Avd. Matematisk statistik

KTH Matematik

EXAM FOR SF2943/SF2945 TIME SERIES ANALYSIS/TIDSSERIEANALYS
THURSDAY 31 MAY 2012, 14.00–19.00 HRS.

Examiner: Tobias Rydén, tel. 790 8469

Allowed aids: Formulas and survey, Time series analysis (without notes!). Pocket calculator.

Notation introduced should be defined and explained. Solutions, arguments and calculations must be clear and motivated well enough to make them easy to follow.

Each correct solution counts for 10 points. Pass (grade E) requires 25 points. Students who obtain 23 or 24 points will be offered the option to do an additional small exam to possibly raise their grade to E. Students wanting to take this option must contact the examiner within a week after the results from the exam have been published.

Solutions in Swedish are of course welcome!

The exam will be marked no later than 11 June, and the results will be available through *Mina sidor*.

Problem 1

When recording a stationary time series $\{X_t\}$, due to faulty measurement equipment each data point is with some (small) probability p erroneously substituted by 0. This happens independently for each time-point t.

Show that the actually recorded data series, $\{Y_t\}$ say, is a stationary time series and express its autocovariance function in terms of that of $\{X_t\}$. (10 p)

Problem 2

In an application studied at the company TIMESYS, the data being observed is well described by the stationary time series model

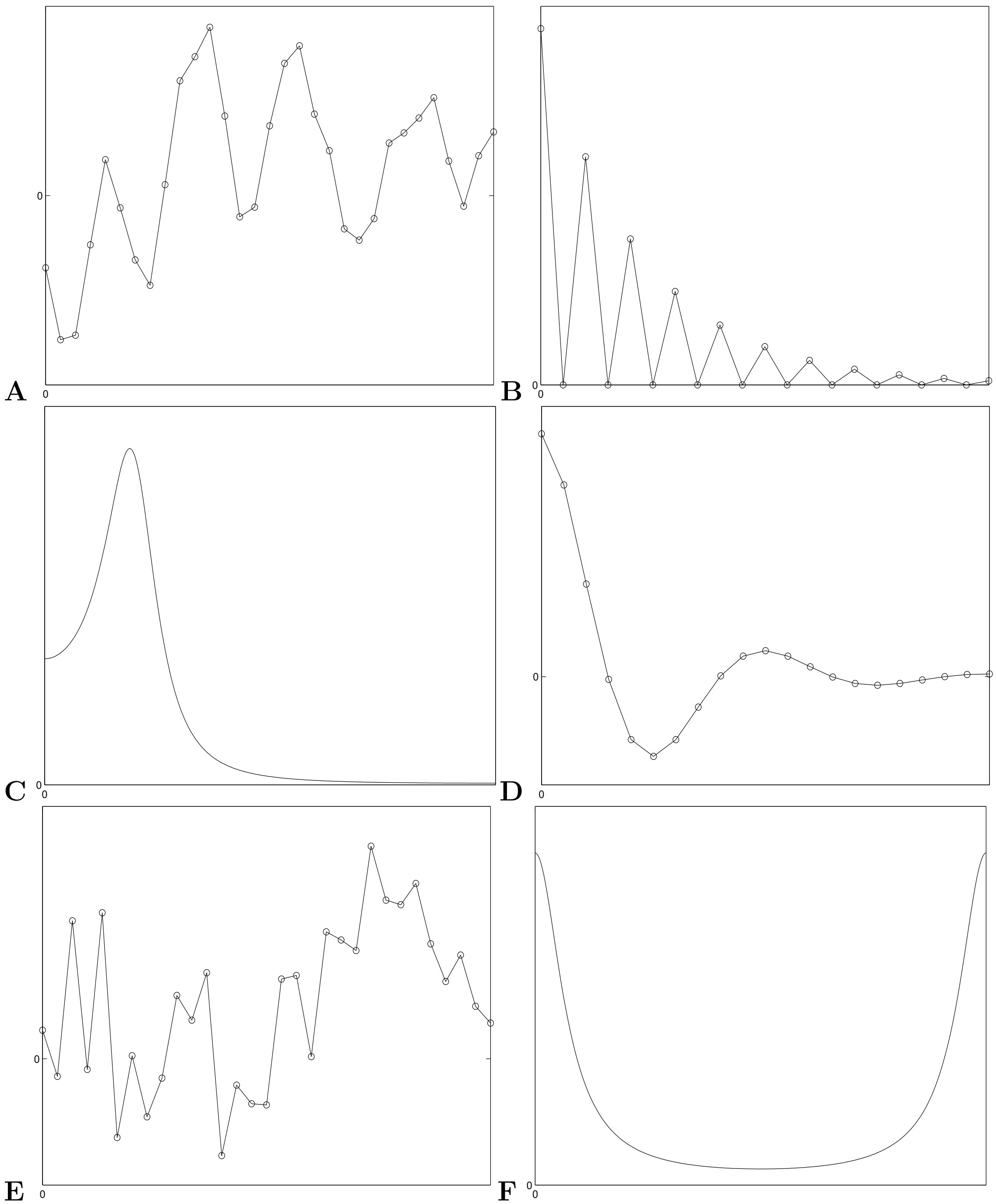
$$X_t = Z_t - 0.4Z_{t-1} + 0.6Z_{t-2},$$

