Exercises and Lecture Notes
STK 4060, Spring 2022

Version 0.05, 31-Jan-2022

Nils Lid Hjort

Department of Mathematics, University of Oslo

Abstract

These are Exercises and Lecture Notes for the course on time series, modelling and analysis, STK 4060 (Master level) or STK 9060 (PhD level), for the spring semester 2022. The collection will grow during the course.

1. The variance or an average of correlated variables

A classical and crucial result from traditional statistics is that if $x_1, \ldots, x_n$ are independent with the same distribution, then $\mathrm{Var}\,\bar{x}_n = \sigma^2/n$, for the data average $\bar{x} = (1/n)\sum_{i=1}^n x_i$, where σ^2 is the variance of a single observation. This is rather different for models with dependence. Suppose now that $x_1, \ldots, x_n$ is a stationary sequence, with $\mathrm{cov}(x_k, x_{i+h}) = \sigma^2\rho(|h|)$, for some correlation function $\rho(h) = \mathrm{corr}(x_i, x_{i+h})$.

(a) Show that

$$\mathrm{Var}\,\bar{x}_n = \frac{\sigma^2}{n}\Big\{1 + 2\sum_{h=1}^n (1 - h/n)\rho(h)\Big\} = \frac{\sigma^2}{n}\sum_{h=-n}^{n}(1 - |h|/n)\rho(j).$$

(b) For the special case of $\rho(h) = \rho^h$, called autocorrelation of order 1, show that

$$\mathrm{Var}\,\bar{x}_n = \frac{\sigma^2}{n}\Big\{1 + 2\sum_{h=1}^{n-1}\rho^h - (1/n)\sum_{h=1}^{n-1} h\rho^h\Big\} = \frac{\sigma^2}{n}\Big\{\frac{1+\rho}{1-\rho} + O(1/n)\Big\}.$$

With a positive autocorrelation, therefore, the variance of $\bar{x}_n$ becomes clearly bigger than under independence.

(c) Suppose you observe such a stationary time series $x_1, \ldots, x_n$, with autocorrelation function $\rho(h) = \rho^h$ for $h = 1, 2, 3, \ldots$, and with unknown mean μ, variance σ^2, and autocorrelation parameter $\rho \in (-1, 1)$. If you do the traditional $\bar{x}_n \pm 1.96\, s_n/\sqrt{n}$ interval for μ, recommended in 99 statistics books, with s_n the empirical standard deviation, what will be its confidence coverage level?

(d) Give estimators for μ, σ, ρ, constructed from the observed time series.

(e) Give a more careful and appropriate 95 percent confidence interval, taking autocorrelation into account. Note in particular that such a confidence interval *is wider* than the traditional one, when the autocorrelation is positive.

2. An autoregressive time series model

Construct a time series $x_1, x_2, \ldots, x_n$ as follows, via i.i.d. $\varepsilon_1, \ldots, \varepsilon_n$ being standard normal. Let $x_1 = \varepsilon_1$ and then $x_{t+1} = \rho x_t + \varepsilon_{t+1}$ for $i = t, 2, \ldots$, where ρ is a value inside $(-1, 1)$.

(a) Take $n = 100$ and $\rho = 0.345$, and simulate such a time series in your computer. Check what the `acf(xdata)` does, playing also a bit with other combinations of n and ρ.

(b) Write $\mathcal{F}_t$ for all observed history up to and including time point t. Show that $\mathrm{E}\,(x_{t+1}\,|\,\mathcal{F}_t) = \rho x_t$ and $\mathrm{Var}\,(x_{t+1}\,|\,\mathcal{F}_t) = 1$. Deduce also from this that

$$\mathrm{E}\,x_t = \rho\,\mathrm{E}\,x_{t-1} \quad \text{and} \quad \mathrm{Var}\,x_t = 1 + \rho^2\,\mathrm{Var}\,x_{t-1}.$$

Show that $\mathrm{E}\,x_t = 0$, for all t, and find a formula for the variance of x_t.

