

Regressionsmodelle mit Anwendungen in der Versicherungs- & Finanzwirtschaft SS 2015

Univ.-Prof. Dr. Michael Merz
Universität Hamburg



Übungsaufgaben

Übungsaufgaben

1) Berechnen Sie die Schiefe und Kurtosis folgender Zufallsvariablen:

a) X ist auf dem Intervall $[0, 1]$ gleichverteilt, d.h. besitzt die Dichte

$$f(x) = \begin{cases} 1 & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}.$$

b) X besitzt die Dichte

$$f(x) = \begin{cases} 2x & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}.$$

c) $X \sim N(\mu, \sigma^2)$, d.h. X ist normalverteilt mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ (Hinweis: Verwenden Sie die momenterzeugende Funktion

$$M_X(t) = \mathbb{E}[e^{tX}] = e^{\mu t + \sigma^2 t^2 / 2}.$$

2) Approximieren Sie mit Hilfe des zentralen Grenzwertsatzes die folgenden Wahrscheinlichkeiten:

a) Eine faire Münze wird 1000 mal geworfen. Wie hoch ist die Wahrscheinlichkeit, dass häufiger als 530 mal Kopf fällt?

b) In einem Versicherungsportfolio weisen 70% aller Risiken ein Versicherungsschaden auf. Wie hoch ist die Wahrscheinlichkeit, dass bei 28 unabhängigen Risiken dieses Portfolio mindestens 20 Risiken einen Versicherungsschaden aufweisen?

- 3) Die Zufallsvariablen X_1, X_2, X_3, X_4 seien stochastisch unabhängig und identisch Bernoulli-verteilt mit dem Parameter $p \in (0, 1)$.
- a) Bestimmen Sie die Wahrscheinlichkeits- und Verteilungsfunktion des zweidimensionalen Zufallsvektors $\mathbf{X}_1 = (X_1, X_2)^T$.
 - b) Ermitteln Sie den Erwartungswertvektor und die Varianz-Kovarianzmatrix des vierdimensionalen Zufallsvektors $\mathbf{X}_2 = (X_1, X_2, X_3, X_4)^T$.
- 4) Der Zufallsvektor $\mathbf{X} = (X_1, X_2, X_3)^T$ sei dreidimensional $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -verteilt mit

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \quad \text{und} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix}.$$

- a) Geben Sie die Verteilung von X_1 und von X_1 , bedingt gegeben $X_2 = 3$, an.
- b) Ermitteln Sie die Verteilung des zweidimensionalen Zufallsvektors $(X_1, X_2)^T$, bedingt gegeben X_3 .
- c) Gegeben sei

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 2 & 0 \end{pmatrix}.$$

Bestimmen Sie die Verteilung von $\mathbf{A}\mathbf{X}$.

- 5) Ein Mitarbeiter in der Marketingabteilung eines Finanzdienstleisters vermutet, dass es einen positiven Zusammenhang zwischen den Ausgaben (in Millionen Euro) für Werbung und dem Absatz (in Millionen Euro) des beworbenen Finanzproduktes gibt. In den letzten drei Jahren wurden folgende Zahlen beobachtet:

Absatz y	Ausgaben x
5	2
8	3
9	4

Es wird im Folgenden davon ausgegangen, dass ein linearer Zusammenhang zwischen dem Absatz und den Ausgaben vorliegt.

- Bestimmen Sie die Designmatrix \mathbf{X} und den KQ-Schätzer $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$.
- Bestimmen Sie eine Schätzung für die Varianz-Kovarianzmatrix des KQ-Schätzers $\hat{\beta}$.
- Ermitteln Sie den Wert des Bestimmtheitsmaßes R^2 und des adjustierten Bestimmtheitsmaßes R_a^2 .
- Testen Sie die Signifikanz der erklärenden Variablen x bei einem Signifikanzniveau von $\alpha = 0,05$ und $\alpha = 0,01$.
- Ermitteln Sie für den Regressionskoeffizienten β_1 das 95%-Konfidenzintervall. Kommentieren Sie das Ergebnis kurz.

- 6) Es soll der Zusammenhang zwischen einer abhängigen Variablen y und zwei erklärenden Variablen x_1 und x_2 untersucht werden. Hierzu wird das lineare Regressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

mit $\text{Var}(\varepsilon) = \sigma^2$ betrachtet. Es liegen die folgenden Beobachtungen vor:

y	x_1	x_2
1	2	0
2	0	0
4	0	1
1	0	-1

- a) Berechnen Sie den KQ-Schätzer für $\beta = (\beta_0, \beta_1, \beta_2)^T$.
b) Bestimmen Sie die Residuen $\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3, \hat{\varepsilon}_4)^T$ und R^2 .
c) Ermitteln Sie Schätzungen für σ^2 und $\text{Cov}(\hat{\beta}, \hat{\beta})$.
- 7) Bestimmen Sie in Abhängigkeit von $\mathbf{y} = (y_1, \dots, y_9)^T$ den KQ-Schätzer für das lineare Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad \text{für } i = 1, \dots, 9$$

mit $\text{Var}(\varepsilon_i) = \sigma^2$ und den Beobachtungen $x_1 = x_2 = x_3 = -1$,
 $x_4 = x_5 = x_6 = 0$ und $x_7 = x_8 = x_9 = 1$ für die erklärenden Variablen.