where $\{Z_t\}$ is zero mean white noise with variance 3. In this particular application one wishes to predict, at any time t, the sum of the two following observations, i.e. $X_{t+1} + X_{t+2}$.

TIMESYS employee Albert argues that 'since any X_s with $s < t-1$ is uncorrelated with X_{t+1} and X_{t+2}, these variables will not be of any use for the prediction.' Then he asks for the best (in the MSE sense) linear predictor of the above sum, using X_t and X_{t-1} only.

Solve the above problem, i.e. compute the best linear predictor of $X_{t+1} + X_{t+2}$ using X_t and X_{t-1}. Compute the MSE of the predictor as well. (10 p)

Problem 3



The six plots A–F above are a relisation, the autocorrelation function (ACF), and the spectral density (SD) for each of two different AR(2) processes.

Which plots are realisations, ACFs, and SDs, respectively? Group the plots into two triplets, each containing a realisation, an ACF and an SD, such that the plots in each triplet correspond to the same time series. Motivate your results. (10 p)

Problem 4

Using a sample of size 1,000 from a stationary time series, the following estimates the of autocovariance and partial autocorrelation functions were computed:

h	0	1	2	3	4	5	6	7	8
$\widehat{\gamma}(h)$	0.596	–0.276	–0.024	0.011	0.027	–0.008	–0.044	0.025	–0.031
$\widehat{\text{PACF}}(h)$	1	–0.463	–0.325	–0.224	–0.098	–0.045	–0.122	0.009	–0.090

(a) Based on the above estimates, choose between the model classes AR(p) and MA(q) for this data, and suggest a suitable model order p or q. Motivate your choices properly. (4 p)

(b) Estimate the model parameters $\phi_1, \dots, \phi_p$ or $\theta_1, \dots, \theta_q$, as well as the innovation variance σ^2, using the above information. (6 p)

Problem 5

(Continuation of Problem 2). Albert's colleague Beth argues that Albert is wrong in claiming that data X_s with $s < t - 1$ are useless for the prediction, and she says that a predictor that takes more historical data into account would improve the MSE. Albert says 'Well, maybe, but the improvement won't be large.'

Address this dispute by computing the mean squared error (MSE) of the best (in the MSE sense) linear predictor of $X_{t+1} + X_{t+2}$, using (potentially) all X_s with $s \leq t$. (10 p)

Note: Of course it is unrealistic that data for all $s \leq t$ is available, since it is an infinite amount, but you may view this assumption as a simplifying approximation.

Good luck!

SOLUTIONS TO EXAM FOR SF2943/SF2945 TIME SERIES ANALYSIS/TIDSSERIE-ANALYS, THURSDAY 31 MAY 2012

Formulas and survey will be referred to as FS.

