Faculteit der Exacte Wetenschappen
Afdeling Wiskunde
Vrije Universiteit

Statistical Models–written exam
December 10, 2002

1.
Describe briefly, but as completely as possible, a three-factor analysis of variance model with interactions.

2.
a. When is a time series $\{X_t\}$ stationary?
b. Describe briefly two models for a stationary time series.
c. What is (are) the main difference(s) between the (partial) autocorrelations of the two models which you have just described under b?
d. In Figure 1 on the next page the sample autocorrelation and the sample partial autocorrelation function of a time series are given. Which of the models AR, MA and ARMA do you think is best for this time series in view of these plots? Why? Give an estimate of the order of the chosen process.

3.
Describe the role of the deviance in a statistical analysis based on a generalized linear model. (It is not asked to give a formula.)

4.
Let

$$Y_i = f(x_i, \theta) + \varepsilon_i, \qquad i = 1, \dots, n$$

be a non-linear regression model with $Y_1, \dots, Y_n$ independent random variables, x_i a k-dimensional vector of explanatory variables for Y_i ($i = 1, \dots, n$), ε_i random measurement error with expectation 0 and variance σ^2 ($i = 1, \dots, n$), θ an unknown p-dimensional parameter vector and f a non-linear function of θ.
a. What is the definition of the least-square estimator $\hat{\theta}$ of θ?
b. Let $\widehat{\text{Cov}}(\hat{\theta}) = \hat{\sigma}^2(\hat{V}^T\hat{V})^{-1}$ be an estimator of the covariance matrix of $\hat{\theta}$. Give an approximate 95% confidence interval for θ_l, $l = 1, \dots, p$.
c. One has two data sets of which one thinks that the same non-linear model is applicable. The idea is that the parameter values are the same for both data sets, except for the value of θ_1. For both data sets approximate 95% confidence intervals for θ_1 are determined. One finds (4.32,4.68) and (4.59,5.19) for the first and second set, respectively, according

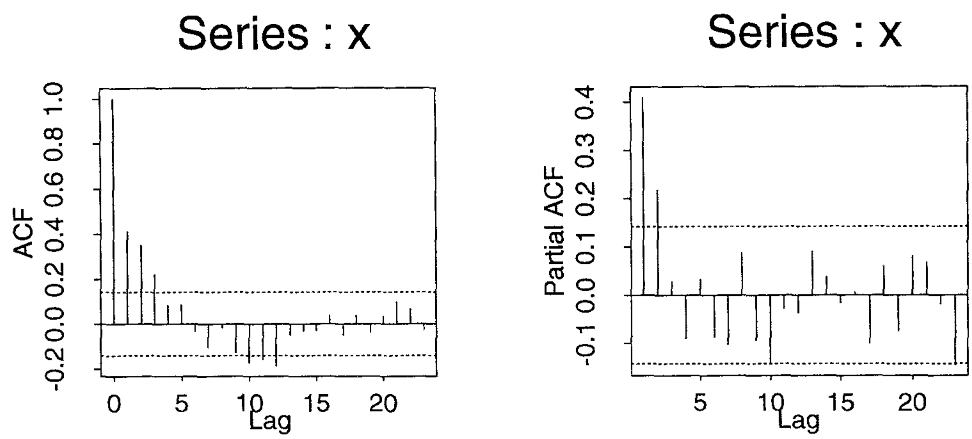

Figure 1: Sample autocorrelation and sample partial autocorrelation function of a time series.

to the classical theory; one finds (4.41,480) and (4.83,5.90) for the first and second set, respectively, with the bootstrap. What do you conclude from these intervals? Explain your conclusions.

5.

Below some practical situations are sketched. Choose for each case one or more suitable models out of the following set of models: linear regression model, analysis of variance model, non-linear model, generalized linear model, time series. Explain your choice, give the model and mention at least which variables are the responses and which the explanatory variables.

a. Overweight people following a particular diet show an exponential decrease of their lipid tissue during the period of the dietary intake. To investigate how long such a diet should be followed, one person has his weight measured on a daily basis during one month while he is on the diet. The measurement errors are small compared to the person's weight.

b. In an industrial process small errors occur irregularly. To investigate the cause of the errors, the following experiment is performed. Different stocks of raw material are selected. Every stock is divided in two parts. For each stock one part is treated with the standard method, and the other part with a modified method in which the temperature is decreased for a while. Before the treatment the purity of each stock as a whole is measured. For each part it is reported whether or not small errors have occurred during the treatment.

c. An insurance company investigates for different types of airplanes whether the number of reported damages to the planes depends on variables like year of construction, aggregate month service and number of flying hours.

Number of points to be earned:

1: 5	**2.a: 2**	**3: 4**	**4.a: 1**	**5.a: 3**
	2.b: 4		**4.b: 2**	**5.b: 3**
	2.c: 2		**4.c: 4**	**5.c: 3**
	2.d: 3			

Grade of exam = (total number of points)/4 + 1