- 8) Es wird ein lineares Modell mit sechs zu schätzenden Parametern β_0, \dots, β_5 betrachtet.

- a) Ermitteln Sie den Wert der F -Statistik für die Testsituation

$$H_0 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T = \mathbf{0} \quad \text{gegen} \quad H_1 : (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T \neq \mathbf{0}$$

und den Fall, dass $n = 40$ Beobachtungen vorliegen und das Bestimmtheitsmaß $R^2 = 0,2$ beträgt. Beurteilen Sie anschließend, ob die Nullhypothese H_0 bei einem Signifikanzniveau von $\alpha = 0,05$ bzw. $\alpha = 0,1$ zu verwerfen ist.

- b) Es liegen nun $n = 400$ Beobachtungen vor. Beurteilen Sie, ob die Nullhypothese H_0 jetzt zu verwerfen ist.

- c) Der KQ-Schätzer für β sei nun durch $\hat{\beta} = (2, 2, 3, 3, 4, 1)^T$ gegeben. Berechnen Sie das 95%- und 99%-Konfidenzintervall für die Linearkombination $\mathbf{c}^T \hat{\beta}$ mit $\mathbf{c}^T = (1, 2, 1, 4, 5, 3)$ und $\hat{\sigma}(\mathbf{c}^T \hat{\beta}) = 0,9$ für den Fall, dass $n = 40$ Beobachtungen vorliegen.

- d) Der KQ-Schätzer für β sei wieder durch $\hat{\beta} = (2, 2, 3, 3, 4, 1)^T$ gegeben und $n = 40$. Ferner gelte für den Vektor mit den Werten der erklärenden Variablen $\mathbf{x}_* = (1, 1, 3, 2, 2, 1)^T$ und $\hat{\sigma}(\mathbf{x}_*^T \hat{\beta}) = 0,3$. Als Schätzung für den Varianzparameter (mittlerer quadratische Fehler) σ^2 liegt der Wert $\hat{\sigma}^2 = 4$ vor. Berechnen Sie mittels diesen Angaben für y das $(1 - \alpha)$ -Prognoseintervall zum Signifikanzniveau $\alpha = 0,95$.

- 9) Die folgende Tabelle enthält für die zwölf Risikoklassen (i, j) eines Autohaftpflichtportfolios die Gesamtschadenhöhen S_{ij} :

Fahrzeuggewicht i	Alter j			
	21-30 ($j = 1$)	31-40 ($j = 2$)	41-50 ($j = 3$)	51-60 ($j = 4$)
leicht ($i = 1$)	1859	1872	1430	2028
mittel ($i = 2$)	2376	1872	1716	2340
schwer ($i = 3$)	2772	2184	2002	2730

- Stellen Sie ein allgemeines lineares Modell für die logarithmierten Gesamtschadenhöhen $\ln(S_{ij})$ auf und wählen Sie dabei die Risikoklasse $(1, 1)$ als Referenzklasse.
- Passen Sie das allgemeine lineare Modell an die logarithmierten Gesamtschadenhöhen $\ln(S_{ij})$ an (Lognormales-Verfahren) und beurteilen Sie die Anpassungsgüte.
- Bestimmen Sie Schätzungen \hat{p}_{ij} für die erwartete Gesamtschadenhöhe in den zwölf Risikoklassen.

- 10) Es wird wieder die Situation aus Aufgabe 5) betrachtet.
- Ermitteln Sie eine Schätzung für die Varianz-Kovarianzmatrix der Residuen $\hat{\varepsilon}$ und Schätzungen für die Varianzen der Residuen $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \hat{\varepsilon}_3$.
 - Berechnen Sie die standardisierten Residuen r_i für $i = 1, 2, 3$ und beurteilen Sie damit und einer gängigen Daumenregel, ob es sich bei den Beobachtungen y_1, y_2, y_3 um Ausreißer handelt.
 - Beurteilen Sie anhand einer gängigen Daumenregel für die Hebelwerte, ob die Beobachtungen y_1, y_2, y_3 selbst-schätzend sind. Erläutern Sie ferner, ob die Anwendung dieser Daumenregel in diesem Fall sinnvoll ist.
- 11) Weisen Sie nach, dass für die Hebelwerte h_{ii} im linearen Modell $y = \beta_0 + \beta_1 x + \varepsilon$

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad \text{und} \quad \bar{h} := \frac{1}{n} \sum_{i=1}^n h_{ii} = \frac{2}{n}$$

gilt.

