Schriftliche Prüfung im Grundwissen

Angewandte Stochastik

Klausur mit Lösungen

gemäß Prüfungsordnung 4
der Deutschen Aktuarvereinigung e. V.

am 19. Juni 2020

Hinweise:

- Als Hilfsmittel sind Seminarunterlagen und Aufgaben in Papierform, handschriftliche Notizen im Rahmen der normalen Schulung sowie ein nicht programmierbarer Taschenrechner zugelassen.
- 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 mit Lösungen besteht aus 28 Seiten.
- Alle Antworten sind zu begründen und bei Rechenaufgaben muss der Lösungsweg ersichtlich sein.

Mitglieder der Prüfungskommission:

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

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

Ein Lebensversicherungsunternehmen LVU möchte einen Teil seines Versichertenbestandes auf einen anderen Lebensversicherer übertragen. Hierzu wird der bisherige Gesamtbestand V in die unabhängigen Teilbestände A (verbleibt bei LVU) und B (wird übertragen) aufgeteilt. Als Aktuar des Lebensversicherungsunternehmens LVU werden Sie um Antworten zu folgenden Fragen gebeten:

(a) [20 Punkte]
Sind die biometrischen Risiken in den Beständen A und B gleich?

Als Unterlagen haben Sie für beide Bestände und die letzten 5 Jahre folgende Informationen zur Verfügung:

(A) Versichertenbestand mit individuellen Angaben (Geschlecht, Geburtsdatum, Versicherungsbeginn, Versicherungssumme, Deckungskapital) jeweils zum Beginn und zum Ende des Geschäftsjahres,

(B) Liste der Abgänge im jeweiligen Geschäftsjahr mit den Angaben unter (A) und zusätzlich nur den Abgangsgrund Tod oder Ablauf der Versicherung (Storno wird vernachlässigt).

(i) [8 Punkte]

(A) [3 Punkte] Welche Methode zur Bestimmung der Sterblichkeiten q_{Ax} bzw. q_{Bx} der beiden Bestände ist hier nicht anwendbar? Begründen Sie Ihre Antwort.

(B) [5 Punkte] Stellen Sie eine der möglichen Methoden dar, mit der Sie die Sterblichkeiten q_{Ax} bzw. q_{Bx} der beiden Bestände ermitteln können.

Gehen Sie davon aus, dass die Altersbestimmung nach der Kalenderjahrsmethode (Alter x = Kalenderjahr - Geburtsjahr) erfolgt.

(ii) [12 Punkte] Führen Sie zur Beurteilung der Sterblichkeitsunterschiede zwischen den Beständen A und B für die folgenden beobachteten Sterblichkeiten den Iterationstest mit $\alpha = 5\%$ durch.

Alter x	30	31	32	33	34	35	36	37	38	39	40	41	42
q_{Ax}	1,0%	1,2%	0,5%	1,0%	1,5%	1,2%	1,1%	0,9%	1,4%	1,3%	1,6%	1,7%	1,6%
q_{Bx}	1,1%	2,0%	1,0%	1,1%	1,4%	1,1%	0,9%	1,3%	1,5%	1,4%	1,8%	2,0%	1,8%

Die Dichte $f(t)$ der Binomial-Verteilung lautet für p=0,5 und n=10, 11, 12 (Werte in %):

t	0	1	2	3	4	5	6	7	8	9	10	11	12
n=10	0,10	0,98	4,39	11,72	20,51	24,61	20,51	11,72	4,39	0,98	0,10		
n=11	0,05	0,54	2,69	8,06	16,11	22,56	22,56	16,11	8,06	2,69	0,54	0,05	
n=12	0,02	0,29	1,61	5,37	12,08	19,34	22,56	19,34	12,08	5,37	1,61	0,29	0,02

(b) [10 Punkte]

Gehen Sie davon aus, dass die Schäden in den Beständen A, B und V normalverteilt sind mit Erwartungswert μ_w und Varianz σ_w^2 (w=A, B, V). Zum Ausgleich des Schwankungsrisikos soll zusätzlich zum Erwartungswert der Schäden auf Grundlage des Value-at-Risk zur Sicherheitswahrscheinlichkeit 0,95 ein Sicherheitskapital gebildet werden.

