Schriftliche Prüfung im Fach

Angewandte Stochastik
Klausur mit Lösungen

gemäß Prüfungsordnung 5
der Deutschen Aktuarvereinigung e. V.

am 16. Mai 2025

Hinweise:

- Als Hilfsmittel für die Klausur sind der Leitfaden und die im Leitfaden angegebene Literatur, Seminarunterlagen (Skript und Aufgabensammlung) sowie frühere Klausuren jeweils in Papierform sowie ein nicht programmierbarer Taschenrechner zugelassen. Hinweise zu den MC-Fragen finden Sie in der jeweiligen Aufgabenstellung.
- Die Gesamtpunktzahl beträgt 180 Punkte. Die Klausur ist bestanden, wenn mindestens 90 Punkte erreicht werden.
- Bitte prüfen Sie die Ihnen vorliegende Prüfungsklausur auf Vollständigkeit. Die Klausur besteht aus 30 Seiten.
- Alle Antworten sind zu begründen und bei Rechenaufgaben muss der Lösungsweg ersichtlich sein.
- Bitte vermeiden Sie bei der Lösungserstellung die nicht zusammenhängende Streuung der Lösungen zu den einzelnen Aufgabenteilen.
- Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers (m/w/d) verzichtet.

Mitglieder der Prüfungskommission:

Prof. Torsten Becker, Dr. Richard Herrmann,
Prof. Christian Heumann, Dr. Stefan Pilz,
Prof. Viktor Sandor, Dr. Dominik Schäfer

Aufgabe 1. *[Lebensdauermodelle] [30 Punkte]*

Es wird ein Bestand von Rentnern untersucht, in dem alle Zugänge bereits bis zum Alter 70 erfolgt sind. Ein Storno der Rentenversicherung ist nicht möglich. Ausgewertet werden die Alter 70 bis 77 bezüglich der Sterblichkeit. Informationen liegen zu dem Bestand von Versicherten und den Todesfälle eines Jahres vor. Die Angaben umfassen sowohl die Anzahlen als auch die individuellen Beträge der Jahresrenten (im folgenden als „Renten" bezeichnet) jeweils für den Bestand und die Todesfälle des Jahres nach Alter zusammengefaßt („Rentensumme", vgl. nachstehende Tabelle). Die q_x bezeichnen die rechnungsmäßigen Ausscheidewahrscheinlichkeiten. Die Altersbestimmung erfolgt als Differenz zwischen Beobachtungsjahr und Geburtsjahr.

		Bestand		erwartete Todesfälle		eingetretene Todesfälle	
Alter x	q_x	Anzahl	Rentensumme	Anzahl	Rentensumme	Anzahl	Rentensumme
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
70	0,145	100	10000	14,5	1450	20	1420
71	0,145	100	10000	14,5	1450	13	1170
72	0,145	100	10000	14,5	1450	16	1216
73	0,145	100	10000	14,5	1450	14	1540
74	0,145	100	10000	14,5	1450	13	1040
75	0,145	100	10000	14,5	1450	15	1395
76	0,145	100	10000	14,5	1450	13	1183
77	0,145	100	10000	14,5	1450	13	936

(a) *[10 Punkte]* Ermitteln Sie mit der Geburtsjahrmethode die Sterblichkeiten für jedes der Alter 70 bis 77 sowohl auf Basis der Anzahlen als auch auf Basis der Rentensummen.

(b) *[20 Punkte]* Überprüfen Sie anhand des Vorzeichentests auf dem 10-Prozent-Niveau, ob die rechnungsmäßigen Wahrscheinlichkeiten q_x nicht mehr angemessen sind und zwar

  (i) *[14 Punkte]* auf Basis der Anzahlen des Bestandes und der Todesfälle,

  (ii) *[4 Punkte]* auf Basis der Rentensummen des Bestandes und der Todesfälle.

  (iii) *[2 Punkte]* Geben Sie eine mögliche Begründung für die unterschiedlichen Ergebnisse unter (i) und (ii) an.

Für die Binomialverteilung gelten folgende Quantile:

Bin(8;0,5)	
k	F(k)
0	0,004
1	0,035
2	0,145
3	0,363
4	0,637
5	0,855
6	0,965
7	0,996
8	1,000

Bin(7;0,5)	
k	F(k)
0	0,008
1	0,063
2	0,227
3	0,500
4	0,773
5	0,938
6	0,992
7	1,000

Bin(6;0,5)	
k	F(k)
0	0,016
1	0,109
2	0,344
3	0,656
4	0,891
5	0,984
6	1,000

Lösung

(a) *[10 Punkte]* Aufgrund der Altersbestimmung ergeben sich bei der Geburtsjahrmethode die Sterbehäufigkeiten für $x = 70, \ldots, 77$

auf Basis der Anzahlen

$$q_x^A = \frac{|T_x|}{|L_x|}$$

auf Basis der Rentensummen

$$q_x^R = \frac{R(T_x)}{R(L_x)}$$

mit

$|T_x|$=Anzahl der Todesfälle im Alter x
$|L_x|$=Anzahl der Lebenden im Alter x
$R(T_x)$=Rentensumme der Todesfälle im Alter x
$R(L_x)$=Rentensumme der Lebenden im Alter x

x	q_x	$\lvert L_x \rvert$	$\lvert T_x \rvert$	q_x^A	$R(L_x)$	$R(T_x)$	q_x^R
70	0,145	100	20	0,200	10000	1420	0,142
71	0,145	100	13	0,130	10000	1170	0,117
72	0,145	100	16	0,160	10000	1216	0,122
73	0,145	100	14	0,140	10000	1540	0,154
74	0,145	100	13	0,130	10000	1040	0,104
75	0,145	100	15	0,150	10000	1395	0,140
76	0,145	100	13	0,130	10000	1183	0,118
77	0,145	100	13	0,130	10000	936	0,094

(b) (i) *[14 Punkte]* Überprüfung anhand der Anzahlen

Formulierung von Hypothese $\mathbf{H_0}$ und Alternative $\mathbf{H_1}$:

$\mathbf{H_0}$: Die rechnungsmäßigen Sterbewahrscheinlichkeiten q_x und die tatsächlichen Sterbehäufigkeiten q_x^A stimmen überein.

$\mathbf{H_1}$: Die rechnungsmäßigen Sterbewahrscheinlichkeiten q_x und die tatsächlichen Sterbehäufigkeiten q_x^A sind verschieden.

Bei Gültigkeit der Hypothese $\mathbf{H_0}$ kann man folgern, dass die beobachteten Wahrscheinlichkeiten (Häufigkeiten) „mal größer und mal kleiner“ als die rechnungsmäßigen Wahrscheinlichkeiten sind. Ist die Differenz zwischen beobachteten und rechnungsmäßig erwarteten Wahrscheinlichkeiten häufig positiv (oder häufig negativ), so deutet dies auf Unterschiede zwischen den rechnungsmäßigen und den tatsächlichen Sterbewahrscheinlichkeiten (Sterbehäufigkeiten) hin.

Teststatistik des Vorzeichentests:

$$T_A = \sum_{x=70}^{77} 1_{\{(q_x - q_x^A) > 0\}}$$

Es ist $n = 8$ und unter der Gültigkeit der Hypothese $\mathbf{H_0}$ ist T_A binomialverteilt mit Wahrscheinlichkeit $\frac{1}{2}$ sowohl für eine positive als auch für eine negative Differenz, d.h. es gilt

$$T_A \sim \mathcal{B}(8, \frac{1}{2})$$

Zu vorgegebenem Signifikanzniveau $\alpha = 0,1$ wird ein Schwellenwert n_α so bestimmt, dass die Hypothese $\mathbf{H_0}$ abgelehnt wird, wenn der Wert der Teststatistik n_α nicht überschreitet oder den Wert $8 - n_\alpha$ überschreitet (zweiseitiger Test).

n_α wird bestimmt aus

$$P(T_A \leq n_\alpha) = \sum_{j=0}^{n_\alpha} \binom{8}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{8-j} = \left(\frac{1}{2}\right)^8 \sum_{j=0}^{n_\alpha} \binom{8}{j} \leq \frac{\alpha}{2}$$

Wegen

$$f(0) + f(1) = 0,004 + 0,031 = 0,035 < \frac{\alpha}{2} = 0,05$$

und

$$f(0) + f(1) + f(2) = 0,004 + 0,031 + 0,109 = 0,145 > \frac{\alpha}{2} = 0,05$$

ist $n_\alpha = 1$ und $8 - n_\alpha = 7$.