- 12) Weisen Sie nach, dass die Binomialverteilung zur Exponential-Dispersions-Familie gehört.
- 13) Weisen Sie nach, dass die Gammaverteilung zur Exponential-Dispersions-Familie gehört.

- 14) In einem verallgemeinerten linearen Modell ist die Nullhypothese

$$H_0 : \beta_j = 0 \quad \text{gegen} \quad H_1 : \beta_j \neq 0$$

mit Hilfe der Wald-Statistik zu testen. Für die ML-Schätzung von β_j und deren Varianz gilt

$$\hat{\beta}_j = 2 \quad \text{und} \quad \text{Var}(\hat{\beta}_j) = 1,5.$$

Berechnen Sie den p -Wert und geben Sie an, ob die Nullhypothese zum Signifikanzniveau von $\alpha = 0,05$ zu verwerfen ist.

- 15) Betrachtet wird ein verallgemeinertes lineares Modell mit $k = 5$ erklärenden Variablen und $n = 10$ Beobachtungen. Zu testen ist

$$H_0 : \beta_1 = \beta_2 \quad \text{gegen} \quad H_1 : \beta_1 \neq \beta_2.$$

Für die Log-Likelihoodfunktion resultieren die Werte

$$\ln \left(L \left(\hat{\beta}_{H_0}; y_1, \dots, y_n \right) \right) = 0,5 \quad \text{und} \quad \ln \left(L \left(\hat{\beta}_{H_1}; y_1, \dots, y_n \right) \right) = 2,5.$$

Untersuchen Sie, ob die Nullhypothese $H_0 : \beta_1 = \beta_2$ zum Signifikanzniveau $\alpha = 0,05$ abzulehnen ist.

- 16) Gegeben sei ein verallgemeinertes lineares Modell und zu testen sei die allgemeine lineare Hypothese

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d} \quad \text{gegen} \quad H_1 : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$$

mit $\text{rang}(\mathbf{C}) = 2$. Für die Log-Likelihoodfunktion resultieren die Werte

$$\ln \left(L \left(\hat{\boldsymbol{\beta}}_{H_0}; y_1, \dots, y_n \right) \right) = 0,3 \quad \text{und} \quad \ln \left(L \left(\hat{\boldsymbol{\beta}}_{H_1}; y_1, \dots, y_n \right) \right) = 0,9.$$

Prüfen Sie, ob die Nullhypothese H_0 zum Signifikanzniveau $\alpha = 0,05$ bzw. $\alpha = 0,01$ zu verwerfen ist.

- 17) Bisher wurde in einem Unternehmen ein verallgemeinertes lineares Modell M_1 mit $k = 3$ erklärenden Variablen eingesetzt. Das Unternehmen steht vor der Entscheidung, ob zukünftig ein erweitertes Modell M_2 verwendet werden sollte, bei dem zu den bisherigen drei erklärenden Variablen zwei weitere hinzugefügt wurden. D.h. zu testen ist

$$H_0 : M_1 \quad \text{gegen} \quad H_1 : M_2.$$

Bei Vorliegen von $n = 20$ Beobachtungen betragen die skalierten Devianzen

$$\Delta_{M_1}^* = 7,2 \quad \text{und} \quad \Delta_{M_2}^* = 1,1.$$

Entscheiden Sie anhand dieser Informationen, ob die Nullhypothese $H_0 : M_1$ zum Signifikanzniveau $\alpha = 0,05$ abzulehnen ist.

- 18) Der Datensatz `Driver_testing_results` enthält für 26 Gruppen (geordnet von 18 bis 30 Jahren und Geschlecht m/f) die Anzahl der Teilnehmer, die eine Fahrprüfung erfolgreich bestanden haben.

	AGE	NUMBER	PASS	PASS_RATE	GENDER
1	18	25	5	0.20	m
2	19	37	14	0.38	m
3	20	38	18	0.47	m
4	21	42	18	0.43	m
5	22	37	22	0.59	m
6	23	36	18	0.50	m
7	24	45	30	0.67	m
8	25	33	19	0.58	m
9	26	27	18	0.67	m
10	27	31	23	0.74	m
11	28	33	23	0.70	m
12	29	27	23	0.85	m
13	30	26	24	0.92	m
14	18	20	11	0.55	f
15	19	22	13	0.59	f
16	20	29	15	0.52	f
17	21	37	19	0.51	f
18	22	31	15	0.48	f
19	23	30	19	0.63	f
20	24	36	20	0.56	f
21	25	42	29	0.69	f
22	26	40	23	0.58	f
23	27	34	22	0.65	f
24	28	28	19	0.68	f
25	29	38	26	0.68	f
26	30	42	27	0.64	f

D.h. zur Analyse der Erfolgsquote $PASS_RATE = PASS/NUMBER$ liegen die Merkmale (erklärenden Variablen)

Bezeichnung	Variable
y	PASS_RATE
x_1	GENDER
x_2	AGE

vor.

