Room 3: Garden

STAT 464/864 ~ **Fall 2024**

Discrete Time Series Analysis | *Skye P. Griffith* ~ Queen's University

Deadline for Original Problems: Friday, November 1st (midnight)

Galaxy Grading (Modified from *Specifications Grading,* Linda Nilson, 2014)

You need 4 Stars to complete this Room. Choose 4 Original Stars to submit.
The other 2 will be your Alternate Stars, which you may complete later to boost your score.
Alternate Stars may only replace Original Stars from the same "Room."

Graduate students must also complete 1 **comet.** Choose 1 Original Comet to submit.
The other will be your Alternate Comet. The same rules apply.

Formatting

Your submission **must** be a rendered Quarto PDF document, otherwise it will not be graded.

On your computer, make a folder called `464_Room3` (or `864_Room3`, obviously.)
Create a Quarto file called `464_Room3.qmd` file in that folder.
Store all components of this Room in that folder, including datasets and image files.

- **R-code** must be typed and executed in "chunks".
- **Commentary** must be typed in-document.
- **Mathematical Proofs** may be typed, or hand-written and inserted into your Quarto document. The syntax to do so is: `![Caption](image.png){width=100%}`
- **Use headers** (hashtags) to organize your work. See lines 22-34 of W1_solutions.qmd.

Learning Outcomes

1. Forecast time-series data based on ARMA models, using R.
2. Find optimal (minimum MSE) linear predictors, given various time series models, particularly in the context of h-step-ahead prediction.
3. Determine the MSE (mean squared error) of such a linear predictor.
4. Understand the difference between forecasting based on finitely-many versus infinitely-many past values.

The Toronto Heat Super-Galaxy (Stars 1 & 2)

Data	Hourly Toronto Temperature
Times Sampled	May 1, 2003 (00:00) – May 7, 2003 (23:00)

Data:

Download the dataset Toronto_Room3.csv from Github, and save it where? That Room 3 folder you just made. Make sure the data and the .qmd file are both in there. While you're at it, make sure the .png files for your handwritten Stars are in there too. Now LOAD the data!

```
toronto <- read.csv("Toronto_Room3.csv")
```

Plotting Etiquette

1. Line-plots. Not just points.
2. Create scientifically meaningful labels.
3. Use colour to distinguish different lines on the same plot.

Star 1) *Forecasting*

1. From the data, create a time series consisting of the hourly Toronto temperature. Truncate this series to the dates `2003-05-01 (00:00) - 2003-05-07 (23:00)`.
2. Use the `forecast()` function from the `itsmr` to predict the temperature for the next 24 hours. That is, from `2003-05-08 (00:00)` to `2003-05-08 (23:00)`. Base your prediction on an ARMA(1,1) model. The model parameters should:
   - account for a seasonal component with a period of 24 hours
   - account for a quadratic trend

Star 2) *Plotting + Commentary*

Create a single plot of the following:

1. The time series you created in part 1 of Star 1.
2. The predicted values you obtained from part 2 of Star 1).
3. The upper and lower confidence bands for the predicted values as dashed line plots. (I recommend using `lty = 2`)
4. The actual temperature data from `2003-05-08 (00:00)` to `2003-05-08 (23:00)` as a dashed line plot.

Commentary: How well did you think you predicted May 8th's temperature?

Star 3: Harmony Forecast Galaxy

Suppose you have two uncorrelated, zero-mean RVs with unit variance: A_1 and B_1. Let ω be a fixed frequency in the interval $[0, \pi]$. Then define

$$X_t = A_1 \cos(\omega t) + A_2 \sin(\omega t) \qquad t = 0, \pm 1, ...$$

Find the following:

1. $P_2 X_3$ and its mean squared error.
2. $\tilde{P}_n X_{n+1}$ and its mean squared error.

Star 4: Puzzle Piece Galaxy

Let X_1, X_2, X_4, X_5 be taken from the MA(1) model

$$X_t = Z_t + \theta Z_{t-1}$$
$$\{Z_t\} \sim wn(0, \sigma^2)$$

Find the following:

2. $P(X_3|X_4, X_5)$: Find the best linear predictor of X_3 in terms of X_4 and X_5.
3. $P(X_3|X_4, X_5)$: Find the best linear predictor of X_3 in terms of X_1, X_2, X_4, and X_5.
4. Compute the MSE for each of these estimates.

The ARMA(1,1) Super-Galaxy (Stars 5 and 6)

Let $\{Y_t\}$ be the ARMA(1,1) process:

$$Y_t - \frac{1}{2}Y_{t-1} = W_t + \frac{1}{2}W_{t-1}$$
$$W_t \sim wn(0, \sigma^2)$$

Use the formula for the autocovariance of an ARMA(1,1) process (see p.78 text, or the lecture slides, with ϕ, θ and σ^2 as defined in this problem) to complete the following. Give actual numbers for the coefficients of the linear predictions, *and* for the mean squared errors.

Star 5)

1. Find $P(X_2 \mid X_1)$: the best linear predictor of X_2 in terms of X_1.
2. Compute the MSE of $P(X_2 \mid X_1)$.

Star 6)

1. Find $P(X_n \mid X_1, X_2)$, the best linear predictor of X_n in terms of X_1 and X_2, for $n > 3$.
2. Compute the MSE of $P(X_n \mid X_1, X_2)$.
3. Compute the limit of this mean squared error as $n \to \infty$.

Comet 1: Galaxy of the infinite past

Let $\{X_t\}$ be a stationary series satisfying $X_t = Z_t - \theta Z_{t-1},\ t \in \mathbb{Z}$, where Z_t is white noise with variance σ^2. Assume θ lies strictly within the unit disk

1. Show that the best linear predictor $\tilde{P}_n X_{n+1}$ of X_{n+1}, based on $\{X_j\}_{j=-\infty}^{n}$ is

$$\tilde{P}_n X_{n+1} = -\sum_{j=1}^{\infty} \theta^j X_{n+1-j}$$

2. What is the MSE of $\tilde{P}_n X_{n+1}$?

Comet 2: Galaxy of the finite past

Let $\{X_t\}$ be the time series $X_t = Z_t - Z_{t-1}$, where $\{Z_t\}$ is white noise with variance σ^2.

1. Find the best linear predictor $P_n X_{n+1}$ of X_{n+1} in terms of $X_1, \dots, X_n$.
2. What is the corresponding MSE?