Die Standardabweichung für Bestand V beträgt σ_V = 10 Mio und für Bestand A σ_A= 6 Mio; das 95%-Quantil der Standardnormalverteilung beträgt $u_{0,95}$ = 1,645.

Welches Sicherheitskapital ist dann nach der Übertragung für den Bestand B erforderlich?

Lösung

(a) (i) (A) Die Verweildauermethode ist nicht anwendbar. Zur Anwendung der Verweildauermethode fehlen in den Bestandsdaten der Todeszeitpunkt und das Ablaufdatum.

(B) Folgende Methoden kommen in Betracht:

- Geburtsjahrmethode
- Sterbejahrmethode
- Sterbeziffernverfahren

(nur 1 Methode wird verlangt)

Bezeichne

L eine Personengesamtheit, unter Risiko stehend

T Tote aus L

$q = \frac{|T|}{|L|}$ relative Sterbehäufigkeit = rohe Sterbewahrscheinlichkeit

$L' = L \setminus T$ = überlebende (Teilgesamtheit der Lebenden)

B = Beobachtungszeitraum [1.1.2004 , 1.1.2009)

G = Geburtsjahr

$L_x(B, G)$ die Personen aus der Personengesamtheit, die im Beobachtungszeitraum B das x. Lebensjahr vollenden (könnten) und im Geburtsjahr G geboren wurden,

$T_x(B, G)$ die Personen aus $L_x(B, G)$, die im Beobachtungszeitraum sterben.

Geburtsjahrmethode

Aufgrund der Kalenderjahrmethode zur Altersbestimmung wird jedem Versicherten unabhängig vom unterjährigen exakten Todesdatum in Abhängigkeit vom Geburtsjahr das Alter x = Kalenderjahr – Geburtsjahr zugeordnet.

Die rohe Sterbewahrscheinlichkeit nach der Geburtsjahrmethode ist definiert durch:

$$q_x^G = \frac{|T_x(B, G)|}{|L_x(B, G)|}$$

Sterbejahrmethode

Bei der Sterbejahrmethode werden sämtliche Todesfälle eines Alters x berücksichtigt. Bei Verwendung der Kalenderjahrmethode zur Altersbestimmung ist die Sterbejahrmethode mit der Geburtsjahrmethode identisch.

Die rohe Sterbewahrscheinlichkeit nach der Sterbejahrmethode ist definiert durch:

$$q_x^S = \frac{|T_x(B,\ G)|}{|L_x(B,\ G)|}$$

Sterbeziffernverfahren

Im Gegensatz zur Geburtsjahr- und zur Sterbejahrmethode werden beim so genannten Sterbeziffernverfahren die Bestandsveränderungen innerhalb des Beobachtungszeitraums näherungsweise durch die Durchschnittsbildung der Personengesamtheit am Anfang und am Ende jedes Jahres der Beobachtungsperiode berücksichtigt. Zu- und Abgänge, die innerhalb eines Beobachtungsjahres stattfinden, können aber auch hier nicht miteinbezogen werden.

Die Sterbeziffer lautet

$$k_x = \frac{|T_x(B,G)|}{\sum_{i=1}^n \frac{1}{2}\left(\left|L_x^A(B_i,G)\right| + \left|L_x^E(B_i,G)\right|\right)},$$

wobei L_x^A und L_x^E den Bestand der x-Jährigen am Anfang bzw. am Ende eines Jahres innerhalb des Beobachtungszeitraums bezeichnen.

k_x kann jedoch nicht als Wert für die relative Sterbehäufigkeit angesetzt werden: Denn da in L_x^A und L_x^E nicht die x-Jährigen erfasst sind, die bereits vor diesen Stichtagen gestorben sind, würde die Sterbeziffer eine zu hohe Sterblichkeit anzeigen.