- Passen Sie an die Zielvariable $PASS_RATE$ ein Bernoulli-GLM mit der sog. Logit-Linkfunktion $g(\mathbb{E}[y]) = \ln\left(\frac{\mathbb{E}[y]}{1-\mathbb{E}[y]}\right)$, den beiden erklärenden Variablen $GENDER$ und AGE sowie deren Interaktion $AGE * GENDER$ an. Verwenden Sie dabei die Variable $NUMBER$ als Gewicht (Größe/Volumina) w der einzelnen Gruppen.
- Beurteilen Sie anhand der Wald-Statistik, welche erklärenden Variablen einen statistisch signifikanten Einfluss auf die erwartete Erfolgsquote haben, und interpretieren Sie den resultierenden linearen Prädiktor.
- Ermitteln Sie eine Schätzung für die Erfolgsquote der Gruppe der 22-jährigen Frauen.
- Stellen Sie getrennt für Frauen und Männer in Abhängigkeit der Variablen AGE die beobachteten und geschätzten Erfolgsquoten in einer Abbildung dar.

- 19) Die folgende Tabelle enthält die beobachteten Schadenanzahlen N_{ij} und Schadenfrequenzen R_{ij} für die zwanzig Risikoklassen (i,j) eines Autohaftpflichtportfolios:

Hubraumklasse		Geschlecht		
		M	F	Gesamt
1	Anzahl Risiken V_{ij}	574	409	983
	Anzahl Schaden N_{ij}	32	20	52
	Frequenz R_{ij}	0,0557	0,0489	0,0529
2	Anzahl Risiken V_{ij}	11429	14255	25684
	Anzahl Schaden N_{ij}	702	832	1534
	Frequenz R_{ij}	0,0614	0,0584	0,0597
3	Anzahl Risiken V_{ij}	19049	15896	34945
	Anzahl Schaden N_{ij}	1237	950	2187
	Frequenz R_{ij}	0,0649	0,0598	0,0626
4	Anzahl Risiken V_{ij}	36862	17285	54147
	Anzahl Schaden N_{ij}	2587	1248	3835
	Frequenz R_{ij}	0,0702	0,0722	0,0708
5	Anzahl Risiken V_{ij}	35627	13467	49094
	Anzahl Schaden N_{ij}	2617	896	3513
	Frequenz R_{ij}	0,0735	0,0665	0,0716

6	Anzahl Risiken V_{ij}	52717	9804	62521
	Anzahl Schaden N_{ij}	3800	704	4504
	Frequenz R_{ij}	0,0721	0,0718	0,0720
7	Anzahl Risiken V_{ij}	23920	3642	27562
	Anzahl Schaden N_{ij}	1770	283	2053
	Frequenz R_{ij}	0,0740	0,0777	0,0745
8	Anzahl Risiken V_{ij}	13818	2064	15882
	Anzahl Schaden N_{ij}	1117	142	1259
	Frequenz R_{ij}	0,0808	0,0688	0,0793
9	Anzahl Risiken V_{ij}	14341	1822	16163
	Anzahl Schaden N_{ij}	1170	157	1327
	Frequenz R_{ij}	0,0816	0,0862	0,0821
10	Anzahl Risiken V_{ij}	5917	862	6779
	Anzahl Schaden N_{ij}	535	76	611
	Frequenz R_{ij}	0,0904	0,0882	0,0901
Gesamt	Anzahl Risiken V_{ij}	214254	79506	293760
	Anzahl Schaden N_{ij}	15567	5308	20875
	Frequenz R_{ij}	0,0727	0,0668	0,0711

- a) Es wird angenommen, dass die Schadenanzahlen N_{ij} stochastisch unabhängig und Poisson-verteilt $N_{ij} \sim \Pi(V_{ij}a_i b_j)$ sind für $i = 1, 2, \dots, 10$ und $j = 1, 2$. Mit dem sog. Marginalsummenverfahren resultieren für die Parameter a_i und b_j für $i = 1, 2, \dots, 10$ und $j = 1, 2$ die Schätzungen

$$\hat{\mathbf{a}} = \begin{pmatrix} 0,0536 \\ 0,0608 \\ 0,0635 \\ 0,0715 \\ 0,0722 \\ 0,0724 \\ 0,0748 \\ 0,0796 \\ 0,0824 \\ 0,0905 \end{pmatrix} \quad \text{und} \quad \hat{\mathbf{b}} = \begin{pmatrix} 1,0000 \\ 0,9684 \end{pmatrix}.$$

Diese Schätzungen werden bei der Verwendung des Marginalsummenverfahrens gerade so bestimmt, dass die folgenden sog. Marginalsummenbedingungen erfüllt sind:

$$\sum_{j=1}^2 V_{ij} \widehat{a}_i \widehat{b}_j = \sum_{j=1}^2 N_{ij} \quad \text{für alle } i = 1, \dots, 10$$

$$\sum_{i=1}^{10} V_{ij} \widehat{a}_i \widehat{b}_j = \sum_{i=1}^{10} N_{ij} \quad \text{für alle } j = 1, 2$$

Überprüfen Sie anhand eines Likelihood-Quotienten-Tests, ob bei einem Signifikanzniveau von $\alpha = 5\%$ ein signifikanter Unterschied zwischen Männern und Frauen besteht.