Problem 1

Let $\mu_X = E(X_t)$. Write $I_t = 1$ if Y_t is recorded correctly, i.e. $Y_t = X_t$, and $I_t = 0$ otherwise, i.e. $Y_t = 0$. Thus $Y_t = I_t X_t$, and $\{I_t\}$ is a sequence of IID random variables with $P(I_t = 0) = p$. We have

$$E(I_t) = 0 \times p + 1 \times (1-p) = 1-p.$$

Now, because of the independence,

$$E(Y_t) = E(I_t X_t) = E(I_t)E(X_t) = (1-p)\mu_X$$

and, for $h \leq 0$,

$$E(Y_t Y_{t+h}) = E(I_t X_t I_{t+h} X_{t+h}) = E(I_t)E(I_{t+h})E(X_t X_{t+h}) = (1-p)^2 E(X_t X_{t+h})$$

while

$$E(Y_t^2) = E(I_t^2 X_t^2) = E(I_t X_t^2) = E(I_t)E(X_t^2) = (1-p)E(X_t^2) = (1-p)(\sigma_X^2 + \mu_X^2)$$

say (since $I_t^2 = I_t$).

Thus

$$\mathrm{Var}(Y_t) = E(Y_t^2) - E(Y_t)^2 = (1-p)(\sigma_X^2 + \mu_X^2) - (1-p)^2\mu_X^2$$

while, for $h \neq 0$,

$$\begin{aligned} \mathrm{Cov}(Y_t, Y_t + h) &= E(Y_t Y_{t+h}) - E(Y_t)E(Y_{t+h}) = (1-p)^2 E(X_t X_{t+h}) - (1-p)^2 \mu_X^2 \\ &= (1-p)^2 \mathrm{Cov}(X_t, X_{t+h}) = (1-p)^2 \gamma_X(h). \end{aligned}$$

The above shows that neither $E(Y_t)$ nor $\mathrm{Cov}(Y_t, Y_t + h)$ depends on t, so $\{Y_t\}$ is stationary with variance $(1-p)(\sigma_X^2 + \mu_X^2) - (1-p)^2\mu_X^2$ and covariance function $(1-p)^2\gamma_X(h)$ for $h \neq 0$.

Problem 2

Fix t and put $U = X_{t+1} + X_{t+2}$. A linear predictor of U using X_t and X_{t-1} is of the form $a_1 X_t + a_2 X_{t-1}$, and the best linear predictor is given by $\boldsymbol{a} = [a_1\ a_2]^T$ such that $\boldsymbol{a} = \Gamma^{-1}\boldsymbol{\beta}$, where $\Gamma = \mathrm{Cov}([X_t\ X_{t-1}]^T)$ is the covariance matrix of the covariates, and $\boldsymbol{\beta} = \mathrm{Cov}([X_t\ X_{t-1}]^T, U)$ is the vector of covariances between the covariates and the variable being predicted (see FS, Section 7).

To compute these covariances, we need the ACVF of $\{X_t\}$. This process is an MA(2)-process, and its ACVF γ is given by

$$\begin{aligned}
\gamma(0) &= \mathrm{Var}(X_t) = \mathrm{Var}(Z_t - 0.4Z_{t-1} + 0.6Z_{t-2}) = 1^2\sigma^2 + (-0.4)^2\sigma^2 + 0.6^2\sigma^2 = 1.52\sigma^2,\\
\gamma(1) &= \mathrm{Cov}(Z_{t+1}, Z_t) = \mathrm{Cov}(Z_{t+1} - 0.4Z_t + 0.6Z_{t-1}, Z_t - 0.4Z_{t-1} + 0.6Z_{t-2})\\
&= -0.4 \times 1\sigma^2 + 0.6(-0.4)\sigma^2 = -0.64\sigma^2,\\
\gamma(2) &= \mathrm{Cov}(Z_{t+2}, Z_t) = \mathrm{Cov}(Z_{t+2} - 0.4Z_{t+1} + 0.6Z_t, Z_t - 0.4Z_{t-1} + 0.6Z_{t-2})\\
&= 0.6 \times 1\sigma^2 = 0.6\sigma^2,\\
\gamma(h) &= 0 \quad \text{for } h > 2,
\end{aligned}$$

with $\sigma^2 = 3$ and where we used bilinearity of the covariance operation (i.e., linearity in both arguments) and that $\{Z_t\}$ is WN.