Stattdessen definiert man

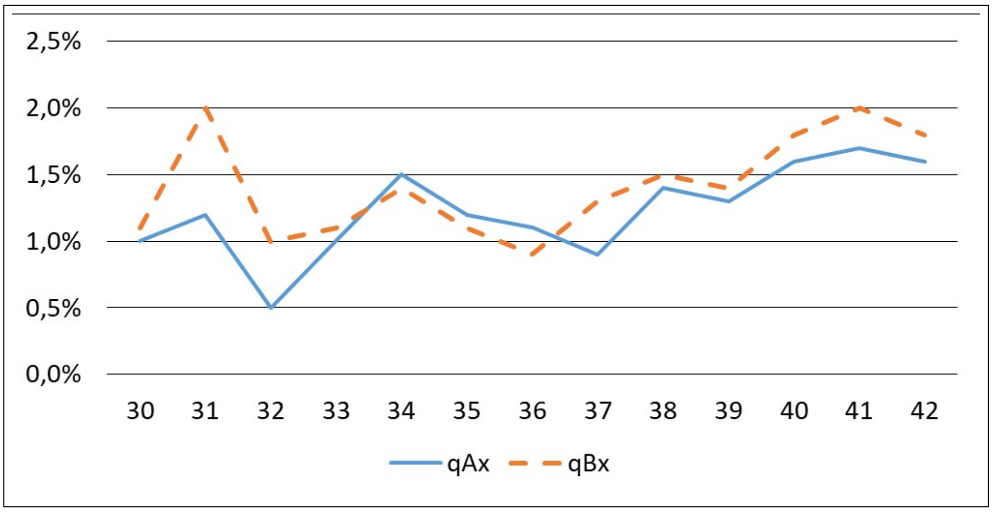
$$q_x^z = \frac{2k_x}{2+k_x}\ .$$

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

$\mathbf{H_0}$: Die Sterbewahrscheinlichkeiten stimmen in beiden Versichertenbeständen überein.

$\mathbf{H_1}$: Die Sterbewahrscheinlichkeiten in beiden Versichertenbeständen sind verschieden.

Signifikanzniveau $\alpha = 0,05$



Bei Gültigkeit der Hypothese, dass in beiden Beständen die Sterbewahrscheinlichkeiten übereinstimmen, kann man folgern, dass die beobachteten Wahrscheinlichkeiten des einen Bestandes „mal größer und mal kleiner“ als die des anderen sind, d.h. dass häufige Vorzeichenwechsel auftreten. Wenige Vorzeichenwechsel deuten auf Unterschiede zwischen den Sterblichkeiten der Bestände hin.

Teststatistik:

$$T = \sum_{x=31}^{42} 1_{\{Sign(q_{Ax}-q_{Bx}) \neq Sign(q_{Ax-1}-q_{Bx-1})\}}$$

Es ist $n = 12$ und unter der Gültigkeit der Nullhypothese H_0 ist T binomialverteilt mit Wahrscheinlichkeit $\frac{1}{2}$ für jeden Vorzeichenwechsel, d.h. es gilt

$$T \sim \mathcal{B}(12, \frac{1}{2})$$

Zu vorgegebenem Signifikanzniveau α wird ein Schwellenwert n_α so bestimmt, dass die Nullhypothese abgelehnt wird, wenn der Wert der Teststatistik n_α unterschreitet (einseitiger Test).

Dabei wird n_α bestimmt aus

$$P(T < n_\alpha) = \sum_{j=0}^{n_\alpha - 1} \binom{12}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{12-j} = \left(\frac{1}{2}\right)^{12} \sum_{j=0}^{n_\alpha} \binom{12}{j} \leq \alpha$$

Wegen

$$f(0)+f(1)+f(2)=0,0002+0,0029+0,0161=0,0192<\alpha$$

und

$$f(0)+f(1)+f(2)+f(3)=0,0002+0,0029+0,0161+0,0537=0,0729>\alpha$$

ist $n_\alpha=3$.

Aufgrund der angegebenen Sterblichkeiten ist $T=2$, so dass für $\alpha=5\%$ T in den Ablehnungsbereich fällt und die Hypothese $\mathbf{H_0}$ abgelehnt wird.

(b) Bezeichne S_w den Schaden im Bestand w für w=V, A, B.