Die Hypothese $\mathbf{H_0}$ wird abgelehnt, wenn die Teststatistik T_A in den Ablehnungsbereich $A = \{0, 1, 7, 8\}$ fällt.

Aufgrund der angegebenen Anzahlen von erwarteten und tatsächlichen Todesfällen ist $T_A = 3$, so dass für $\alpha = 10\%$ T_A nicht in den Ablehnungsbereich A fällt und die Hypothese $\mathbf{H_0}$ nicht abgelehnt wird. Aufgrund des durchgeführten Tests besteht kein Anlass zur Änderung der Rechnungsgrundlagen.

(ii) *[4 Punkte]* Überprüfung anhand der Rentensummen der Todesfälle

Anstelle der Anzahl der verstorbenen Rentner des Alters x sollen jetzt die Rentensummen der verstorbenen Rentner im Alter x betrachtet werden. Die Formulierung von Hypothese und Alternative aus (i) ist deshalb anzupassen.

$\mathbf{H_0}$: Die rechnungsmäßigen Sterbewahrscheinlichkeiten q_x und die tatsächlichen Sterbehäufigkeiten q_x^R stimmen überein.

$\mathbf{H_1}$: Die rechnungsmäßigen Sterbewahrscheinlichkeiten q_x und die tatsächlichen Sterbehäufigkeiten q_x^R sind verschieden.

Bei Gültigkeit der Hypothese $\mathbf{H_0}$ kann man folgern, dass die beobachteten Rentensummen der verstorbenen „mal größer und mal kleiner“ als die Rentensummen der rechnungsmäßigen Sterbefälle sind. Ist die Differenz zwischen beobachteten und rechnungsmäßig erwarteten Rentensummen häufig positiv (oder häufig negativ), so deutet dies auf Unterschiede zwischen den rechnungsmäßigen und den tatsächlichen Rentensummen der verstorbenen Rentner hin. Die Teststatistik des Vorzeichentests lautet jetzt

$$T_R = \sum_{x=70}^{77} 1_{\{(q_x - q_x^R) > 0\}}$$

Die Verteilung von T_R ist wie die von T_A: $T_R \sim \mathcal{B}(8, \frac{1}{2})$, so dass ebenfalls $n_\alpha = 1$ und $1 - n_\alpha = 7$ gilt mit unverändertem Ablehnungsbereich $A = \{0, 1, 7, 8\}$.

Aufgrund der angegebenen Rentensummen von erwarteten und tatsächlichen Todesfällen ist $T_R = 1$, so dass für $\alpha = 10\%$ T_R in den Ablehnungsbereich A fällt und die Hypothese $\mathbf{H_0}$ abgelehnt wird. Aufgrund des durchgeführten Tests können die Rechnungsgrundlagen nicht beibehalten werden.

(iii) *[2 Punkte]* Begründung für die unterschiedlichen Ergebnisse in (i) und (ii)

Bei der Überprüfung auf Basis der Anzahlen sind die tatsächlichen Anzahlen der Todesfälle in 3 Altern (70, 72 und 75) höher und in 5 Altern niedriger als erwartet; auf Basis der Rentensummen sind die Rentensummen der tatsächlichen Todesfälle Rentensummen nur im Alter 75 höher und in allen anderen Altern niedriger als die erwarteten Rentensummen. Auf Basis der Anzahlen ergibt sich keine Ablehnung der Hypothese. Die Berücksichtigung der Rentensummen führt zu der Entscheidung, dass die rechnungsmäßig angenommenen Sterblichkeiten die Realität nicht ausreichend abbilden. Ein Grund hierfür kann sein, dass Versicherte mit höheren

Renten eine geringere Sterblichkeit aufweisen als Rentner mit niedrigeren Renten.

Aufgabe 2. *[Deskriptive Statistik, Abhängigkeitsmaße] [30 Punkte]*

(a) *[10 Punkte]* In Abbildung 1 sind vier Scatterplots von Zufallsvektoren $(X_i, Y_i)^\top$, $i = 1, \ldots, 4$ zu sehen. Einer ist der Scatterplot einer Copula, ein weiterer der Scatterplot eines Zufallsvektors mit standardnormalverteilten Rändern.

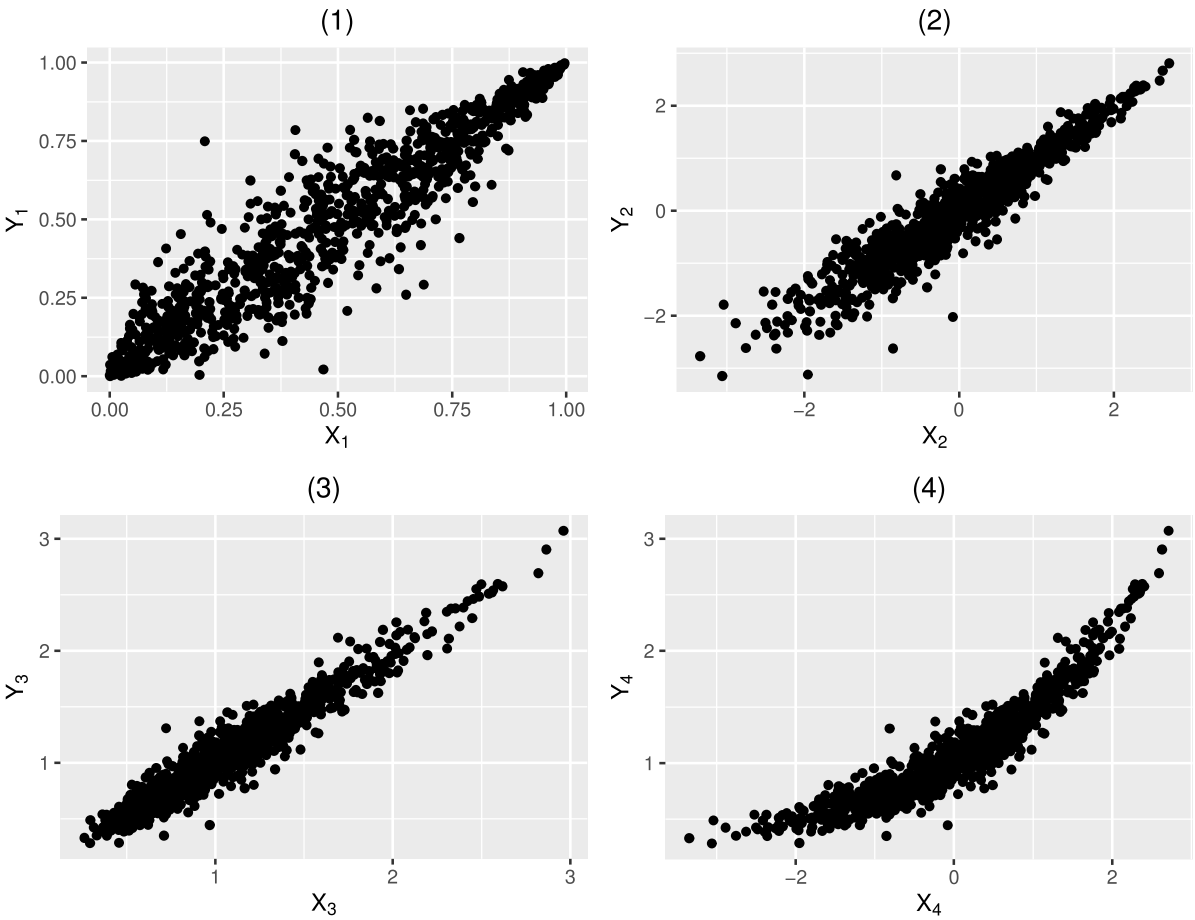


Abbildung 1: Scatterplots von 1000 Simulationen von zweidimensionalen Zufallsvektoren

(i) *[3 Punkte]* Welcher der Scatterplots gehört zur Copula? Begründen Sie ihre Antwort.