- b) Passen Sie an die Frequenz R_{ij} ein Poisson-GLM mit der Log-Linkfunktion $g(\mathbb{E}[R_{ij}]) = \ln(\mathbb{E}[R_{ij}])$ sowie den beiden erklärenden Variablen **Hubraumklasse** und **Geschlecht** an. Verwenden Sie dabei V_{ij} als Gewicht (Größe/Volumina) für die einzelnen Gruppen. Was ist zu beobachten?

20) Betrachtet wird die Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = U_1 \sin(2\pi\delta t) + U_2 \cos(2\pi\delta t) \quad \text{für } t \in \mathbb{Z},$$

wobei U_1 und U_2 stochastisch unabhängig sind mit

$$\mathbb{E}[U_1] = \mathbb{E}[U_2] = 0 \quad \text{und} \quad \text{Var}(U_1) = \text{Var}(U_2) = \sigma^2.$$

Zeigen Sie, dass diese Zeitreihe schwach stationär ist und für die Autokovarianzfunktion

$$\gamma(h) = \sigma^2 \cos(2\pi\delta h) \quad \text{für } h \in \mathbb{Z}$$

gilt.

(Hinweis: Verwenden Sie $\sin(\alpha)\sin(\beta) + \cos(\alpha)\cos(\beta) = \cos(\alpha - \beta)$)

21) Weisen Sie nach, dass die Zeitreihe $(X_t)_{t \in \mathbb{N}}$ mit

$$X_t = \sin(2\pi Ut) \quad \text{für } t \in \mathbb{N}$$

und einer auf dem Intervall $(0, 1)$ gleichverteilten Zufallsvariablen U schwach stationär ist.

(Hinweis: Verwenden Sie $\sin(\alpha)\sin(\beta) = \frac{1}{2}(\cos(\alpha - \beta) - \cos(\alpha + \beta))$)

22) Betrachtet wird der stochastische Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \beta_1 + \beta_2 t + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

sowie $\beta_1, \beta_2 \in \mathbb{R}$ und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

- Untersuchen Sie, ob $(X_t)_{t \in \mathbb{Z}}$ schwach stationär ist.
- Zeigen Sie, dass $(\nabla X_t)_{t \in \mathbb{Z}}$ mit $\nabla X_t = X_t - X_{t-1}$ schwach stationär ist.
- Weisen Sie für den Prozess $(Y_t)_{t \in \mathbb{Z}}$ mit

$$Y_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t-j} \quad \text{für } t \in \mathbb{Z}$$

nach, dass $\mathbb{E}[Y_t] = \beta_1 + \beta_2 t$ gilt.

- Bestimmen Sie die Autokovarianzfunktion von $(Y_t)_{t \in \mathbb{Z}}$.

23) Betrachtet wird der stochastische Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \varepsilon_t \varepsilon_{t-1} \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{IID}(0, \sigma^2)$. Bestimmen Sie die Mittelwert- und Autokovarianzfunktion von $(X_t)_{t \in \mathbb{Z}}$ und geben Sie an, ob $(X_t)_{t \in \mathbb{Z}}$ schwach stationär ist.

- 24) Gegeben sei der stochastische Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1} \quad \text{für } t \in \mathbb{Z}$$

sowie $\theta \in \mathbb{R}$ und $\varepsilon_t \sim \text{IID}(0, \sigma^2)$. Durch $M_\varepsilon(s) := \mathbb{E}[e^{s\varepsilon_t}]$ ist die momentenerzeugende Funktion der Zufallsvariablen ε_t für $t \in \mathbb{Z}$ gegeben.

- a) Drücken Sie die momenterzeugende Funktion

$$M_{(X_1, \dots, X_n)}(s_1, \dots, s_n) = \mathbb{E} \left[\exp \left(\sum_{t=1}^n s_t X_t \right) \right]$$

des Zufallvektors $(X_1, \dots, X_n)^T$ mit Hilfe von $M_\varepsilon(s)$ aus.

- b) Zeigen Sie mit Hilfe des Aufgabenteils a), dass $(X_t)_{t \in \mathbb{Z}}$ eine stark stationäre Zeitreihe ist.
- 25) Es seien $(X_t)_{t \in \mathbb{Z}}$ und $(Y_t)_{t \in \mathbb{Z}}$ zwei schwach stationäre und unkorrelierte Prozesse. Zeigen Sie, dass dann auch $(Z_t)_{t \in \mathbb{Z}}$ mit $Z_t := X_t + Y_t$ für $t \in \mathbb{Z}$ schwach stationär ist.
- 26) Weisen Sie nach, dass absolute Summierbarkeit stets quadratische Summierbarkeit impliziert. D.h. es gilt

$$\sum_{j=0}^{\infty} |\eta_j| < \infty \quad \implies \quad \sum_{j=0}^{\infty} \eta_j^2 < \infty.$$