S_A und S_B sind stochastisch unabhängig und es gilt $S_V=S_A+S_B$. Für die Verteilungen gilt $S_w\sim\mathcal{N}(\mu_w,\sigma_w^2)$.

Für den Value at Risk gilt dann mit Φ als Verteilungsfunktion der Standard-Normalverteilung

$$\begin{aligned}VaR_{0,05}(S_w)&=\mu_w+\sigma_w\cdot\Phi^{-1}(1-0,05)\\&=\mu_w+\sigma_w\cdot u_{0,95}\end{aligned}$$

Das über den Erwartungswert hinausgehende Sicherheitskapital $S_w^{0,95}$ ist dann

$$S_w^{0,95}=VaR_{0,05}(S_w)-\mu_w=\sigma_w\cdot u_{0,95}$$

Für die Differenz $S_B=S_V-S_A$ der unabhängigen normalverteilten Zufallsvariablen S_V und S_A gilt

$$S_B=S_V-S_A\ \sim\mathcal{N}(\mu_V-\mu_A,\sigma_V^2-\sigma_A^2)$$

$$VaR_{0,05}(S_B)=\mu_V-\mu_A+\sqrt{\sigma_V^2-\sigma_A^2}\ \cdot u_{0,95}$$

und

$$\begin{aligned}S_B^{0,95}=VaR_{0,05}(S_B)-(\mu_V-\mu_A)&=\sqrt{\sigma_V^2-\sigma_A^2}\ \cdot u_{0,95}\\&=\sqrt{100-36}\ \cdot 1,645\\&=8\ \cdot 1,645\\&=13,16\end{aligned}$$

Das Sicherheitskapital für den übertragenen Bestand B beträgt 13,16 Mio.

Aufgabe 2. *[Abhängigkeiten und Copulas] [30 Punkte]*

(a) *[8 Punkte]* Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind. Schreiben Sie Ihre Antworten auf die Lösungsblätter, die Sie abgeben. Eine Begründung ist nicht erforderlich.

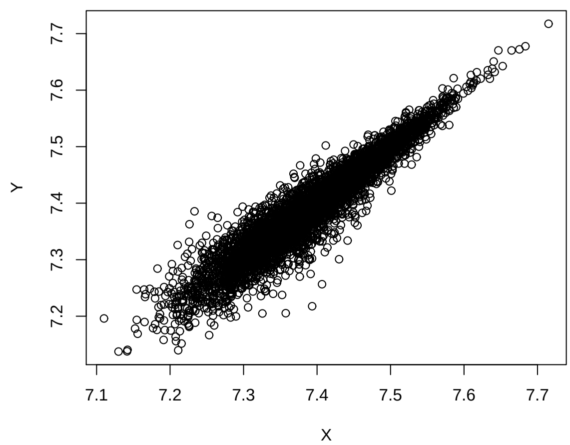
Für jede richtige Antwort gibt es 2 Punkte, für jede falsche Antwort 0 Punkte.

A Ist die gemeinsame Verteilungsfunktion von (X_1, X_2) eine Copula, dann sind X_1, X_2 gleichverteilt.

B Gegeben sei eine bivariate Stichprobe $(x_i, y_i) \in (0, \infty)^2$, $i = 1, \ldots, n$ eines stetig verteilten Zufallsvektors $(X, Y) : \Omega \longrightarrow (0, \infty)^2$. Die empirischen Korrelationskoeffizienten nach Pearson von (x_i, y_i) und der logarithmierten Stichprobe $(\ln(x_i), \ln(y_i))$ sind im Allgemeinen ungleich.

C Gegeben sei eine bivariate Stichprobe $(x_i, y_i) \in (0, \infty)^2$, $i = 1, \ldots, n$ eines stetig verteilten Zufallsvektors $(X, Y) : \Omega \longrightarrow (0, \infty)^2$. Die Rangkorrelationen nach Spearman von (x_i, y_i) und der logarithmierten Stichprobe $(\ln(x_i), \ln(y_i))$ sind im Allgemeinen ungleich.

D In der unteren Abbildung ist der Scatterplot von 5000 simulierten Schäden (X, Y) dargestellt.



Die Copula von (X, Y) ist die Gauß-Copula.