(c) Starting from

$$\begin{aligned} x_2 &= \rho\varepsilon_1 + \varepsilon_2, \\ x_3 &= \rho^2\varepsilon_1 + \rho\varepsilon_2 + \varepsilon_3, \\ x_4 &= \rho^3\varepsilon_1 + \rho^2\varepsilon_2 + \rho\varepsilon_3 + \varepsilon_4, \end{aligned}$$

find a general formula for x_t, expressed in terms of the i.i.d. components $\varepsilon_1, \ldots, \varepsilon_t$. Use this to find and explicit distribution of x_t. Also show

$$\mathrm{Var}\,X_t = 1 + \rho^2 + \rho^4 + \cdots + \rho^{2(t-1)} = \frac{1 - \rho^{2t-1}}{1 - \rho^2},$$

re-proving what you found in (b).

(d) Find the explicit covariance and correlation between x_i and x_{i-1}.

(e) When the time series has been at work for some time, show that

$$\mathrm{Var}\,x_i \to \frac{1}{1-\rho^2}, \quad \mathrm{cov}(x_i, x_{i+1}) \to \frac{\rho}{1-\rho^2}, \quad \mathrm{cov}(x_i, x_{i+2}) \to \frac{\rho^2}{1-\rho^2},$$

etc.

(f) Show that the real acf (the autocorrelation function) becomes $1, \rho, \rho^2, \rho^3, \ldots$,

(g) Simulate a few time series using the above construction, with a few combinations of n and ρ. Verify that with n moderate-to-large, the empirical `acf(xdata)` becomes close to the real $1, \rho, \rho^2, \rho^3, \ldots$.

3. Using regression modelling for the Johnson & Johnson dataset

Consider the dataset called `jj` in the `astsa` package, giving the quarterly earnings of the J & J company, from quarter 1 1960 to quarter 4 1980. One wishes to study how these $y_1, \ldots, y_n$ evolve over time (with $n = 84$ quarters over 21 years), e.g. to predict earnings for the coming year. The task here is to go through some regression models, so to speak before factoring in correlations and specific time series aspects.

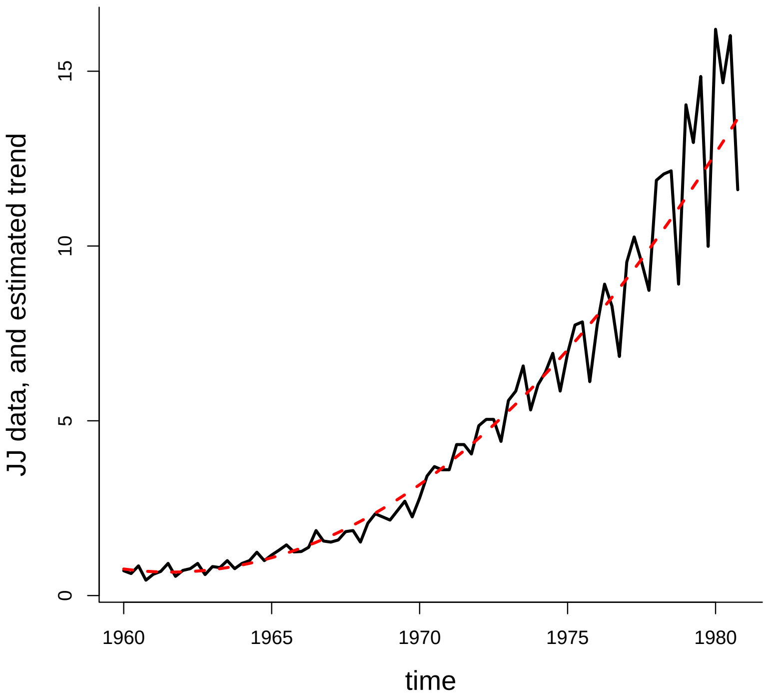

Figure 0.1: The JJ data, with estimated trend, from the five-parameter model.

