Consumer Prices (HICP) and default ratings assigned by the agencies Moody's and Fitch, in the investment grade range, i.e. from Aaa/AAA to Baa3-/BBB-. The data set collects data for the period January 1997 to July 2019, and it has been built mainly using OECD data.

- **(a)** Let us focus on the nominal interest rate for the 10 years Italian government bond. In fact, you may want to consider the *real* interest rate, calculated from the HICP.

Plot the data and comment *briefly* if and why a HMM could be a reasonable model.

Let us indeed use a Hidden Markov Model, with 3 states (representing, say, boom (i.e. less risky, lower interest rates), recession (high risk, high interest rates) and a stable path), and Gaussian emission ditributions, with state-dependent mean and variance.

Provide the MLEs of the unknown paramters of the model (and their standard errors). Comment *briefly*.

Find the optimal state sequence ("deconding") and plot it, comparing it with the data.

- **(b) – (optional)**. HMMs are particularly useful for time series that present change points. However, one may want to go further, trying to improve *prediction* of a possible change point through avialble covariates. To this aim, one may use non-homogeneous HMMs, allowing the transition matrix to depend on covariates. A reference is Zucchini, W., MacDonald, I-L. and Langrock, R. (2016) *Hidden Markov Models for Time Series: an introduction using R*. Chapman and Hall/CRC;and the **R** package *depmixS4* allows this extension, see Visser and Speekenbrink (2010), *Journal of Statistical Software* – both references are posted on BBoard.

You may want to explore this more general class of HMMs for the data under study here.