Project I, Course on Time Series Analysis 2021, StatØ2

Deadline **1 October 2021**

(1) Simulate a sample of size $n = 1000$ from the model $X_t = 0.9X_{t-1} + Z_t$ with (Z_t) iid student t-distributed with 10 degrees of freedom.

(a) Plot the sample ACF for the first 20 lags with the program acf in R. The plot shows 95% asymptotic confidence bands (uniform over the lags) based on the central limit theorem for an iid Gaussian sequence. Calculate lag-wise 95% confidence bands based on 1000 random permutations of the data and show them in the same plot. Compare the two confidence bands.

> Random permutations do not change the distribution of an iid sample. If the data were iid the generated confidence bands would be correct. This method is quick and clean for iid data and does not use any asymptotic theory.

(b) Calculate the Yule-Walker estimators $\widehat{\phi}$ and $\widehat{\sigma}^2$. Determine an asymptotic 95% confidence band for $\widehat{\phi}$ based on the central limit theorem for $\widehat{\phi}$.

(c) The bootstrap is a resampling method for constructing confidence bands, originally for iid data; see the simulation lecture notes. On p. 27 in these notes, a method is described how the bootstrap can be used for an AR(1) process to construct non-asymptotic confidence bands for $\widehat{\phi}$.

The idea is to use the residuals $\widehat{Z}_t = X_t - \widehat{\phi}X_{t-1}$, $t = 1, \ldots, n$ (take $X_0 = 0$). Assume that they are "roughly iid".

(i) Draw a sample of n integers $\pi_1, \ldots, \pi_n$ from $1, \ldots, n$ with equal probabilities $1/n$. Define

$$Z_1^* = \widehat{Z}_{\pi_1}, \ldots, Z_n^* = \widehat{Z}_{\pi_n} .$$

(ii) Define a new sample $X_t^* = \widehat{\phi}X_{t-1}^* + Z_t^*$, $t = 1, \ldots, n$.

(iii) Calculate the Yule-Walker estimator $\widehat{\phi}^*$ from this sample. Repeat steps (i)–(iii) $B = 1000$ times. Then you get the sequence of Yule-Walker estimators $\widehat{\phi}^*(i)$, $i = 1, \ldots, B$.

Show a box-plot of this sequence of estimators.

Calculate the 97.5% and 2.5% quantiles of this sequence of estimators. They constitute a confidence band for $\widehat{\phi}$. Compare with the asymptotic confidence band from (b).

(2) Go to Yahoo Finance and download the S&P 500 index, closing prices, from January 1, 2000, until today. Assume that a business year has 250 days. Plot the annual sample means and variances of the log-returns and of their absolute values. Are these estimates in agreement with the assumption of an ergodic time series?

(3) Calculate the return series (Y_t) and log-return series $(\log(1 + Y_t))$ for the S&P500 closing prices.

(a) Plot the differences $|Y_t - \log(1 + Y_t)|$ and determine their maximum.

(b) Calculate the sample autocorrelations of the resulting log-return time series, their absolute values and squares. Choose the maximum number of lags as 10% of the sample size (function acf in R).

(c) Fit an AR model to the log-returns (function ar.mle or ar.yw); choose its order by using the proposed AIC number.

(d) Simulate from the chosen AR model with iid student noise with 4 degrees of freedom (function arima.sim in R), plot a realization with the same sample size as the log-return series and plot the sample autocorrelations of the simulated series, their absolute values and squares for the same number of lags as for the log-return series. Would you recommend to choose an AR model for this log-return series?

(4) Consider Bartlett's formula for a causal AR(1) process.

(a) Calculate the asymptotic variance w_{hh} of $\rho_{n.X}(h)$ as well as the limit of $\lim_{h\to\infty} w_{hh}$ as $h \to \infty$.

(b) Simulate a sample of size $n = 200$ from the AR(1) process $X_t = 0.8X_{t-1} + Z_t$ for iid standard normal white noise (Z_t) (use arima.sim in R). Draw the sample autocorrelation function for (X_t) with maximal lag $h = 25$ and asymptotic confidence bands for iid white noise (this is standard in the function acf in R). Then draw in the same graph the 95% asymptotic confidence bands based on the calculations of (a), i.e., $\rho_{n,X}(h) \pm 1.96\sqrt{w_{hh}/n}$.

> The asymptotic confidence bands in the ACF plot provided by software correspond to iid white noise. If the sample ACF at a given lag h is outside these bands this is an indication of dependence between X_t and X_{t+h}. The confidence bands given by Bartlett's formula follow the estimated autocorrelations; they are more informative than the bands for iid white noise.

(5) The Wölfer sunspot number series is a standard time series which is available in R (sunspots). First transform the monthly data to annual data by taking annual averages. In this way, one avoids seasonality of the time series.

(a) Calculate the AIC for the sunspot numbers for an AR(p) model, $p = 1, \ldots, 20$, (functions ar.yw or ar.mle in R) and plot the AIC in a graph against p.

(b) Fit an AR(p) model for p that minimizes the AIC (functions ar.yw or ar.mle). Simulate a time series from this AR(p) model (arima.sim) with iid noise and the same sample size as the sunspot numbers. Plot the sunspot numbers and the simulated time series for comparison. Experiment with the distribution of the noise in order to get the size of the data right; by default arima.sim chooses iid standard normal noise.

Possible choices for the noise distribution are 1. take a random permutation of the residuals of the AR(p) process or 2. find (fit?) a suitable distribution for the residuals.

(c) Plot the sample autocorrelation functions of the sunspot numbers and the simulated time series and compare them.