(ii) *[4 Punkte]* In welchem Scatterplot sind die Ränder standardnormalverteilt? Begründen Sie ihre Antwort.

(iii) *[3 Punkte]* Ist der Zufallsvektor in (a) (ii) zweidimensional normalverteilt? Begründen Sie ihre Antwort.

(b) *[8 Punkte]* In Abbildung 2 sind die Boxplots und QQ-Normal-Plots der Ränder zu einem der Scatterplots (1)-(4) Abbildung 1 gegeben.

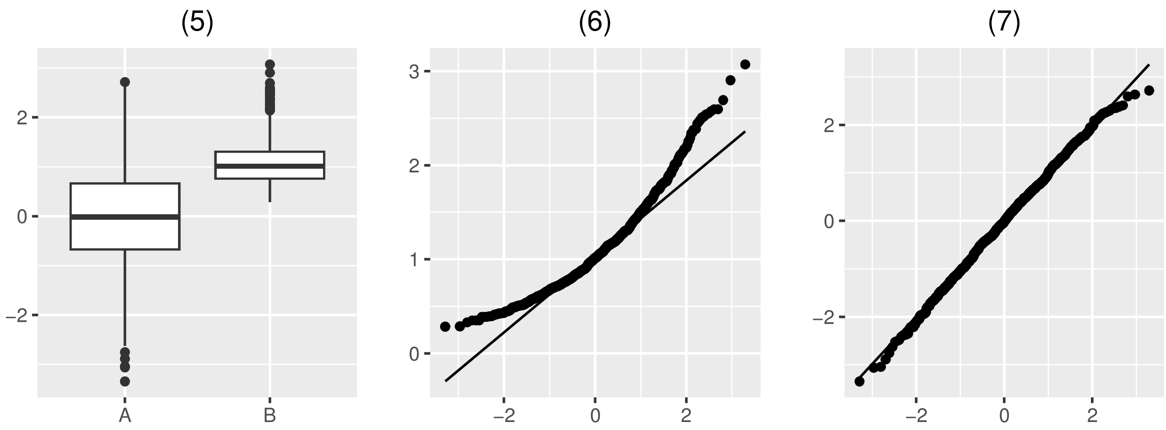


Abbildung 2: Boxplots und QQ-Normalplots der zwei Randverteilungen eines der Scatterplots (1)-(4) in Abbildung 1

(i) *[4 Punkte]* Welcher Zufallsvektor $(X_k, Y_k)^\top$ ist in (5)-(7) dargestellt, welche Zufallsvariable ist A bzw. B in (5)? Welche Zufallsvariable ist in (6) und welche in (7) dargestellt? Begründen Sie ihre Antworten.

(ii) *[4 Punkte]* Was schließen Sie aus den Plots in Abbildung 2 bezüglich der Verteilung der beiden Zufallsvariablen? Begründen Sie ihre Antwort.

(c) *[8 Punkte]* Sei $\theta > 0$ und $\varphi : (0,1] \longrightarrow [0,\infty)$, $\varphi(t) = \frac{1}{\theta}\left(\frac{1}{t^\theta} - 1\right)$.

(i) *[4 Punkte]* Beweisen Sie, dass φ der Erzeuger einer archimedischen Copula ist.

(ii) *[4 Punkte]* Die von φ erzeugte Copula ist die Clayton-Copula (ein Beweis ist nicht verlangt). Beweisen Sie, dass für die untere Tailabhängigkeit

$$\lambda_L = \frac{1}{2^{1/\theta}}$$

gilt.

(d) *[4 Punkte]* Konstruieren Sie einen Zufallsvektor $(X, Y)^\top$ mit $X, Y \sim N(0,1)$ und der unteren Tailabhängigkeit $\lambda_L = \frac{1}{2}$. Ist $(X, Y)^\top$ zweidimensional normalverteilt?

Lösung Aufgabe 2 [(a,i) 3, (a,ii) 4, (a, iii) 3, (b,i) 4, (b,ii) 4, (c,i) 4, (c,ii) 4, (d) 4]

(a) (i) Die Copula ist in (1) abgebildet. Das erkennt man daran, dass X_1, Y_1 nur Werte in $[0, 1]$ annimmt. Bei (2)-(4) werden auch Werte angenommen, die größer als 1 sind.

(ii) Nur in (2) bewegen sich die Werte von X_2, Y_2 in $[-3, 3]$ und die Ränder sind standardnormalverteilt. In (1) handelt es sich um die Copula, also $X_1, Y_1 \sim U[0, 1]$, in (3) nehmen X_3, Y_3 nur positive Werte an, in (4) nimmt Y_4 nur positive Werte an, ist also nicht standardnormalvereilt.

(iii) In (2) liegt eine obere Tailabhängigkeit $\lambda_U > 0$ vor, da die Punktewolke in der rechten oberen Ecke zusammenläuft. Daher ist $(X_2, Y_2)^\top$ nicht zweidimensional normalverteilt, da sonst $\lambda_U = 0$ wäre.

(b) (i) Aufgrund der angenommenen Werte, handelt es sich um (4) und es gilt $A = X_4$, $B = Y_4$. In (6) ist der QQ-Plot von Y_4 und in (7) von X_4 zu sehen. Beim QQ-Normal-Plot werden auf der y-Achse die Stichprobenwerte aufgetragen, auf der x-Achse die Quantile von $N(0, 1)$. Da Y_4 nur positive Werte annimmt ist also (6) der QQ-Plot von Y_4.

(ii) Da die Ausgleichsgerade in (7) die Identitätsgerade ist, kann man von $X_4 \sim N(0, 1)$ ausgehen. Dafür spricht auch der symmetrische Boxplot (5), links. In (6) sind die Abweichungen der Punkte von der Ausgleichsgeraden sehr groß, auch der Boxplot in (5) ist unsymmetrisch. Daher ist Y_4 nicht normalverteilt, weitere Aussagen zur Verteilung sind nicht möglich.

(c) (i) Zu zeigen ist, dass φ streng monoton fallend und konvex ist, $\varphi(1) = 0$ und $\lim_{t\downarrow 0} \varphi(t) = \infty$ gilt.

$$\begin{aligned}
\lim_{t\downarrow 0} \varphi(t) &= \frac{1}{\theta}\left(\lim_{t\to 0} \frac{1}{t^\theta} - 1\right) = \infty \\
\varphi(1) &= \frac{1}{\theta}\left(1^{-\theta} - 1\right) = 0 \\
\varphi'(t) &= \frac{1}{\theta}\left(-\theta t^{-\theta-1}\right) = -t^{-\theta-1} < 0 \\
\varphi''(t) &= (\theta + 1)t^{-\theta-2} > 0
\end{aligned}$$

(ii) Mit der Clayton-Copula $C^{\mathrm{Cl}}(u, v) = \left(\frac{1}{u^{1/\theta}} + \frac{1}{v^{1/\theta}} - 1\right)^{-1/\theta}$ und $\lambda_L = \lim_{t\downarrow 0} \dfrac{C(t, t)}{t}$ folgt

$$\lim_{t\downarrow 0}\frac{C(t,t)}{t}=\lim_{t\downarrow 0}\frac{1}{t}\left(2\left(\frac{1}{t}\right)^{\theta}-1\right)^{-\frac{1}{\theta}}=\lim_{t\downarrow 0}\frac{1}{t}\left(\frac{2-t^{\theta}}{t^{\theta}}\right)^{-1/\theta}=\lim_{t\downarrow 0}\frac{1}{t}\cdot\left(\frac{t^{\theta}}{2-t^{\theta}}\right)^{1/\theta}$$
$$=\lim_{t\downarrow 0}\frac{1}{t}\frac{t}{\left(2-t^{\theta}\right)^{1/\theta}}=\lim_{t\downarrow 0}\frac{1}{\left(2-t^{\theta}\right)^{1/\theta}}=\frac{1}{2^{1/\theta}}.$$

(d) Wählt man bei der Clayton-Copula $\theta=1$, dann ergibt sich in (c) (ii) $\lambda_L=\frac{1}{2}$. Sei Φ die Verteilungsfunktion von $N(0,1)$. Dann ist

$$F(x,y)=C^{CL}_{\frac{1}{2}}(\Phi(x),\Phi(y)),\ (x,y)\in\mathbb{R}^2$$

mit dem Satz von Sklar die Verteilungsfunktion eines Zufallsvektors mit $N(0,1)$ verteilten Rändern. Die Tailwahrscheinlichkeit hängt nur von der Copula ab, es gilt also $\lambda_U=\frac{1}{2}$. Wäre $(X,Y)^{\top}$ zweidimensional normalverteilt, dann wäre $\lambda_U=0$.