27) Multiplizieren Sie die beiden folgenden Lag-Polynome aus und benennen Sie die zugehörigen Prozesse minimaler Ordnung:

a) $(1 - 0,2L)(1 - 0,6L)X_t = (1 + 0,5L)\varepsilon_t$

b) $(1 - 0,7L)(1 - 0,4L^2)(1 - L)X_t = (1 + 0,5L)(1 - 0,4L^2)\varepsilon_t$

28) Im Folgenden gelte $\varepsilon_t \sim \text{WN}(0, \sigma^2)$. Benennen Sie die untenstehenden stochastischen Prozesse:

a) $X_t = \varepsilon_t + 2\varepsilon_{t-1}$

b) $X_t = \frac{13}{10}X_{t-1} + \varepsilon_t$

c) $X_t = \frac{13}{10}X_{t-1} - \frac{2}{5}X_{t-2} + \varepsilon_t$

d) $X_t = \frac{13}{10}X_{t-1} - \frac{2}{5}X_{t-2} + \varepsilon_t - \frac{3}{10}\varepsilon_{t-1}$

e) $X_t = \frac{1}{5}X_{t-1} + \frac{4}{5}X_{t-2} + \varepsilon_t$

f) $X_t = \frac{1}{5}X_{t-1} + \frac{4}{5}X_{t-2} + \varepsilon_t - \frac{3}{2}\varepsilon_{t-1} + \frac{1}{2}\varepsilon_{t-2}$

Untersuchen Sie die Prozesse ferner auf Parameterredundanz, schwache Stationarität, Kausalität und Invertierbarkeit.

29) Bestimmen Sie für den AR(2)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = -0,9X_{t-2} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ das zugehörige Autoregressive Polynom $\Phi_2(z)$ sowie dessen Nullstellen in Polardarstellung. Stellen Sie die Autokorrelationsfunktion $\rho(h)$ als Funktion dieser Nullstellen dar und interpretieren Sie das Ergebnis.

- 30) Ermitteln Sie für den AR(1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = 0,8X_{t-1} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ die Varianz des Durchschnitts

$$\frac{X_1 + X_2 + X_3 + X_4}{4}.$$

- 31) Ermitteln Sie mittels der Methode der unbestimmten Koeffizienten die kausale Darstellung der beiden stochastischen Prozesse

a) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + \varepsilon_t$

b) $X_t = 1,3X_{t-1} - 0,4X_{t-2} + \varepsilon_t - 0,2\varepsilon_{t-1}$

mit $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

(Hinweis: Verwenden Sie, dass die Differenzengleichung 2. Ordnung

$$0,4\psi_{j-1} - 1,3\psi_j + \psi_{j+1} = 0 \quad \text{für } j \in \mathbb{N}$$

die Lösung

$$\psi_j = c_1 \left(\frac{1}{z_1}\right)^j + c_2 \left(\frac{1}{z_2}\right)^j$$

besitzt, wobei z_1 und z_2 die Nullstellen des Polynoms $\Phi_2(z) = 1 - 1,3z + 0,4z^2$ sind und $c_1 + c_2 = 1$ gilt.)

32) Betrachtet wird ein ARMA(2, 1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ sowie

$$\Phi_2(L) = 1 - 1,3L + 0,4L^2 \quad \text{und} \quad \Theta_1(L) = 1 + 0,7L.$$

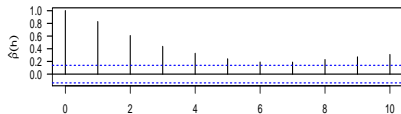
- Berechnen Sie mittels der Methode der unbestimmten Koeffizienten die kausale Darstellung von $(X_t)_{t \in \mathbb{Z}}$.
(Hinweis: Verwenden Sie dabei den Hinweis zu Aufgabe 12)
- Berechnen Sie mit Hilfe des Ergebnisses aus Aufgabenteil a) die Autokovarianzfunktion von $(X_t)_{t \in \mathbb{Z}}$.

33) Benennen Sie die folgenden vier stochastischen Prozesse:

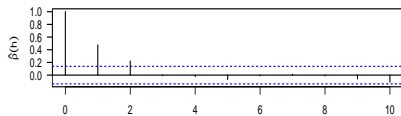
- $X_t = -0,5X_{t-1} + 0,3X_{t-2} + \varepsilon_t$
- $X_t = \varepsilon_t + 0,5\varepsilon_{t-1} + 0,3\varepsilon_{t-2}$
- $X_t = -0,8X_{t-1} + \varepsilon_t + 0,3\varepsilon_{t-1}$
- $X_t = 0,8X_{t-1} + \varepsilon_t + 0,3\varepsilon_{t-1}$

Ordnen Sie diese stochastischen Prozesse den folgenden vier empirischen Autokorrelationsfunktionen und empirischen partiellen Autokorrelationsfunktionen zu und begründen Sie ihre Zuordnung kurz.