Now

$$\Gamma = \begin{pmatrix} \mathrm{Cov}(X_t, X_t) & \mathrm{Cov}(X_t, X_{t-1}) \\ \mathrm{Cov}(X_{t-1}, X_t) & \mathrm{Cov}(X_{t-1}, X_{t-1}) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix} = \begin{pmatrix} 1.52 & -0.64 \\ -0.64 & 1.52 \end{pmatrix} \sigma^2,$$

and

$$\begin{aligned}
\boldsymbol{\beta} &= \begin{pmatrix} \mathrm{Cov}(X_t, U) \\ \mathrm{Cov}(X_{t-1}, U) \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(X_t, X_{t+1} + X_{t+2}) \\ \mathrm{Cov}(X_{t-1}, X_{t+1} + X_{t+2}) \end{pmatrix}\\
&= \begin{pmatrix} \gamma(1) + \gamma(2) \\ \gamma(2) + \gamma(3) \end{pmatrix} = \begin{pmatrix} -0.04 \\ 0.6 \end{pmatrix} \sigma^2.
\end{aligned}$$

The coefficients of the optimal predictor are given by

$$\boldsymbol{a} = \Gamma^{-1}\boldsymbol{\beta} = \frac{1}{1.52^2 - 0.64^2} \begin{pmatrix} 1.52 & 0.64 \\ 0.64 & 1.52 \end{pmatrix} \begin{pmatrix} -0.04 \\ 0.6 \end{pmatrix} = \begin{pmatrix} 0.1700 \\ 0.4663 \end{pmatrix},$$

i.e. the desired optimal predictor is $0.1700X_t + 0.4663X_{t-1}$.

The MSE of this predictor is (cf. FS, Section 7.1)

$$\begin{aligned}
\mathrm{Var}(U) - \boldsymbol{\beta}^T\Gamma^{-1}\boldsymbol{\beta} &= \mathrm{Var}(X_{t+1} + X_{t+2}) - (-0.04\ 0.6) \begin{pmatrix} 0.1700 \\ 0.4663 \end{pmatrix} \sigma^2\\
&= \mathrm{Var}(X_{t+1}) + \mathrm{Var}(X_{t+2}) + 2\,\mathrm{Cov}(X_{t+1}, X_{t+2}) - 0.2730\sigma^2\\
&= 2\gamma(0) + 2\gamma(1) - 0.2730\sigma^2 = 1.76\sigma^2 - 0.2730\sigma^2 = 1.4870\sigma^2 = 4.461.
\end{aligned}$$

Problem 3

Plots C and F have a continuous index on the x-axis, while the other plots have a discrete index. Since an SD is a function of a continuous variable (angular frequency or frequency), plots C and F must be SDs. Moreover, an ACF is maximal (and equal to one) at lag zero. Plots A and E are not maximal at time-index 0, and can hence not be ACFs. Thus plots A and E are realisations, and plots B and D are AFCs.

Summing up: A and E are realisations, B and D are ACFs, and C and F are PSDs.

A, D and C belong together (note the pronounced peak in SD C, associated with the slow oscillation in the ACF D, and corresponding slow oscillation in realisation A).

Also, E, B and F belong together (note that the SD F has large energy at high frequencies, corresponding to rapid oscillations in ACF B and realisation E).

Problem 4

(a) The estimates $\widehat{\gamma}(h)$ are relatively large for $h = 0$ and $h = 1$, and much smaller for larger h. We can compare to the significance threshold (under IID data) $1.96/\sqrt{n} = 1.96/\sqrt{1000} = 0.062$, which is exceeded only for $h \le 1$.

The estimated PACFs do not decay in the same way (look for instance at $h = 6$).

Thus, it seems reasonable to suggest an MA(q) model, for which me know that $\gamma(h) = 0$ for $|h| > q$. The data indicates $q = 1$.

For an AR(p) model the PACF is 0 for $|h| > p$, and hence we would have to choose a much larger model order to fit the data this way. To choose a model that is compatible with the data but not unnecessarily complicated, we suggest an MA(1) model.