(b) *[6 Punkte]* Gegeben sei die Clayton-Copula $C : [0, 1]^2 \longrightarrow [0, 1]$, $C(u, v) = (u^{-2} + v^{-2} - 1)^{-\frac{1}{2}}$.

(i) *[4 Punkte]* Der Zufallsvektor (X, Y) besitze die Copula C, $X \sim \mathcal{LN}(0, 1)$, $Y \sim \mathcal{E}(1)$. Bestimmen Sie $P(X \leq 1, Y \leq \ln(2))$.

(ii) *[2 Punkte]* Sind X, Y unabhängig?

(c) *[16 Punkte]* Sei $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ zweidimensional normalverteilt mit $\rho \in [-1, 1]$. Sei $Y_i := \exp(X_i)$ und $\rho(Y_1, Y_2)$ die Korrelation von Y_1, Y_2.

(i) *[2 Punkte]* Welches ist die Copula von $(Y_1, Y_2)^\top$? Begründen Sie Ihre Antwort.

(ii) *[4 Punkte]* Begründen Sie, dass $Y_i \sim \mathcal{LN}(0, 1)$ und $Y_1 Y_2 \sim \mathcal{LN}(0, 2 + 2\rho)$ gilt.

(iii) *[6 Punkte]* Beweisen Sie, dass

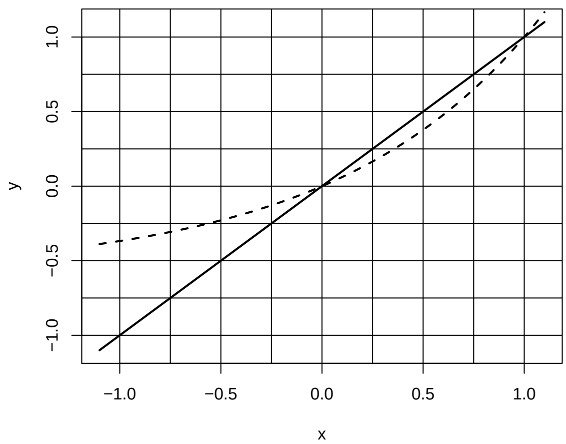
$$\rho(Y_1, Y_2) = \frac{e^\rho - 1}{e - 1}$$

gilt.

(iv) *[2 Punkte]* Sei $f : [-1, 1] \longrightarrow \mathbb{R}$, $f(t) = \dfrac{e^t - 1}{e - 1}$. Es gelten die folgenden Ungleichungen (vgl. Abbildung, ein Beweis ist nicht notwendig):

$$f(t) < t,\ t \in (0, 1) \text{ und} \tag{1.1}$$

$$f(t) \geq -\frac{1}{e},\ t \in [-1, 1]. \tag{1.2}$$



Verwenden Sie (iii), (1.1) und (1.2) um die folgenden Ungleichungen zu beweisen:

$$\rho(Y_1, Y_2) < \rho \qquad \text{falls } \rho \in (0, 1), \tag{1.3}$$

$$-\frac{1}{e} \leq \rho(Y_1, Y_2) \qquad \text{für alle } \rho \in [-1, 1]. \tag{1.4}$$

(v) *[2 Punkte]* Gibt es Zufallsvariablen Z_1, Z_2 mit $Z_1, Z_2 \sim \mathcal{LN}(0, 1)$, $(\ln(Z_1), \ln(Z_2))$ bivariat normalverteilt und $\rho(Z_1, Z_2) = -0,5$? Begründen Sie Ihre Antwort.

Lösung

(a) *[8 Punkte]* Eine Begründung war nicht erforderlich. Dennoch wird sie hier gebeben, damit die Antworten nachvollzogen werden können.

(A) Wahr.
Jede Copula ist die gemeinsame Verteilung zweier Zufallsvariablen mit auf $[0,1]$ gleichverteilten Komponenten.

(B) Wahr.
Nur linear affine Transformationen verändern die Pearson-Korrelation nicht.

(C) Falsch.
Die Ränge ändern sich bei einer Transformation mit einer streng monoton wachsenden Funktion nicht, also auch nicht der Rangkorrelationskoeffizient, da dieser nur die Ränge betrachtet.