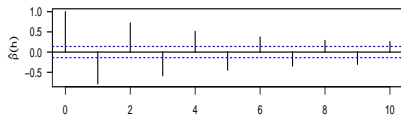
Prozess 1



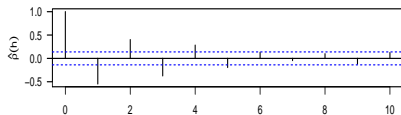
Prozess 2



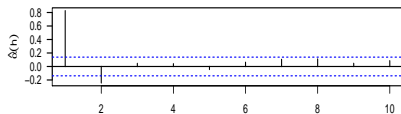
Prozess 3



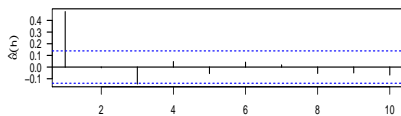
Prozess 4



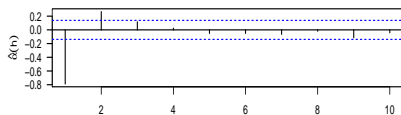
Prozess 1



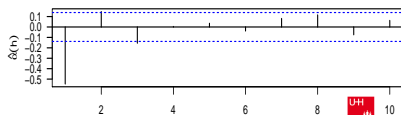
Prozess 2



Prozess 3



Prozess 4



34) Betrachtet wird der stochastische Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = -2X_{t-1} - \frac{5}{4}X_{t-2} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

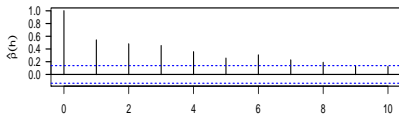
- Benennen Sie den Prozess und geben Sie den zugehörigen Autoregressiven-Operator und Moving-Average-Operator an. Stellen Sie $(X_t)_{t \in \mathbb{Z}}$ mit Hilfe dieser beiden Operatoren dar.
- Untersuchen Sie den Prozess auf Parameterredundanz, schwache Stationarität, Kausalität und Invertierbarkeit.

35) Benennen Sie die folgenden vier stochastischen Prozesse:

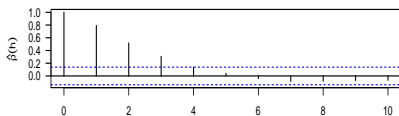
- $X_t = 0,9X_{t-1} - 0,1X_{t-2} + \varepsilon_t$
- $X_t = \varepsilon_t + 0,7\varepsilon_{t-1} + 0,5\varepsilon_{t-2}$
- $X_t = 0,4X_{t-1} + 0,3X_{t-2} + \varepsilon_t$
- $X_t = 0,8X_{t-1} - 0,2X_{t-2} + \varepsilon_t - 0,3\varepsilon_{t-1} + 0,5\varepsilon_{t-2}$

Ordnen Sie diese stochastischen Prozesse den folgenden vier empirischen Autokorrelationsfunktionen und empirischen partiellen Autokorrelationsfunktionen zu und begründen Sie ihre Zuordnung kurz.

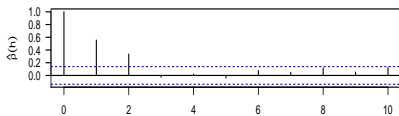
Prozess 1



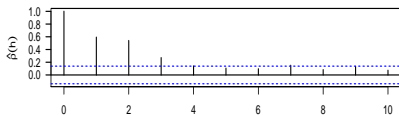
Prozess 2



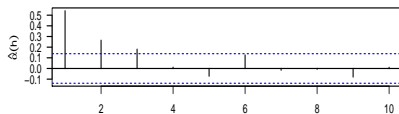
Prozess 3



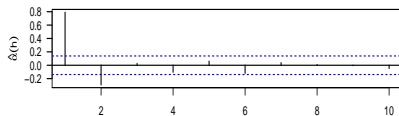
Prozess 4



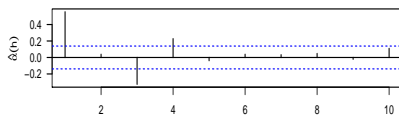
Prozess 1



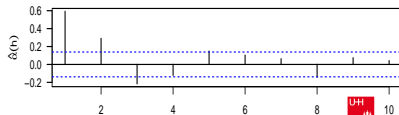
Prozess 2



Prozess 3



Prozess 4



- 36) Für eine schwach stationäre Zeitreihe $(X_t)_{t \in \mathbb{Z}}$ mit $\mathbb{E}[X_t] = 0$ soll zum Zeitpunkt t die Zufallsvariable X_{t+k} für $k \in \mathbb{N}$ prognostiziert werden.
- a) Als Prädiktor für X_{t+k} wird aX_t mit $a \in \mathbb{R}$ verwendet. Zeigen Sie, dass der Mean square error of prediction (MSEP)

$$\text{MSEP}(a) := \mathbb{E} \left[(X_{t+k} - aX_t)^2 \right]$$

für $a = \rho(k)$ minimiert wird.

