

Opgave 1.

- a. Beschrijf kort, maar zo volledig mogelijk, een loglineair model.
- b. Geef een voorbeeld van een situatie waarin een dergelijk model kan worden gebruikt.

Opgave 2.

- a. Beschrijf kort de AR, MA en ARMA modellen voor een tijdreeks.
- b. Geef de definitie van de autocorrelatiefunctie van een stationair proces $\{X_t\}$.
- c. Wat is (zijn) het (de) essentiële verschil(len) tussen de autocorrelaties en de partiële autocorrelaties van de modellen die onder a. zijn beschreven?
- d. In Figuur 1 op de volgende bladzijde zijn de (steekproef) autocorrelatie- en (steekproef) partiële autocorrelatiefunctie van een tijdreeks gegeven. Welk van de modellen AR, MA en ARMA past op grond van deze plaatjes volgens jou het best bij deze tijdreeks? Schat ook de orde van het proces. Beargumenteer je keuze.
- e. Beschrijf kort hoe de orde van het proces op een meer formele manier kan worden geschat.

Opgave 3.

We beschikken over twee niet-lineaire regressiefuncties om onze data mee te beschrijven:

$$f(x, \theta) = \theta_1 \exp(\theta_2 x) + \theta_3 (1 - \exp(-\theta_4 x)), \quad \theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T, \quad (1)$$

en

$$f(x, \theta) = \theta_1 \exp(\theta_2 x), \quad \theta = (\theta_1, \theta_2)^T. \quad (2)$$

- a. Geef de algemene vorm van een niet-lineair regressiemodel.
- b. Veronderstel vervolgens dat de meetfouten in de data normaal verdeeld zijn. Beschrijf een toets om te bepalen welk van de twee regressiefuncties de data het best beschrijft.

Opgave 4.

Gegeven zijn de waarden van n realisaties van een responsvariabele met mogelijke waarden 0 en 1. Men wenst het verband te onderzoeken tussen deze responsvariabele en vier andere variabelen, te weten twee continue variabelen gemeten op een ratio schaal en twee variabelen gemeten op een nominale schaal met, respectievelijk, twee en vier categorieën.

- a. Stel een logistisch regressiemodel op voor dit probleem.

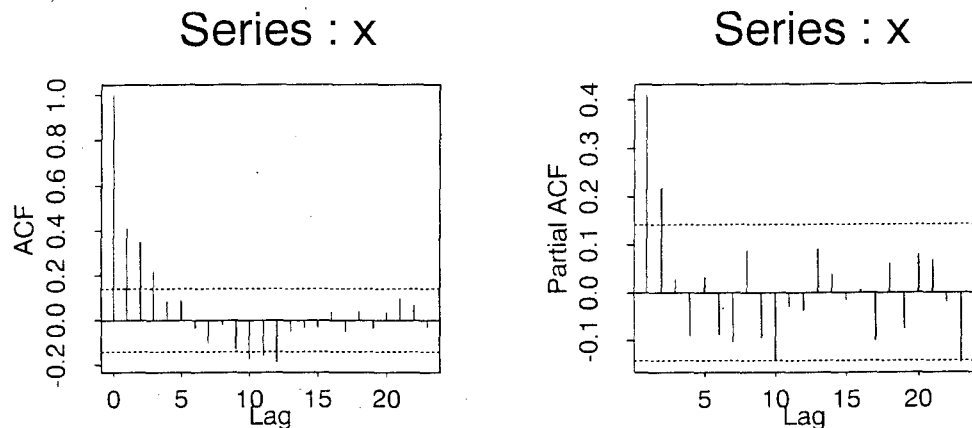


Figure 1: Autocorrelatie en partiële autocorrelatie functie van een tijdreeks.

- b. Hoe kan men per variabele toetsen of deze variabele in het model opgenomen dient te worden?

Opgave 5.

Hieronder worden enkele praktijksituaties geschetst. Kies voor elke situatie één of meer geschikte modellen. Kies hierbij uit: lineair regressiemodel, variantie-analyse model, niet-lineair model, gegeneraliseerd lineair model, tijdreeksmodel. Beargumenteer je antwoord. Beschrijf het gekozen model zo volledig mogelijk. Vermeld hierbij in elk geval wat de respons- en verklarende variabelen zijn.

- a. In een experiment ter vergelijking van de produktie van drie verschillende merken machines, werden van elk merk vijf machines gedurende een bepaalde tijd getest. De totale produktie van een machine werd uitgedrukt in kg produkt.
- b. Een fabrikant van chocolade paaseitjes wil nagaan of het toevoegen van een bepaalde smaakstof de smaak van de eitjes ten goede komt. De fabrikant heeft de keuze uit de smaakstoffen A, B en C. Een panel van 8 personen proeft de eitjes. Elk panellid krijgt van elk van de vier soorten (met toevoeging A, B of C en zonder toevoeging) 4 eitjes te proeven in een willekeurige volgorde die onafhankelijk is van de volgorde waarin de andere panelleden hun eitjes krijgen aangeboden. De panelleden worden geacht de smaak van een geproefd eitje te waarderen met een (geheel) cijfer tussen 1 en 10.

- c. Bij de afwegingen of koninginndag dit jaar wel of niet op de geplande wijze zal worden gevierd, wil de beslissende instantie een schatting maken van de hoeveelheid mensen die op een bepaalde plek langs zullen komen. Hiertoe vraagt zij gegevens op over bezoekersaantallen bij de verschillende evenementen tijdens de twintig voorgaande koninginndagen.

Normering:

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

$$\text{Eindcijfer tentamen} = \text{totaal}/4 + 1$$