Aufgabe 3. *[36 Punkte]* Induktive Statistik

Bei einer zufällig ausgewählten Teilmenge von $n = 5000$ Versicherten eines Bestands wurde ein neues Versicherungsprodukt beworben. Anschließend wurde evaluiert, ob die Werbeaktion erfolgreich war ("Ja") oder nicht ("Nein"). Sei π_i die Wahrscheinlichkeit, dass die versicherte Person i das neue Produkt erwirbt. Die individuellen Wahrscheinlichkeiten sollen modelliert werden. Dazu stehen die Merkmale `Alter` (metrisch, in Jahren) und `Region` (drei Kategorien: `A, B, C`) zur Verfügung.

(a) *[2 Punkte]* Definieren Sie eine geeignete Zielvariable Y und deren Verteilung für ein verallgemeinertes lineares Regressionsmodell.

(b) *[2 Punkte]* Stellen Sie die Likelihood als Funktion der π_i, $i = 1, \dots, n$ auf.

(c) *[16 Punkte]* Die Ausgabe des Regressionsmodells sieht folgendermaßen aus:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.1578     0.3439    -9.2   <2e-16 ***
alter         0.0383          ?     5.1    3e-07 ***
regionB      -1.3379     0.1279   -10.5   <2e-16 ***
regionC       1.3520     0.0869    15.6   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 5045.3  on 4999  degrees of freedom
Residual deviance: 4528.4  on 4996  degrees of freedom
AIC: 4536
```

(i) *[4 Punkte]* Geben Sie die Regressionsgleichung mit kanonischer Linkfunktion an, welche zur gegebenen Ausgabe führt. Geben Sie dazu genau an, welche Kodierung für die Region verwendet wird. Geben Sie auch die Responsefunktion an.

(ii) *[2 Punkte]* Geben Sie den kodierten Vektor der Kovariablen x_i für eine 35–jährige versicherte Person in Region `B` an.

(iii) *[2 Punkte]* Berechnen Sie den Standardfehler (`Std. Error`) für das Merkmal `Alter`.

(iv) *[1 Punkt]* Interpretieren Sie den geschätzten Koeffizienten für das Merkmal `Alter` hinsichtlich seines quantitativen Einflusses auf die Zielvariable (eine mögliche Interpretation genügt).

(v) *[2 Punkte]* Interpretieren Sie die geschätzten Koeffizienten für das Merkmal `Region`.

(vi) *[3 Punkte]* Welche Wahrscheinlichkeit schätzt man für eine 40–jährige versicherte Person aus Region `A`, das neue Produkt zu erwerben?

(vii) *[2 Punkte]* Das metrische Alter ist linear in den Prädiktor aufgenommen worden. Nennen Sie zwei weitere Möglichkeiten, das Alter nichtlinear in den Prädiktor aufzunehmen.

(d) *[16 Punkte]* Es wird nun ein weiteres Modell geschätzt:

```
   Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -2.1700     0.4269   -5.08  3.7e-07 ***
alter               0.0165     0.0094    1.76   0.0789 .
regionB            -0.6295     1.1947   -0.53   0.5983
regionC            -2.6186     0.8170   -3.21   0.0014 **
alter:regionB      -0.0154     0.0263   -0.59   0.5570
alter:regionC       0.0882     0.0180    4.89  1.0e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 5045.3  on 4999  degrees of freedom
Residual deviance: 4500.9  on 4994  degrees of freedom
AIC: 4513
```

Wir betrachten im Folgenden die Interaktion von `Alter` und `Region`.

(i) *[4 Punkte]* Formulieren Sie die entsprechenden Hypothesen H_0 und H_1 (verbal und mathematisch), um zu überprüfen, ob die Interaktion von `Alter` und `Region` einen Einfluss auf die Wahrscheinlichkeit hat, dass die versicherte Person das Produkt erwirbt.

(ii) *[2 Punkte]* Schlagen Sie eine einen Test vor, für den Sie mit den gegebenen R-Ausgaben die Teststatistik berechnen können (*kurze Begründung*).

(iii) *[2 Punkte]* Geben Sie die Zahl der Freiheitsgrade (Df) des entsprechenden Tests (*kurze Begründung*) an.

(iv) *[1 Punkt]* Welche (approximative) Verteilung hat die Teststatistik?

(v) *[5 Punkte]* Berechnen Sie die Teststatistik!

(vi) *[2 Punkte]* Geben Sie die Testentscheidung zum Niveau $\alpha = 0.05$ an (*kurze Begründung*).

Hinweis: Folgende 95%-Quantile der $\chi^2(df, 0.95)$-Verteilung sind gegeben:

$$\chi^2_{1;0.95} = 3.84 \quad \chi^2_{2;0.95} = 5.99 \quad \chi^2_{4994;0.95} = 5159.52$$

.

Lösung

(a) *[2 Punkte]* Eine geeignete Zielvariable Y ist binär: $Y_i = 1$, falls Versicherter i das Produkt erwirbt, $Y_i = 0$ sonst. Als Verteilung wählt man die Bernoulli-Verteilung.

(b) *[2 Punkte]* Likelihood als Funktion der π_i:

$$L(\pi_1, \dots, \pi_{5000}) = \prod_{i=1}^{5000} \pi_i^{y_i}(1-\pi_i)^{(1-y_i)} .$$

(c) *[16 Punkte]*

(i) *[4 Punkte]* Logistisches Regressionsmodell mit `Alter` und `Region` als Haupteffekte. `Region` ist kategorial, man wählt die Dummy-Kodierung. Da die `Region` drei Kategorien hat, müssen 2 Dummy–Variablen verwendet werden. In der Ausgabe wird `Region A` als Referenzkategorie verwendet.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \text{Alter} + \beta_2 \text{Region}_B + \beta_3 \text{Region}_C .$$

Die Responsefunktion lautet:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 \text{Alter} + \beta_2 \text{Region}_B + \beta_3 \text{Region}_C)}{1 + \exp(\beta_0 + \beta_1 \text{Alter} + \beta_2 \text{Region}_B + \beta_3 \text{Region}_C)}$$

(ii) *[2 Punkte]* `Region A` ist auch in der R-Ausgabe die Referenzkategorie. Eine 35–jährige versicherte Person in `Region B` hat deshalb den kodierten Vektor:

Intercept	Alter	Dummy 1 (Region B)	Dummy 2 (Region C)
1	35	1	0

oder kurz $x_i = (1, 35, 1, 0)$.

(iii) *[2 Punkte]* Der geschätzte Standardfehler ist:

$$\text{Std. Error} = \frac{Estimate_{Alter}}{z_{Alter}} = 0.0383/5.1 = \mathbf{0.0075098} .$$

(iv) *[1 Punkt]* 3 mögliche Interpretationen:

- Erhöht sich das Alter um 1 Jahr, so erhöht sich die logarithmierte Chance (für Kauf des Produkts vs. Nichtkauf),

$$\log\left(\frac{\pi}{1-\pi}\right) ,$$

additiv um $\beta_1 = 0.0383$.

- Erhöht sich das Alter um 1 Jahr, erhöht sich die Chance (Odds, für Kauf des Produkts vs. Nichtkauf) (multiplikativ) um den Faktor $\exp(\beta_1) = 1.039$.
- Der Odds Ratio ist $\exp(\beta_1) = 1.039$, d.h. die Chance einer $(x+1)$–Jahre alten versicherten Person (für Erwerb des Produkts vs. Nichterwerb) ist um den Faktor $\exp(\beta_1) = 1.039$ höher als die Chance einer x–Jahre alte versicherten Person.

(v) *[2 Punkte]* Die Interpretation ist immer bezogen auf die Referenzkategorie (`Region A`) zu machen:

Die Chance für Kauf vs. Nichtkauf ist in Region B um den Faktor

$$\exp(-1.3379) = 0.2624$$

niedriger als in Region A.

Die Chance für Kauf vs. Nichtkauf ist in Region C dagegen um den Faktor

$$\exp(1.3520) = 3.8651$$

höher als in Region A.

(vi) *[3 Punkte]* Prädiktor $\eta_i = -3.1578 + 40 \cdot 0.0383 = -1.6258$. Damit:

$$P(Y_i = 1) = \frac{1}{1+\exp(1.6258)} = \frac{\exp(-1.6258)}{1+\exp(-1.6258)} = 0.1644 \; .$$

Die Kaufwahrscheinlichkeit wird auf etwa 0.1644 geschätzt.

(vii) *[2 Punkte]* Folgende Alternativen sind möglich: Aufnahme von Transformationen des Alters, also log(Alter) oder Alter^2 oder Generalisierte Additive Modelle (GAM) oder feinere Kategorisierung.

(d) *[16 Punkte]*

(i) *[4 Punkte]* Verbal: H_0: Es besteht keine Interaktion zwischen Alter und Region. H_1 : Es besteht eine Interaktion zwischen Alter und Region. Mathematisch:

$$H_0 : \beta_{Alter,Region_B} = \beta_{Alter,Region_C} = 0$$

versus

$$H_1 : \beta_{Alter,Region_B} \neq 0 \quad \text{oder} \quad \beta_{Alter,Region_C} \neq 0$$

(oder im nicht ausschließenden Sinn, d.h mindestens einer der Koeffizienten ist von Null verschieden).

(ii) *[2 Punkte]* Likelihood-Quotienten-Test (LQT). Die Teststatistik kann mit Hilfe der beiden AIC Werte der Modelle ohne und mit Interaktion berechnet werden.

(iii) *[2 Punkte]* Die Zahl der Freiheitsgrade ist 2. Begründung: Alter ist metrisch und Region hat 3 Kategorien. Damit ist die Zahl der benötigten Dummy-Variablen für die Interaktion $\cdot(3-1)=2$. Dies ist die Differenz der Anzahl der Parameter im Modell mit Interaktion im Vergleich zum Modell ohne Interaktion.

(iv) *[1 Punkt]* Die Teststatistik ist asymptotisch χ^2-verteilt mit 2 Freiheitsgraden.

(v) *[5 Punkte]*