(D) Falsch.
Es liegt wohl eine positive untere Tailabhängigkeit vor, außerdem ist die Punktwolke unsymmetrisch. Bei der Gauß-Copula liegt Symmetrie vor, die Tailabhängigkeiten verschwinden.

(b) *[6 Punkte]*

(i) *[4 Punkte]* Mit dem Satz von Sklar gilt

$$P(X \le x, Y \le y) = C(F_X(x), F_Y(y)) = C(P(X \le x), P(Y \le y)).$$

In diesem konkreten Fall gilt

$$P(X \le 1) = P(\ln(X) \le \ln(1)) = \Phi(0) = 0,5$$
$$P(Y \le \ln(2)) = 1 - e^{-\ln(2)} = 0,5$$
$$P(X \le 1, Y \le \ln(2)) = C\left(\frac{1}{2}, \frac{1}{2}\right) = (4+4-1)^{-\frac{1}{2}} = \frac{\sqrt{7}}{7} \approx 0,38.$$

(ii) *[2 Punkte]* X, Y sind nicht unabhängig, da

$$P(X \le 1)P(Y \le \ln(2)) = \frac{1}{4} \ne \frac{\sqrt{7}}{7}.$$

(c) *[16 Punkte]*

(i) *[2 Punkte]* Die Copula von $(X_1, X_2)^\top$ ist die Gauß-Copula C_ρ^{Ga}. Da ln streng monoton wächst, ändert sich die Copula nicht, also ist die Copula von $(Y_1, Y_2)^\top$ ebenfalls C_ρ^{Ga}.

(ii) *[4 Punkte]* Wegen $\ln(Y_i) = X_i \sim \mathcal{N}(0,1)$ gilt $Y_i \sim \mathcal{LN}(0,1)$ nach Definition. Ferner gilt

$$\ln(Y_1 Y_2) = X_1 + X_2.$$

Die Zufallsvariable $X_1 + X_2$ ist als Linearkombination der Komponenten normalverteilt, wegen der Eigenschaften der mehrdimensionalen Normalverteilung. Damit ist $Y_1 Y_2$ lognormal-verteilt. Die Parameter ergeben sich aus

$$\begin{aligned} Var(Y_1 + Y_2) &= Var(Y_1) + 2Cov(Y_1, Y_2) + Var(Y_2) = 2 + 2\rho. \\ E(Y_1 + Y_2) &= E(Y_1) + E(Y_2) = 0. \end{aligned}$$

(iii) *[6 Punkte]* Wegen $Y_1 Y_2 \sim \mathcal{LN}(0, 2+2\rho)$, $Y_i \sim \mathcal{LN}(0,1)$ gilt mit den Formeln für Erwartungswert und Varianz der Lognormalverteilung

$$\begin{aligned} E(Y_1 Y_2) &= e^{0+(2+2\rho)/2} = e^{1+\rho} \\ Cov(Y_1, Y_2) &= E(Y_1 Y_2) - E(Y_1)E(Y_2) = e^{1+\rho} - e^{\frac{1}{2}+\frac{1}{2}} = e^{1+\rho} - e = e(e^\rho - 1) \\ Var(Y_i) &= e^{2\cdot 0+1}\left(e^{1^2} - 1\right) = e\,(e-1) \\ \rho(Y_1, Y_2) &= \frac{e(e^\rho - 1)}{e\,(e-1)} = \frac{e^\rho - 1}{e-1}. \end{aligned}$$

(iv) *[2 Punkte]* Es gilt

$$\rho(Y_1, Y_2) = f(\rho)$$

und die Behauptungen folgen direkt aus den beiden Ungleichungen (1.1) und (1.2).

(v) *[2 Punkte]* Angenommen, es gibt solche Z_1, Z_2. Wegen (1.4) gilt dann $\rho(Z_1, Z_2) \geq -\frac{1}{e} \approx -0,37$ im Widerspruch zu $\rho(Z_1, Z_2) = -0.5$.