(a) Write $x_t = t - 1960$, for $t = 1, \ldots, n$. Fit the rather simple classic linear regression model, with $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$, with the ε_t taken i.i.d. $\mathrm{N}(0, \sigma^2)$. Look at the fitted trend $\widehat{m}_1(t) = \widehat{\beta}_0 + \widehat{\beta}_1 x_t$, alongside data, to check that this model is far too simple. For the practice, check also the residuals $r_{1,t} = y_t - \widehat{m}_1(t)$; these will vary too much, indicating again that this model is too coarse.

(b) A rather better model is to include a quadratic term for the trend. Fit the regression model $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$, again with the ε_t taken i.i.d. from the $\mathrm{N}(0, \sigma^2)$. Plot the estimated trend $\widehat{m}_2(t) = \widehat{\beta}_0 + \widehat{\beta}_1 x_t + \widehat{\beta}_2 x_t^2$ alongside data, examine the residuals $r_{2,t} = y_t - \widehat{m}_2(t)$, and comment on what you find.

(c) You learn from the above that the trend function is adequately described by such a parabola, but that that variance of data is not constant; it increases over time. So try the variance heteroscedastic model

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \sigma_t \varepsilon_t, \quad \text{for } t = 1, \ldots, n, \quad \text{with } \sigma_t = \exp\{\gamma_0 + \gamma_1 (x_t - \bar{x})\},$$

and with the ε_t now being i.i.d. and standard normal. The model has three parameters for the mean and two for the variance. Show that the log-likelihood function for this five-parameter model can be expressed as

$$\ell(\theta) = \sum_{t=1}^{n} \{-\log \sigma_t - \tfrac{1}{2}(y_t - \beta_0 - \beta_1 x_t - \beta_2 x_t^2)/\sigma_t^2 - \tfrac{1}{2}\log(2\pi)\},$$

in terms of the full parameter vector θ.

(d) Find the maximum likelihood (ML) estimates, say $\widehat\theta_{\rm ml}$, by numerically maximising the log-likelihood function. Compute also approximate standard errors, for the five parameter estimates, via the general normal approximation theorem for parametric models,

$$\widehat\theta_{\rm ml} \approx_d {\rm N}_p(\theta, \widehat\Sigma), \quad \text{with } \widehat\Sigma = \widehat J^{-1}. \tag{0.1}$$

Here $\widehat J = -\partial^2 \ell(\widehat\theta_{\rm ml})/\partial\theta\partial\theta^{\rm t}$, the Hesse matrix of second order derivatives, computed at the the ML position. Using `nlm` in `R` you get the Hesse matrix for free, along with the numerical optimisation, using something like

$$\texttt{hello} = \texttt{nlm}(\texttt{minuslogL}, \texttt{starthere}, \texttt{hessian} = \texttt{T})$$

followed, pretty generically and very usefully, by

```
ML = hello$estimate
Jhat = hello$hessian
se = sqrt(diag(solve(Jhat)))
showme = cbind(ML,se,ML/se)
print(round(showme,4))
```

(e) Produce a version of Figure 0.1.

(f) Once you have the basic code up and running it is relatively easy to try out other variations of such models. Try to put in a cyclic term, perhaps $\beta_4 \cos(2\pi t/4)$, and again look at both the residuals and the acf.

4. Understanding the empirical acf, under independence

Suppose $x_1, x_2, \ldots$ are really independent, with mean zero and variance one. What happens then, with the `acd(xdata)`? Below, write $x_{a,b}$ for the average of values $x_a, \ldots, x_b$.

(a) Consider first $A_n = (1/n)\sum_{t=1}^{n-1} x_t x_{t+1}$. Show that A_n has mean zero and variance $(n-1)/n^2$, i.e. approximately $1/n$.