- $AIC = -2 \cdot \text{loglikelihood} + 2 \cdot \text{Anzahl der geschätzten Parameter}$ und damit

$$2 \cdot \text{loglikelihood} = 2 \cdot \text{Anzahl der geschätzten Parameter} - AIC$$

- Modell ohne Interaktion: $AIC = 4536$, Modell mit Interaktion: $AIC = 4513$

- Damit ist die Teststatistik LQT (Zweimal die Differenz der Loglikelihood des Modells mit Interaktion und des Modells ohne Interaktion):

$$\text{LQT} = 2 \cdot 6 - 4513 - (2 \cdot 4 - 4536) = 27$$

(vi) *[2 Punkte]* Die Testentscheidung ist: H_0 ablehnen (Interaktion), da $27 > \chi^2_{2;0.95} = 5.99$

Aufgabe 4. *[24 Punkte]* Zeitreihenanalyse

(a) *[3 Punkte]* Geben Sie an, welche Annahmen für Zeitreihenmodelle hinsichtlich

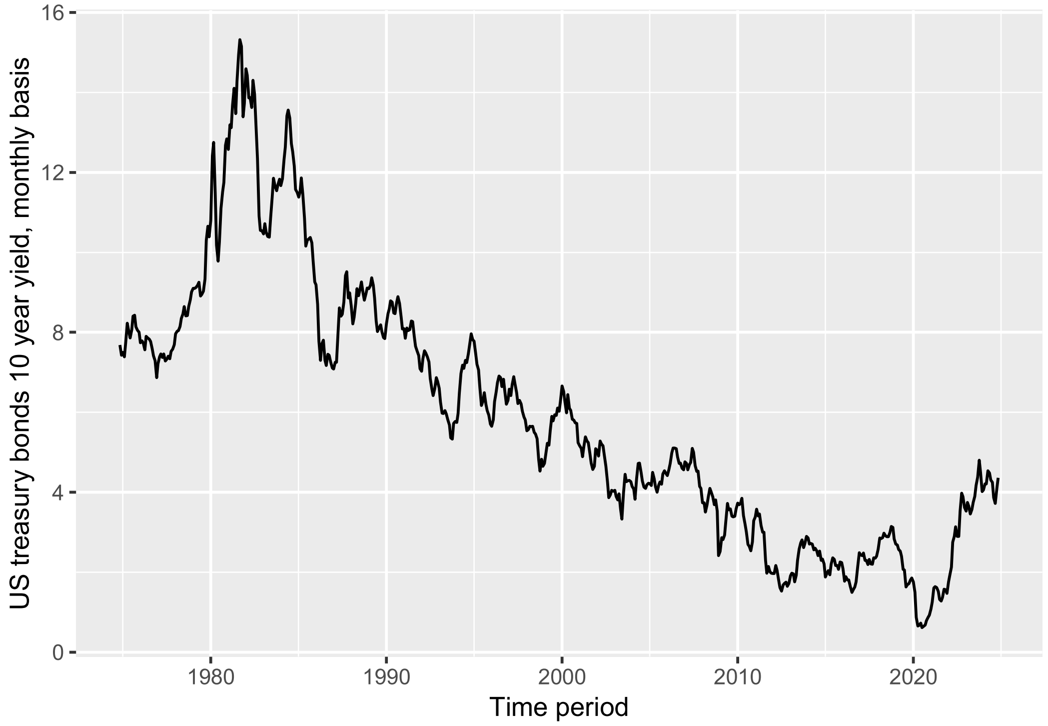
A) des Abstands zwischen zwei zeitlich aufeinanderfolgenden Messungen,
B) des Skalenniveaus der gemessenen Variable,
C) der Korreliertheit oder Unkorreliertheit zwischen den Messungen

getroffen werden.

(b) *[6 Punkte]* Entscheiden Sie, ob die folgenden Aussagen richtig oder falsch sind. Begründen Sie Ihre Entscheidungen kurz und im Falle von falschen Aussagen geben Sie die zudem die richtige Aussage an. *Für jede richtige [falsche] Antwort gibt es zwei [Null] Punkte*.

A) Bei AR(p) Modellen kann man zukünftige Werte schätzen und der Parameter p gibt die Anzahl der latenten Komponenten des AR-Modells an.
B) Der random walk $y_t = y_{t-1} + u_t$, u_t iid N(0,1), ist stationär.
C) Die partielle Autokorrelation ist ein Spezialfall der Autokorrelation.

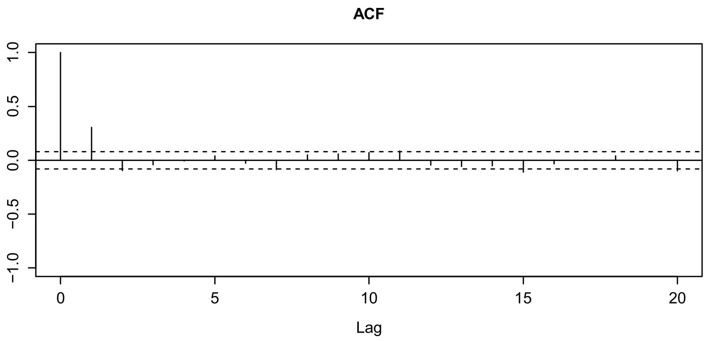
(c) (i) *[3 Punkte]* Die folgende Grafik enthält auf Monatsbasis die 10-jährige Rendite (in Prozent) von U.S. Treasury Securities (Quelle: Federal Reserve Bank of St. Louis). Nennen Sie die drei Charakteristika, die geeignet sind den Verlauf der Zeitreihe zu beschreiben!(Abschätzung anhand der Grafik).

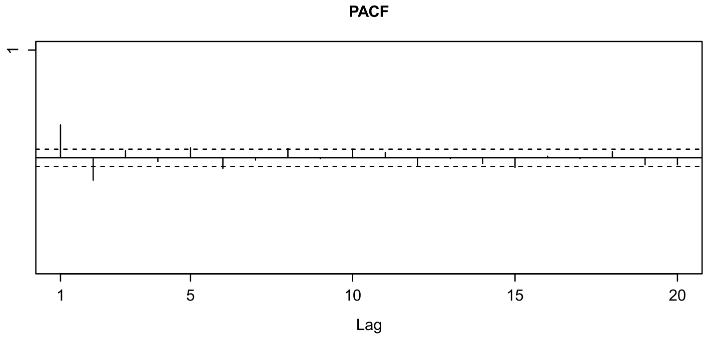


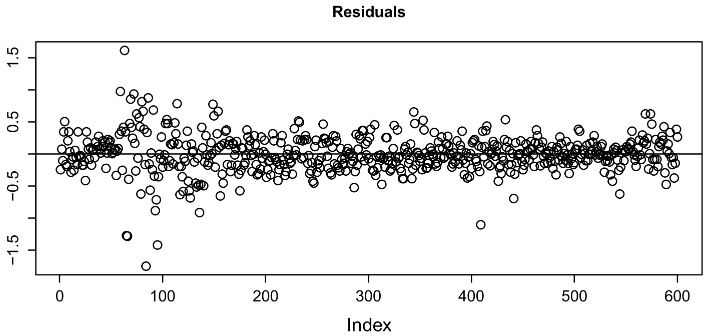
(ii) *[1 Punkt]* Nennen Sie einen Grund warum die Zeitreihe nicht stationär ist.

(iii) *[4 Punkte]* Welche Transformationen bzw. Bereinigungen schlagen Sie vor um eine stationäre Zeitreihe zu erreichen? (zwei Vorschläge mit kurzer Begründung)

(iv) *[3 Punkte]* Für die bereinigte Zeitreihe wurde die Autokorrelationsfunktion bzw. die partielle Autokorrelationsfunktion bzw. die Residuen berechnet und visualisiert. Interpretieren Sie die Grafiken (jeweils 1 Punkt). Bitte beachten Sie, dass die PACF erst ab *lag=1* definiert ist.







(v) *[2 Punkte]* Sie sollen für die bereinigte Zeitreihe ein Modell schätzen. Machen Sie einen Vorschlag und begründen ihn kurz!

(vi) *[2 Punkte]* Wie beurteilen Sie den Einsatz von diesem Zeitreihenmodell (d.h. auf Basis der bereinigten Zeitreihe) für die kurzfristige Prognose (mit kurzer Begründung)?

Lösung

(a) *[3 Punkte]*

A) Die Werte sind zu äquidistanten - gleichabständigen Zeitpunkten vorhanden.
B) Die Werte sind Realisierungen einer metrisch skalierten Variable.
C) Die Werte sind korreliert (abhängig würde auch akzeptiert werden).

(b) *[6 Punkte]*

A) falsch, p gibt die Anzahl der vergangenen Beobachtungen (Parameter) an, die zur Modellierung der gegenwärtigen Beobachtung herangezogen werden ebenso zulässig: AR Modell hat keine latenten Komponenten, sondern das MA-Modell
B) falsch, da die Varianz als Funktion in t zunimmt.
C) falsch, es handelt sich um verwandte Konzepte der Abhängigkeit: bedingte versus marginale Betrachtung.

(c) *[3 Punkte]*

(i) i. Trend(s) vorhanden.

ii. Volatilität vermutlich konstant.

iii. Saisonalität nicht sichtbar bzw. nicht vorhanden.

iv. auch akzeptabel: gleichabständige Beobachtungen bzw. metrisch skaliertes Merkmal. Die Bedingungen sind erfüllt.

(ii) *[1 Punkte]* Die Zeitreihe ist nicht stationär; Gründe: Trend, lokale Pattern (ein Grund genügt).

(iii) *[4 Punkte]* Transformationen

i. Trendbereinigung durch Differenzen erster Ordnung.

ii. Gleitende Durchschnitte.

(iv) *[3 Punkte]* Interpretation

i. Autokorrelationsfunktion (ACF). Die ACF misst (unter Annahme der schwachen Stationarität der Zeitreihe) die paarweisen Korrelationen zwischen zwei Beobachtungen zu den Lags (Zeitabständen) 0, 1, 2, 3, usw. Bei Lag 0 hat die ACF immer den Wert Eins. In der Grafik ist dies gut sichtbar. Die gestrichelten waagrechten Linien um die Null-Linie sind die Annahmebereiche für einen entsprechenden Test mit H_0: Autokorrelation zu Lag l ist 0. Lediglich bei Lag 1 ist die Autokorrelation deutlich signifikant von 0 verschieden, danach schwanken die Autokorrelationen in diesem Annahmebereich.

ii. Partielle Autokorrelationsfunktion: Abgesehen von Lag=1, 2 streuen die Werte nahe dem Wert Null (Lag 0 wird hier in der Grafik nicht abgetragen, die Funktion beginnt bei Lag 1), d.h. sie liegen wieder im Annahmebereich des entsprechendn Tests mit H_0 : PACF(l)= 0. Nur für Lag 1 und 2 ist die PACF(1) und PACF(2) signifikant von Null verschieden. Die partielle ACF gibt die um Zwischenwerte bereinigte Korrelation an.

iii. Residuen-Plot: Die Residuen streuen um den Wert Null.

(v) *[2 Punkte]* Die Werte der partiellen Autokorrelationsfunktion suggerieren ein Modell mit autoregressiver Komponente (z.B. AR(2), und die Werte der Autokorrelationen legen zudem nahe eine moving average Komponente mitzuverwenden, z.BN. MA(1). Dies wäre dann ein ARMA Modell.

(vi) *[2 Punkte]* Aufgrund der Struktur in der ACF, PACF und Residuen sollte die Verwendung eines Zeitreihenmodelles hilfreich sein für die Modellierung des bedingten Erwartungswertes bzw. der Abhängigkeitsstruktur und daher die kurzfristige Prognose verbessern.

