Exam - stochastic processes

CO2 concentration in atmosphere

R's built-in dataset `co2` is a time series of atmospheric CO2 concentration at the Mauna Loa observatory. We will analyse this dataset below.

Explorative data analysis

- Start as always by plotting the data.

- Is this a second order stationary time series (explain what it means)?

- Use `decompose` to make a decomposition of the data to remove any trend and seasonal component and explain the method of how this is done.

- Save the random component as `co2rand` (omit any `NA` values), plot the correlogram of `co2rand` and explain the correlogram: What is it used for and how is it interpreted? What is assumed about the underlying process?

Auto-regressive model of order 1

- Manually fit a AR(1) model using `lm` without an intercept (**note:** you need to change the chunk option `eval` in the Rmd-file to `TRUE` for this to be executed, but the chunk will only work if have defined `co2rand` as instructed previously):

```
n <- length(co2rand)
y <- co2rand[2:n]
x <- co2rand[1:(n-1)]
fitlm <- lm(y ~ x - 1)
summary(fitlm)
```

- Explain the output of the last command above. Is there significant autocorrelation? What is the estimated lag 1 autocorrelation coefficient?

- Write down the equation expressing the fitted model.

- What is the theoretical acf for this model? Comment on this in relation to the correlogram you plotted.

- Based on the data $x_1, \ldots, x_n$ what is the predicted value for x_{n+1} (do not use the `predict()` function, but calculate it manually)?

- What is the predicted value for x_{n+10} (do not use the `predict()` function, but calculate it manually)?

- Save the model residuals (use `residuals(fitlm)`) and plot the correlogram.

- Is the AR(1) model a good fit?

Higher order autoregressive moving average (ARMA) models

- How do we define higher order AR(p) processes?

- Use `lm` as above to estimate an AR(2) model for `co2rand`. Is the lag 2 autocorrelation coefficient significant according to a `summary` of the fitted model?

- What is the partial autocorrelation function and how is it useful in relation to AR(p) processes?

- How is a MA(q) process defined?

- Try to fit a collection of ARMA(p,q) models for p and q at most 2, and find the best fitting one based on AIC. (Hint: If you called the models `ar1`, `ar2`, `ma1`, `ma2`, `arma11`, `arma12`, `arma21`, and `arma22` you can compare them all in a single call to `AIC`; AIC(ar1, ar2, ma1, ma2, arma11, arma12, arma21, arma22))

- Write down the parameter estimates of the final model, and check whether it is a good fit to the data.

- Give 95% confidence intervals for the parameter estimates of the final model selected by AIC. (Hint: Use `confint`.)

Prediction

- Make a prediction with an approximate 95% prediction interval for the next value of the random component x_{n+1} based on this model.

- If there was no autocorrelation in the random component the approximate 95% prediction interval is given by $\bar{x} \pm 2\sqrt{s^2(1 + 1/n)}$. Calculate this prediction interval and compare it with the one obtained above. What is the difference? Try to explain this.