ST5209/X Assignment 3

Due 17 Mar, 11.59pm

Set up

1. Make sure you have the following installed on your system: $\LaTeX$, R4.2.2+, RStudio 2023.12+, and Quarto 1.3.450+.
2. Pull changes from the course repo.
3. Create a separate folder in the root directory of the repo, label it with your name, e.g. `yanshuo-assignments`
4. Copy the assignment1.qmd file over to this directory.
5. Modify the duplicated document with your solutions, writing all R code as code chunks.
6. When running code, make sure your working directory is set to be the folder with your assignment .qmd file, e.g. `yanshuo-assignments`. This is to ensure that all file paths are valid.[1]

Submission

1. Render the document to get a .pdf printout.
2. Submit both the .qmd and .pdf files to Canvas.

1. Holt-Winters, residuals, and forecast accuracy

Consider the antidiabetic drug sales time series which can be loaded using the following code snippet.

```
diabetes <- read_rds("../_data/cleaned/diabetes.rds") |>
  select(TotalC)
```

a. Fit the following exponential smoothing models on the entire time series:

    - Holt-Winters with multiplicative noise and seasonality,

[1] You may view and set the working directory using `getwd()` and `setwd()`.

- Holt-Winters with a log transformation, with additive noise and seasonality,
- Holt-Winters with multiplicative noise and seasonality, and damping.

b. Make ACF plots for the innovation residuals of these three models. Does this suggest that the residuals are drawn from a white noise model?
c. Calculate the p-value from a Ljung-Box test on the residuals with lag $h = 8$. Does this suggest that the residuals are drawn from a white noise model? What does this mean about the fitted model?
d. Perform time series cross-validation for the three methods, using `.init = 50` and `.step = 10`, and with the forecast horizon $h = 4$. Which method has the best RMSSE? How many data points is the error averaged over in total?

2. ETS parameters, prediction intervals

The dataset `hh_budget` contains annual indicators of household budgets for a few countries.

a. Fit `ETS` on the time series comprising savings as a portion of household budgets in Canada. Do not specify what type of ETS model to fit, instead allowing the function to perform automatic model selection.
b. Which ETS model was selected?
c. What are the fitted parameters of the model?
d. Based on the fitted parameters, what simple forecasting method is this similar to?
e. Based on your answer to d), how does the width of the prediction interval change as a function of the forecast horizon h? Make a plot to verify this.

3. Moving averages and differences

Consider the linear trend model

$$X_t = \beta_0 + \beta_1 t + W_t.$$

Define a time series (Y_t) by taking a moving average of (X_t) with a symmetric window of size 7. Define another times series (Z_t) by taking a difference of (X_t).

a. What is the mean function for (Y_t)? What is the ACVF for (Y_t)?
b. What is the mean function for (Z_t)? What is its ACVF?
c. What is the CCF of (Y_t) and (Z_t)?
d. Are (Y_t) and (Z_t) jointly stationary?

4. Sample vs population ACF

Consider the signal plus noise model

$$X_t = \sin(2\pi t/5) + W_t.$$

a. What is the ACF of (X_t)?
b. Simulate a time series $X_1, X_2, \dots, X_{200}$ from this model and plot its sample ACF.
c. Why does the sample ACF not look like the population ACF function?
d. Why does the asymptotic normality theorem for the ACF not apply?

5. Gaussian processes

Consider a random vector (X_1, X_2, X_3, X_4) that has the joint Gaussian distribution

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \sim N\left(\begin{bmatrix} -0.5 \\ 0 \\ 0.5 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 2 \\ 1 & 1 & 2 & 3 \end{bmatrix} \right)$$

a. What is the marginal variance of X_4?
b. What is the conditional variance of X_4, conditioned on the observations $X_1 = 1, X_2 = 1, X_3 = 1$. Does it depend on these particular values? *Hint: The following code snippet creates matrix in R. You may use `solve()` to find the inverse of a matrix.*
c. What is the conditional mean of X_4, conditioned on the observations $X_1 = 1, X_2 = 1, X_3 = 1$?
d. Write a level 95% prediction interval for X_4 given these observations.

```r
A <- matrix(c(3, 1, 1, 1,
              1, 3, 1, 1,
              1, 1, 3, 2,
              1, 1, 2, 3),
            nrow = 4, ncol = 4, byrow = TRUE)
```