Aufgabe 5. *[Credibility] [30 Punkte]*

(a) *[15 Punkte]* In einem Bayes'schen Credibility-Modell sei ϑ Realisierung eines zufälligen Strukturparameters $\Theta: \Omega \to \{0,1\}$, dessen a-priori-Verteilung die Zähldichte $f_\Theta(\vartheta) := P(\Theta = \vartheta)$ mit $f_\Theta(0) = f_\Theta(1) = \frac{1}{2}$ besitze. Die Schadenhöhe X folge bei gegebenem Strukturparameter $\Theta = \vartheta$ einer Verteilung mit Dichte

$$f_{X|\Theta=\vartheta}(x) = 1 + \vartheta - 2\vartheta x \qquad \text{für } 0 \leq x \leq 1.$$

Gegeben seien zudem die (bei gegebenem $\Theta = \vartheta$) unabhängigen Beobachtungen $X_1, \dots, X_n$ von $n = 5$ Schäden mit den Realisierungen

i	1	2	3	4	5
x_i	0,12	0,03	0,80	0,48	0,18

Berechnen Sie die *allgemeine* Credibility-Prämie. Gehen Sie dazu in folgenden Teilschritten vor und runden Sie Zwischenergebnisse auf 4 Nachkommastellen.

(i) *[5 Punkte]* Berechnen Sie den Wert der gemeinsamen Dichte $f(x_1, \dots, x_n, \vartheta)$ von Schäden und Strukturparameter für $\vartheta = 0$ und $\vartheta = 1$ sowie die gegebenen Beobachtungen $x_1, \dots, x_n$.

(Kontrollergebnisse: $f(x_1, \dots, x_n, 0) = 0{,}5$ und $f(x_1, \dots, x_n, 1) = 1{,}1647$)

(ii) *[5 Punkte]* Berechnen Sie den Wert der Dichte $f_{\Theta|X_1=x_1,\dots,X_n=x_n}(\vartheta)$ der a-posteriori-Verteilung für $\vartheta = 0$ und $\vartheta = 1$ sowie die gegebenen Beobachtungen $x_1, \dots, x_n$.

(Kontrollergebnisse: $f_{\Theta|X_1=x_1,\dots,X_n=x_n}(0) = 0{,}3004$ und $f_{\Theta|X_1=x_1,\dots,X_n=x_n}(1) = 0{,}6996$)

(iii) *[5 Punkte]* Ermitteln Sie den Wert der allgemeinen Credibility-Prämie H^*.

(Hinweis: Ohne Beweis können Sie $H(\vartheta) := E(X|\Theta = \vartheta) = \frac{1}{2} - \frac{\vartheta}{6}$ verwenden.)

(b) *[15 Punkte]* Als a-priori-Verteilung des Strukturparameters Θ wird nun eine Gleichverteilung auf dem Intervall $[0,1]$ angesetzt. Die Schadenhöhe X folge bei gegebenem Strukturparameter $\Theta = \vartheta$ weiterhin der Verteilung mit Dichte

$$f_{X|\Theta=\vartheta}(x) = 1 + \vartheta - 2\vartheta x \qquad \text{für } 0 \leq x \leq 1.$$

Berechnen Sie den Wert der ***linearisierten*** Credibility-Prämie H^{**} für die in (a) gegebenen Beobachtungen $x_1, \ldots, x_n$. Gehen Sie in folgenden Teilschritten vor und runden Sie Zwischenergebnisse auf 4 Nachkommastellen.

(Hinweis: Ohne Beweis können Sie verwenden, dass $H(\Theta) = \frac{1}{2} - \frac{\Theta}{6}$ und $Var(\mathrm{X}|\Theta) = \frac{1}{12} - \frac{\Theta^2}{36}$ ist. Außerdem können Sie verwenden, dass für eine auf dem Intervall $[0,1]$ gleichverteilte Zufallsvariable U gilt: $E(U) = \frac{1}{2}$, $E(U^2) = \frac{1}{3}$ und $Var(U) = \frac{1}{12}$.)

(i) *[6 Punkte]* Berechnen Sie $E(X) = E\big(H(\Theta)\big)$ und $Var\big(H(\Theta)\big)$.

(Kontrollergebnisse: $E(X) = 0{,}4167$ und $Var\big(H(\Theta)\big) = 0{,}0023$)

(ii) *[3 Punkte]* Berechnen Sie $E\big(Var(\mathrm{X}|\Theta)\big)$.

(Kontrollergebnis: $E\big(Var(\mathrm{X}|\Theta)\big) = 0{,}0741$)

(iii) *[6 Punkte]* Berechnen Sie den Wert des Credibility-Faktors z und die linearisierte Credibility-Prämie H^{**}.

Lösung:

(a) (i) Für die gemeinsame Dichte gilt

$$f(x_1, \dots, x_n, \vartheta) = f_\Theta(\vartheta) \cdot \prod_{i=1}^{n} f_{X|\Theta=\vartheta}(x_i).$$

Speziell für $\vartheta = 0$ ist

$$f(x_1, \dots, x_n, 0) = \frac{1}{2} \cdot \prod_{i=1}^{n} 1 = \frac{1}{2}.$$

Für $\vartheta = 1$ ergibt sich

$$f(x_1, \dots, x_n, 1) = \frac{1}{2} \cdot \prod_{i=1}^{n} (2 - 2x_i) = \frac{2^n}{2} \cdot \prod_{i=1}^{n} (1 - x_i) = 2^4 \cdot \ 0{,}072795 = 1{,}1647.$$

(a) (ii) Für die gemeinsame Dichte gilt

$$f_{\Theta|X_1=x_1,\dots,X_n=x_n}(\vartheta) = \frac{f(x_1, \dots, x_n, \vartheta)}{f(x_1, \dots, x_n, 0) + f(x_1, \dots, x_n, 1)}.$$

Somit ist

$$f_{\Theta|X_1=x_1,\dots,X_n=x_n}(0) = \frac{f(x_1, \dots, x_n, 0)}{f(x_1, \dots, x_n, 0) + f(x_1, \dots, x_n, 1)} = \frac{0{,}5}{0{,}5 + 1{,}1647} = 0{,}3004$$

und

$$f_{\Theta|X_1=x_1,\dots,X_n=x_n}(1) = \frac{f(x_1, \dots, x_n, 1)}{f(x_1, \dots, x_n, 0) + f(x_1, \dots, x_n, 1)} = \frac{1{,}1647}{0{,}5 + 1{,}1647} = 0{,}6996.$$

(a) (iii) Gemäß Hinweis ist

$$H(\vartheta) = E(X|\Theta = \vartheta) = \frac{1}{2} - \frac{\vartheta}{6}.$$

Somit ist

$$H^* = \sum_{\vartheta=0}^{1} H(\vartheta) \cdot f_{\Theta|X_1=x_1,\dots,X_n=x_n}(\vartheta) = \frac{1}{2} \cdot 0{,}3004 + \left(\frac{1}{2} - \frac{1}{6}\right) \cdot 0{,}6996 = 0{,}3834.$$

(b) (i) Zunächst ist mit $H(\Theta) = \frac{1}{2} - \frac{\Theta}{6}$

$$E(X) = E\big(H(\Theta)\big) = \frac{1}{2} - \frac{\mathrm{E}(\Theta)}{6} = \frac{1}{2} - \frac{1}{6} \cdot \frac{1}{2} = \frac{5}{12} = 0{,}4167.$$

Zur Berechnung des Credibility-Faktors benötigt man zudem

$$Var\big(H(\Theta)\big) = Var\left(\frac{1}{2} - \frac{\Theta}{6}\right) = \frac{Var(\Theta)}{6^2} = \frac{1}{36} \cdot \frac{1}{12} = \frac{1}{432} = 0{,}0023.$$

(b) (ii) Gemäß Hinweis ist

$$Var(\mathrm{X}|\Theta) = \frac{1}{12} - \frac{\Theta^2}{36},$$

so dass

$$E\big(Var(\mathrm{X}|\Theta)\big) = \frac{1}{12} - \frac{E(\Theta^2)}{36} = \frac{1}{12} - \frac{1}{36} \cdot \frac{1}{3} = \frac{2}{27} = 0{,}0741.$$

(b) (iii) Damit ergibt sich der Credibility-Faktor

$$z = \frac{Var\big(H(\Theta)\big)}{\frac{1}{n}E\big(Var(\mathrm{X}|\Theta)\big) + Var\big(H(\Theta)\big)} = \frac{0{,}0023}{\frac{0{,}0741}{5} + 0{,}0023} = 0{,}1343.$$

Die linearisierte Credibility-Prämie ist