(b) Then go to the proper empirical $B_n = (1/n)\sum_{t=1}^{n-1}(x_t - \bar x_{1,n})(x_{t+1} - \bar x_{1,n})$. Show that

$$B_n = A_n - \frac{n-1}{n}\bar x_{1,n}\bar x_{1,n-1} - \frac{n-1}{n}\bar x_{1,n}\bar x_{2,n} + \frac{n-1}{n}\bar x_{1,n}^2 \doteq A_n - \bar x_{1,n}^2,$$

with $\doteq$ meaning 'good approximation, not affecting limits when n grows'.

(c) Show that B_n, like the simpler A_n, has mean zero and variance approximately equal to $1/n$. Show then that $A_n \to_{\rm pr} 0$, $B_n \to_{\rm pr} 0$, with '$\to_{\rm pr}$' denoting convergence in probability: $\Pr(|B_n| \ge \varepsilon) \to 0$ for each small ε.

(d) Since A_n is a sum of variables with the same distribution, with mean zero, and ${\rm Var}\, A_n \doteq 1/n$, it is natural to expect limiting normality, i.e. $\sqrt{n}A_n \to_d {\rm N}(0,1)$. This does *not* follow from the traditional CLTs (central limit theorems), since x_1x_2 is not independent of x_2x_3, etc. Check with the book's Appendix A.2, however, concerning CLTs for m-dependent variables, and verify that indeed $\sqrt{n}A_n \to_d 1$.

(e) From $\sqrt{n}B_n \doteq \sqrt{n}A_n - \sqrt{n}\bar{x}_{1,n}^2$, show that also $\sqrt{n}B_n \to_d \mathrm{N}(0,1)$, i.e. the same limit distribution.

(f) Now go from 1-step to 2-step, and work through the details for $A_n = (1/n)\sum_{t=1}^{n-2} x_t x_{t+2}$ and

$$B_n = \widehat{\gamma}(2) = (1/n)\sum_{t=1}^{n-2}(x_t - \bar{x}_{1,n})(x_{t+2} - \bar{x}_{1,n}).$$

The main things are that $\widehat{\gamma}(2) \to_{\mathrm{pr}} 0$, the true value of $\gamma(2)$ under independence, and that $\sqrt{n}\widehat{\gamma}(2) \to_d \mathrm{N}(0,1)$.

(g) Generalise properly to the result $\sqrt{n}\widehat{\gamma}(h) \to_d \mathrm{N}(0,1)$, for

$$\widehat{\gamma}(h) = (1/n)\sum_{t=1}^{n-h}(x_t - \bar{x}_{1,n})(x_{t+h} - \bar{x}_{1,n}).$$

(h) So far we've assumed variance $\sigma^2 = 1$, for simplicity of presentation and argumentation. For the general case, show that for a sequence of independent variables, with some mean μ and variance σ^2, we have $\sqrt{n}\widehat{\gamma}(h) \to_d \mathrm{N}(0,\sigma^4)$. Finally show that for

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)} = (1/n)\sum_{t=1}^{n-h}\frac{(x_t - \bar{x}_{1,n})}{\widehat{\sigma}}\frac{(x_{t+h} - \bar{x}_{1,n})}{\widehat{\sigma}} = \frac{\sum_{t=1}^{n-h}(x_t - \bar{x}_{1,n})(x_{t+h} - \bar{x}_{1,n})}{\sum_{t=1}^{n}(x_t - \bar{x}_{1,n})^2},$$

our good friend the acf, we do have the clarifying easy good result $\sqrt{n}\widehat{\rho}(h) \to_d \mathrm{N}(0,1)$.

(i) For such a sequence of i.i.d. variables, show that when one computes the empirical acf, then

$$\Pr\{\widehat{\rho}(h) \in [-1.96/\sqrt{n}, 1.96/\sqrt{n}]\} \to 0.95,$$

for each lag h. This is the reason for the 'magical band' $\pm 1.96/\sqrt{n}$ provided in the standard use of `acf`.