Nachrichtlich: Beweis von (1.1) und (1.2):

Mit der ersten und zweiten Ableitung gilt für alle $t \in [-1, 1]$

$$\begin{aligned} f'(t) &= \frac{e^t}{e-1} > 0 \\ f''(t) &= \frac{e^t}{e-1} > 0 \end{aligned}$$

also ist f monoton wachsend und strikt konvex. Wegen $f(0) = 0$, $f(1) = 1$ liegt f im Intervall $(0, 1)$ unter Identität, da diese die Verbindungsgerade der Punkte $(0, 0)$ und $(1, 1)$ ist. Wegen der strengen Monotonie folgt für alle $t \in [-1, 1]$

$$f(t) \geq f(-1) = \frac{e^{-1} - 1}{e - 1} = \frac{1 - e}{e(e - 1)} = -\frac{1}{e}.$$

Aufgabe 3. *[36 Punkte]* Induktive Statistik

Hinweis: bei allen Ergebnissen genügen 2 Nachkommastellen, als Dezimaltrennzeichen wird der Punkt verwendet.

Zur Untersuchung des Einflusses verschiedener Merkmale auf die Wahrscheinlichkeit für Storno von Versicherungspolicen wurden Daten gesammelt. Die Zielvariable gibt an, ob die entsprechende Police i, $i = 1, \dots, n$ storniert wurde ($Y_i = 1$) oder nicht ($Y_i = 0$).

Ein Modell untersucht die Merkmale `Alter` der versicherten Person und `Zahlungsart` der Prämie (4 Kategorien: `jährlich`, `monatlich`, `halbjährlich`, `quartalsweise`) hinsichtlich ihres Einflusses auf die Wahrscheinlichkeit für Storno. Ein geeignetes Regressionsmodell liefert folgende Ausgabe:

```
Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                -0.915764   0.081819 -11.193  < 2e-16 ***
Alter                      -0.020167   0.001782 -11.317  < 2e-16 ***
Zahlungsartmonatlich       -0.378323   0.073704  -5.133 2.85e-07 ***
Zahlungsarthalbjährlich    -0.141249      ?      -2.345    0.019 *
Zahlungsartquartalsweise   -0.399378   0.049958  -7.994 1.30e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 17653  on 23059  degrees of freedom
Residual deviance: 17478  on 23055  degrees of freedom
AIC: 17488
```

(a) *[2 Punkte]* Welche Verteilung wählen Sie für die Zielvariable Y? (kurze Begründung).

(b) *[2 Punkte]* Die Anzahl der Policen im Regressionsmodell sei n. Wie groß ist n in obigem Modell? (kurze Begründung).

(c) *[4 Punkte]* Die Ausgabe zeigt die Punktschätzungen eines logistischen Regressionsmodells. Geben Sie die entsprechende Link- und Responsefunktion an. Stellen Sie die Likelihood als Funktion der Erwartungswerte $\mu_i = P(Y_i = 1)$ der Zufallsvariablen Y_i, $i = 1, \dots, n$, dar.

(d) *[1 Punkt]* Für das Merkmal `Zahlungsart` wurde die Dummykodierung mit `jährlich` als Referenzkategorie verwendet. Welcher Regressionskoeffizient ergibt sich für die Kategorie `jährlich` damit?

(e) *[2 Punkte]* Berechnen Sie den geschätzten Standardfehler für die Kategorie `halbjährlich` des Merkmals `Zahlungsart`.

(f) *[7 Punkte]* Geben Sie das Modell zur obigen Ausgabe an. Geben Sie den kodierten Vektor der Kovariablen für einen 50–jährigen Versicherungsnehmer mit quartalsweiser Zahlungsart an.

(g) *[3 Punkte]* Interpretieren Sie den geschätzten Koeffizienten für das Merkmal `Alter`, welches in Jahren gemessen ist (*Hinweis:* es genügt eine Interpretation).

(h) *[3 Punkte]* Interpretieren Sie die geschätzten Koeffizienten für das Merkmal `Zahlungsart`. Verwenden Sie Odds Ratios für den Vergleich mit der Referenzkategorie.

(i) *[3 Punkte]* Welche Storno-Wahrscheinlichkeit schätzt man für einen 25–jährigen Versicherungsnehmer mit halbjährlicher Zahlungsart?