$$H^* = z \cdot \overline{X} + (1 - z) \cdot E(X) = 0{,}1343 \cdot 0{,}3220 + (1 - 0{,}1343) \cdot 0{,}4167 = 0{,}4040.$$

Aufgabe 6. *[SDGL und ihre Simulation] [30 Punkte]*

(a) *[2 Punkte]* Geben Sie für das Black Scholes Modell eines Aktienkursprozesses $(S_t)_{t\geq 0}$ die beschreibende SDGL sowie die explizite Lösung an.

(b) *[12 Punkte]* Ein Aktienkursprozess $(S_t)_{t\geq 0}$ werde beschrieben durch die SDGL

$$dS_t = \mu S_t \, dt + \sqrt{v(t)} S_t \, dW_t, \qquad S_0 = s_0$$

mit einer zeitabhängigen Funktion $v(t)$, die die Volatilität $\sqrt{v(t)}$ des Kursprozesses bestimmt. Geben Sie eine Lösung dieser SDGL an, wenn $\mu = 0,4$ und $v(t) = 1 + \cos(t)$ (auftretende Riemann-Integrale sind vollständig auszurechnen). Bestimmen Sie daraus die Verteilung von S_t/S_0 (Angabe der Verteilungsfamilie und der Werte der Verteilungsparameter).

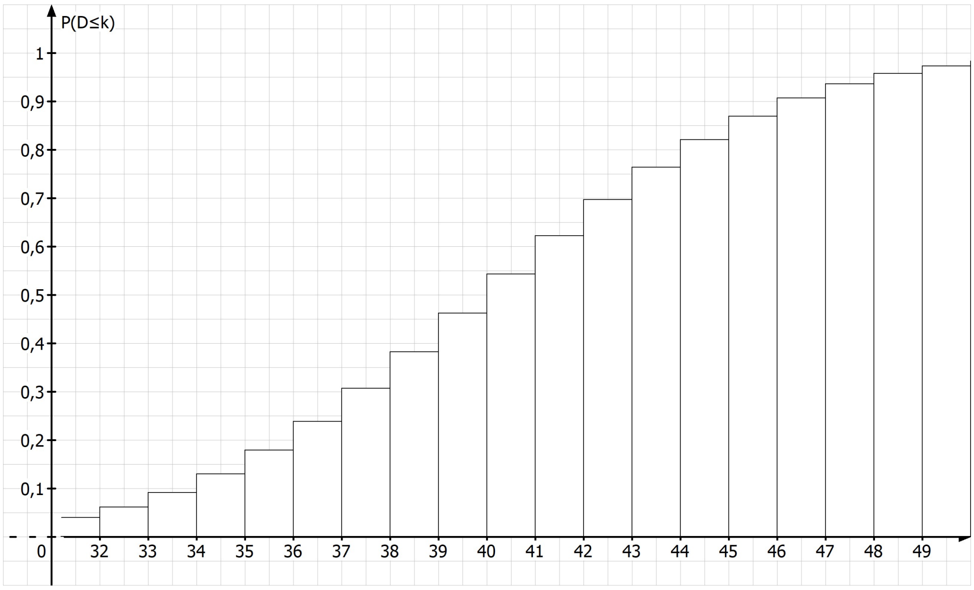
(c) *[6 Punkte]* Die Volatilität $\sqrt{v(t)}$ eines Kursprozesses kann auch durch einen stochastischen Prozess modelliert werden; aus der Funktion $v(t)$ in (b) wird dann der Prozess $(V_t)_{t\geq 0}$. Für eine konkrete Aktie werde dieser beschrieben durch

$$dV_t = 0,2\,(0,3 - V_t)\,dt + 0,5\sqrt{V_t}\,dW_t \qquad (*)$$

Simulieren Sie einen Pfad dieses Prozesses mit der Euler-Methode; die Simulationszeitpunkte seien $t = 0$, $t = 0,1$ und $t = 0,3$, der Startwert des Pfades $v_0 = 0,25$. Verwenden Sie die zwei unabhängigen N(0,1)-Zahlen $z_1 = 0,388$ und $z_2 = -0,928$.

(d) *[2 Punkte]* Die Euler-Methode für eine SDGL vom Typ $(*)$ ist ungeeignet, da das Verfahren unter Umständen irgendwann mit einer Fehlermeldung abbricht. Geben Sie eine plausible Vermutung für den Grund eines solchen Abbruchs an.

(e) *[4 Punkte]* Die tatsächliche Verteilung von $V_{0,1}$ aus Aufgabenteil (c) ist eine χ^2_D-Verteilung, wobei der Freiheitsgrad D selbst wieder eine (diskrete) Zufallsvariable ist. Eine exakte Simulation von $V_{0,1}$ besteht also aus zwei Teilsimulationen, erst wird D simuliert und bei gegebenem $D = d$ anschließend χ^2_d. In folgender Abbildung ist ein Ausschnitt der Verteilungsfunktion von D abgebildet.



(i) *[2 Punkte]* Simulieren Sie einen Wert von D unter Verwendung der Abbildung und der U(0,1)-Zufallszahl $u = 0,34$. Beschreiben Sie Ihr Vorgehen.

(ii) *[2 Punkte]* Für welche $u \in (0,1)$ liefert das Verfahren aus (i) einen Wert für D größer als 40? (ungefähre Zahlenangaben mit zwei Nachkommastellen.)

(f) *[4 Punkte]* Für eine Pfaderzeugung soll nun die exakte Verteilung von $V_{0,1}$ simuliert werden. In einer der Simulationsrunden habe sich für D der Wert 3 realisiert. Simulieren Sie für diesen Fall einen Wert der χ^2_D-Verteilung unter Verwendung der drei unabhängigen N(0,1)-Zahlen $z_1 = 0,108$, $z_2 = -0,344$ und $z_3 = -0,138$.

Lösungsvorschlag

(a) SDGL:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$$

Explizite Lösung:

$$S_t = S_0 \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right) t + \sigma W_t\right)$$

(b) In Verallgemeinerung des Black Scholes Modells hat die SDGL

$$dS_t = \mu(t)\, S_t\, dt + \sigma(t)\, S_t\, dW_t$$

die Lösung

$$S_t = S_0 \exp\left(\int_0^t \left[\mu(s) - \tfrac{1}{2}\sigma^2(s)\right] ds + \int_0^t \sigma(s)\, dW_s\right)$$

Mit den Vorgaben der Aufgabenstellung folgt

$$\begin{aligned} S_t &= S_0 \exp\left(\int_0^t \left[0,4 - \tfrac{1}{2}(1+\cos(s))\right] ds + \int_0^t \sqrt{1+\cos(s)}\, dW_s\right) \\ &= S_0 \exp\left(\left[0,4\, s - \tfrac{1}{2}(s+\sin(s))\right]_{s=0}^{s=t} + \int_0^t \sqrt{1+\cos(s)}\, dW_s\right) \\ &= S_0 \exp\left(-0,1\, t - \tfrac{1}{2}\sin(t) + \int_0^t \sqrt{1+\cos(s)}\, dW_s\right). \end{aligned}$$

Nun ist

$$\int_0^t \sqrt{1+\cos(s)}\, dW_s \sim \mathrm{N}\left(0,\ \int_0^t (1+\cos(s))\, ds\right) = \mathrm{N}\,(0,\ t+\sin(t)).$$

Daraus ergibt sich für die Verteilung von S_t/S_0

$$S_t/S_0 \sim \mathrm{LN}\left(-0,1\, t - \tfrac{1}{2}\sin(t),\ t+\sin(t)\right).$$

(c) Das Euler-Verfahren liefert

$$\widehat{v}_{0,1} = v_0 + 0,2\,(0,3 - v_0)\cdot(0,1-0) + 0,5\sqrt{v_0}\cdot\sqrt{0,1-0}\cdot z_1 = 0,282$$

sowie

$$\widehat{v}_{0,3} = \widehat{v}_{0,1} + 0,2\,(0,3 - \widehat{v}_{0,1})\cdot(0,3-0,1) + 0,5\sqrt{\widehat{v}_{0,1}}\cdot\sqrt{0,3-0,1}\cdot z_2 = 0,172.$$

(d) Es ist möglich, dass im Verlauf des Verfahrens ein $\widehat{v}_{t_i}$ negativ wird (denn es wird ja nicht die exakte Verteilung simuliert, die tatsächlich nur nichtnegative Werte liefert). Dann bricht das Verfahren im nachfolgenden Schritt ab, da unter der Wurzel eine negative Zahl steht.

(e) (i) Inversionsmethode: Ist F die Verteilungsfunktion, dann ist $F^{\leftarrow}(u)$ ein Sample aus F, wenn u eine U(0,1)-Zufallszahl ist.
Zur Ermittlung der Pseudoinversen zeichnet man eine parallele Linie zur x-Achse durch den Punkt $y = 0,34$ ein (0,34 ist in diesem Fall das u). Diese durchdringt den Graphen der Verteilungsfunktion an der Stelle $x = 38$ (vor $x = 38$ ist die Linie über dem Graphen von F, ab $x = 38$ ist sie darunter). Daher ist $F^{\leftarrow}(0,34) = 38$.

(ii) Dies erfüllen alle Werte auf der y-Achse, deren Pseudoinverse größer 40 ist. Die sind laut Grafik (mit zwei Nachkommastellen Genauigkeit) alle $u \in (0,55; 1)$.

(f) Eine χ^2_3-Verteilung ist per Definition die Verteilung einer Zufallsvariablen X der Form

$$X = Z_1^2 + Z_2^2 + Z_3^2,$$

wobei Z_1, Z_2, Z_3 unabhängig und N(0,1)-verteilt sind. Mit den vorgegebenen Werten ergibt sich als Samplewert

$$0,108^2 + (-0,344)^2 + (-0,138)^2 = 0,149.$$