5. A simple moving average process

Suppose $w_0, w_{\pm 1}, w_{\pm 2}, \ldots$ are i.i.d., with finite variance σ^2. Then consider the process

$$x_t = aw_{t-1} + (1-2a)w_t + aw_{t+1},$$

with a a tuning parameter. We call this a moving average process, with window length 3.

(a) Compute the variance of x_t, and also the covariance function $\gamma(h)$ and autocorrelation function $\rho(h)$. Plot the acf for a few values of a, including the equal balance case of $a = 1/3$.

(b) Then do a similar analysis for a 5-window moving average process, of the type

$$x_t = aw_{t-2} + aw_{t-1} + (1-4a)w_t + aw_{t+1} + aw_{t+2}.$$

Again, plot the acf for a few values of a, including the balanced case of $a = 1/5$.

(c) Similarly consider the case of

$$x_t = \rho^2 w_{t-2} + \rho w_{t-1} + w_t + \rho w_{t+1} + \rho^2 w_{t+2}.$$

Find the acf, and plot it, for a few values of ρ.

6. A general stationary normal time series model

Suppose $x_1, \ldots, x_n$ is a stationary normal time series, which means that the full vector has a multinormal distribution; this is also equivalent to saying that all linear combinations arfe normal. Assume it has mean μ, varianec σ^2, and correlation function $\rho(h) = \mathrm{corr}(x_t, x_{t+h})$.

(a) Show that the joint distrisbution of the full series is a $\mathrm{N}_n(\mu\mathbf{1}, \sigma^2 A)$, where $\mathbf{1} = (1, \ldots, 1)^{\mathrm{t}}$ is the vector of 1s, and A the $n \times n$ matrix of $\rho(s-t)$, for $s, t = 1, \ldots, n$; in particular, the diagonal elements are all 1.

(b) Using the basic definition of the multinormal joint density, show that the log-likelihood function can be written

$$\ell(\theta) = -n \log \sigma - \tfrac{1}{2} \log |A| - \tfrac{1}{2}(y - \mu\mathbf{1})^{\mathrm{t}} A^{-1} (y - \mu\mathbf{1})/\sigma^2 - \tfrac{1}{2} n \log(2\pi),$$

wiith θ the parameters involved. If the correlation function is known, then A is known, and θ comprises only μ, σ. For such a case, show that the ML estimators become

$$\widehat{\mu} = \frac{\mathbf{1}^{\mathrm{t}} A^{-1} y}{\mathbf{1}^{\mathrm{t}} A^{-1} \mathbf{1}} \quad \text{and} \quad \widehat{\sigma}^2 = \frac{Q_0}{n}, \quad \text{with } Q_0 = (y - \widehat{\mu}\mathbf{1})^{\mathrm{t}} A^{-1} (y - \widehat{\mu}\mathbf{1}).$$

Check that this leads to familiar formulae in the case of i.i.d. observations, where $A = I_n$, the identity matrix.

(c) If there is a parameter, say λ, in the correlation function, however, we need also $A = A(\lambda)$, and we have

$$\ell(\mu, \sigma, \lambda) = -n \log \sigma + \tfrac{1}{2} \log |A(\lambda)| - \tfrac{1}{2}(y - \mu\mathbf{1})^{\mathrm{t}} A(\lambda)^{-1} (y - \mu\mathbf{1})/\sigma^2 - \tfrac{1}{2} n \log(2\pi),$$

(d) Take e.g. $n = 100$, generate $x_1, \ldots, x_n$ from the standard normal in your computer, and fit the three-parameter model which has unknown μ, σ, λ, where the correlation function is modelled as $\rho(h) = \exp(-\lambda h) = \rho^h$, i.e. with $\rho = \exp(-\lambda)$ the 1-step correlation. Repeat the experiment a few times, to see how well the ML estimators succeed in coming close to the true values.

References

Shumway, R.H. and Stoffer, D.S. (2016). *Time Series Analysis and Its Applications* (4th ed.). Springer, Berlin.