(b) We compute moment estimates, i.e. equate theoretical values of $\gamma(h)$ to their estimated counterparts. For an MA(1) model

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

we know that $\gamma(0) = (1 + \theta^2)\sigma^2$ and $\gamma(1) = \theta\sigma^2$ (do the computation as in Problem 2, or see FS, Section 6.1).

The ACF at lag $h = 1$ is thus $\gamma(1)/\gamma(0) = \theta/(1 + \theta^2)$, and the corresponding estimate is $\widehat{\gamma}(1)/\widehat{\gamma}(0) = -0.276/0.596 = -0.4631$. Thus we need to solve

$$\frac{\theta}{1+\theta^2} = -0.4631 \quad \Leftrightarrow \quad \theta^2 + 2.1594\theta + 1 = 0.$$

This equation has two roots, –1.4868 and –0.6726. Either one will lead to an MA(1) process that fits $\gamma(0)$ and $\gamma(1)$, but we prefer the root yielding an invertible representation, i.e., such that the MA-polynomial $1 + \theta z$ has roots outside the complex unit circle. That root is $\theta = -0.6726$.

It remains to estimate the innovation variance σ^2. The above equations give the estimate $\widehat{\sigma}^2 = \widehat{\gamma}(0)/(1 + \widehat{\theta}^2) = 0.4104$.

Note: data was simulated from an MA(1) process with $\theta_1 = -0.5$ *and* $\sigma^2 = 0.4$.

Problem 5

The roots of the MA polynomial $1 + \theta_1 z + \theta_2 z^2 = 1 - 0.4z + 0.6z^2$, i.e. the solutions to

$$1 - 0.4z + 0.6z^2 = 0 \quad \Leftrightarrow \quad z^2 - \frac{2}{3}z + \frac{5}{3} = 0$$

are $1/3 \pm \sqrt{1/9 - 5/3} = 1/3 \pm i\sqrt{5/3 - 1/9}$, and their squared absolute value is $(1/3)^2 + (5/3 - 1/9) = 5/3$. They are thus outside the complex unit circle, so the MA(2) process under study is invertible.

This means that using all data X_s before some time t (i.e., $s \leq t$), we can reconstruct the innovation Z_t (and of course any Z_s with $s < t$). One may express Z_t as a infinite sum of past X_s (a series whose coefficients are absolutely summable), but the innovation algorithm does this recursively in time. Indeed, with the innovations algorithm run with 'infinite past', it holds that $X_t - \widehat{X}_t = Z_t$.

Now, using the model, we find that

$$\begin{aligned} X_{t+2} + X_{t+1} &= Z_{t+2} - 0.4Z_{t+1} + 0.6Z_t + Z_{t+1} - 0.4Z_t + 0.6Z_{t-1} \\ &= (Z_{t+2} + 0.6Z_{t+1}) + 0.2Z_t + 0.6Z_{t-1}. \end{aligned}$$

Using data X_s with $s \leq t$ we can, as argued above, reconstruct Z_t and Z_{t-1}. The innovations Z_{t+1} and Z_{t+2} are however uncorrelated with (orthogonal to) all data up to time t. Writing $\mathcal{P}_t$ for the projection operator onto the linear span $\text{span}(X_s;\, s \leq t)$ of all X_s with $s \leq t$, we thus have the best linear predictor of $X_{t+2} + X_{t+1}$ in terms of $\{X_s;\, s \leq t\}$ being

$$\mathcal{P}_t(X_{t+2} + X_{t+1}) = 0.2Z_t + 0.6Z_{t-1},$$

even though in practice we need to express the right-hand side in terms of $\{X_s;\, s \leq t\}$. To solve the problem we do not need to do this explicitly however; it suffices to know that we can do it.

Indeed, the prediction error is $Z_{t+2} + 0.6Z_{t+1}$, which is uncorrelated with X_s for each $s \leq t$. The variance of this prediction error is $\sigma^2 + 0.6^2\sigma^2 = 1.36\sigma^2 = 4.08$, and this is thus the MSE of the best linear predictor of $X_{t+2} + X_{t+1}$ in terms of $\{X_s;\, s \leq t\}$.