- b) Zeigen Sie, dass gilt

$$\min_{a \in \mathbb{R}} \text{MSEP}(a) = \gamma(0) \left(1 - \rho^2(k) \right).$$

- 37) Gegeben sei ein kausaler AR(1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \phi_1 X_{t-1} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$. Ermitteln Sie den BLU-Prädiktor \hat{X}_t^{opt} für X_t , wenn die Beobachtungen X_{t-1} und X_{t+1} vorliegen.

38) Betrachtet wird der AR(1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \mu + \phi_1 X_{t-1} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

- a) Mittels der Momentenmethode wurde für ϕ_1 die Schätzung $\hat{\phi}_1 = 0,7$ ermittelt. Berechnen Sie die benötigte Anzahl T an Beobachtungen, für welche das 95%-Konfidenzintervall von $\hat{\phi}_1$ nicht größer als $\hat{\phi}_1 \pm 0,1$ ist.
- b) Für den schwach stationären AR(2)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \quad \text{für } t \in \mathbb{Z}$$

und $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ wurden anhand von $T = 100$ Beobachtungen die Schätzungen $\hat{\rho}(1) = 0,8$ und $\hat{\rho}(2) = 0,5$ sowie

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^{100} X_t = 2$$

ermittelt. Bestimmen Sie hiermit die Momentenschätzungen für die Parameter ϕ_1, ϕ_2 und μ .

39) Betrachtet wird ein schwach stationärer GARCH(1, 1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \sigma_t \varepsilon_t \quad \text{und} \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{für } t \in \mathbb{Z}$$

sowie

$$\varepsilon_t \sim \text{IID}(0, 1) \quad \text{und} \quad \mathbb{E}[X_t^4] < \infty \quad \text{für } t \in \mathbb{Z}.$$

- Erläutern Sie, ob dieser GARCH-Prozess eine ARCH(∞)-Darstellungen besitzt, und ermitteln Sie diese gegebenenfalls.
- Stellen Sie den quadrierten Prozess $(X_t^2)_{t \in \mathbb{Z}}$ als ARMA(1, 1)-Prozess dar.
- Es gelte nun zusätzlich $\varepsilon_t \sim \text{IIN}(0, 1)$. Weisen Sie nach, dass die Kurtosis $K(X_t)$ genau dann existiert, wenn $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$ gilt, und weisen Sie nach, dass dann gilt

$$K(X_t) = 3 + \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2} > 3.$$

(Hinweis: Für $Y \sim \text{N}(0, 1)$ gilt $\mathbb{E}[Y^4] = 3$)

- Bestimmen Sie die bedingte Volatilität $\mathbb{E}[\sigma_{t+h}^2 | \mathcal{F}_t]$ für $h \in \mathbb{N}$ und weisen Sie nach, dass sie für $h \rightarrow \infty$ mit der unbedingten Volatilität $\mathbb{E}[\sigma_{t+h}^2]$ übereinstimmt.

- 40) Weisen Sie nach, dass ein schwach stationärer ARCH(1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \sigma_t \varepsilon_t \quad \text{und} \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 \quad \text{für } t \in \mathbb{Z}$$

die Darstellung

$$X_t^2 = \alpha_0 \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_t^2 \cdots \varepsilon_{t-j}^2$$

besitzt.

- 41) Betrachtet wird ein schwach stationärer GARCH(3,2)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \sigma_t \varepsilon_t \quad \text{und} \quad \sigma_t^2 = \alpha_0 + \sum_{j=1}^3 \alpha_j X_{t-j}^2 + \sum_{j=1}^2 \beta_j \sigma_{t-j}^2 \quad \text{für } t \in \mathbb{Z}$$

sowie

$$\varepsilon_t \sim \text{IID}(0, 1) \quad \text{und} \quad \mathbb{E}[X_t^4] < \infty \quad \text{für } t \in \mathbb{Z}.$$

Ermitteln Sie die ARMA(3,2)-Darstellung des quadrierten Prozesses $(X_t^2)_{t \in \mathbb{Z}}$.

42) Betrachtet wird ein IGARCH(1, 1)-Prozess $(X_t)_{t \in \mathbb{Z}}$ mit

$$X_t = \sigma_t \varepsilon_t \quad \text{und} \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{für } t \in \mathbb{Z}$$

sowie

$$\alpha_1 + \beta_1 = 1 \quad \text{und} \quad \varepsilon_t \sim \text{IID}(0, 1) \quad \text{für } t \in \mathbb{Z}.$$

Zeigen Sie, dass

$$\mathbb{E}[\sigma_{t+h}^2 | \mathcal{F}_{t-1}] = h\alpha_0 + \sigma_t^2 \quad \text{für } h \in \mathbb{N}_0$$

gilt, und interpretieren Sie dieses Ergebnis.