(j) *[2 Punkte]* Das `Alter` ist als metrisches Merkmal linear in das Modell aufgenommen worden. Nennen Sie zwei weitere Möglichkeiten für die Aufnahme des Alters in das Modell.

(k) *[2 Punkte]* Das Alter wird jetzt kategorisiert. Das Modell mit kategorisiertem Alter und Zahlungsart liefert ein Akaike Informationskriterium (AIC) von 17503. Beurteilen Sie, welches der beiden Modelle Sie bevorzugen (das obige Modell oder das Modell mit kategorisiertem Alter).

(l) *[5 Punkte]* In obigem Modell soll mittels eines statistischen Tests überprüft werden, ob die Variable `Zahlungsart` einen signifikanten Einfluss auf die Wahrscheinlichkeit einer Stornierung hat. Welchen Test wenden Sie an? Wie lauten die Hypothesen H_0 und H_1? Die Ausgabe des Tests liefert:

```
            LR Chiquadrat Freiheitsgrade   p-Wert
Zahlungsart     77.416                 3   < 2.2e-16 ***
```

Erläutern Sie (*kurz*) alle Felder der Ausgabe (die Wahrscheinlichkeit für den Fehler 1.Art soll $\alpha = 0.05$ betragen). Wie lautet die Testentscheidung?

Lösung Aufgabe 3

(a) Bernoulli-Verteilung, da die Zielvariable binär ist.

(b) $n = 23060$. Begründung: `Null deviance` bzw. `Residual deviance` sind auf Basis des Modells nur mit Intercept (1 Parameter) bzw. des angegebenen Modells (hier: 5 Parameter incl. Intercept) berechnet. n errechnet sich als angegebene Freiheitsgrade (`degress of freedom`) plus Anzahl der Freiheitsgrade.

(c) Linkfunktion ist der logit-Link

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = x_i^T\beta \ ,$$

die Responsefunktion ist die Umkehrfunktion

$$\mu_i = \frac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)} \ .$$

Die Likelihood lautet

$$L = \prod_{i=1}^{23060} \mu_i^{y_i}(1-\mu_i)^{1-y_i}$$

(d) Die Kategorie `jährlich` ist die Referenzkategorie, daher wird diese Kategorie so kodiert, dass alle Dummyvariablen des Merkmals `Zahlungsart` Null sind. Das lässt sich auch so interpretieren, dass für die Kategorie `jährlich` der entsprechende Koeffizient 0 ist.

(e) Es gilt: Z-Wert = Schätzung/Standardfehler und damit Standardfehler = Schätzung/Z-Wert und damit $se = -0.141249/-2.345 = 0.06$

(f)

$$\begin{aligned}\log\left(\frac{\mu}{1-\mu}\right) &= \beta_0 + \beta_1 Alter + \beta_2 \text{Zahlungsart monatlich} \\ &\quad + \beta_3 \text{Zahlungsart halbjährlich} + \beta_4 \text{Zahlungsart quartalsweise}\end{aligned}$$

Kodierter Vektor: $x = (1, 50, 0, 0, 1)$

(g) Odds Ratio, Alterskoeffizient ist negativ, d.h. erhöht sich Alter um 1 Jahr, verringert sich der Odds für Storno,

$$\frac{\mu}{1-\mu} \ ,$$

um den Faktor $\exp(-0.020167) = 0.98$ (2%). Andere Interpretationen (z.B. additive Veränderung des log Odds um -0.020167) sind ebenfalls möglich.

(h) Beispiel `Zahlungsart quartalsweise`, Referenz ist `Zahlungsart jährlich`: Odds für Storno verringert sich bei quartalsweiser Zahlung im Vergleich zu jährlicher Zahlung um den Faktor $\exp(-0.399378) = 0.67$ (33%). Analog ebenfalls Verringerung des Odds bei monatlicher und halbjährlicher Zahlungsweise im Vergleich zu jährlicher Zahlungsweise. Andere Interpretationen (z.B. additive Veränderung des log Odds, jeweils im Vergleich zur Referenzkategorie) sind ebenfalls möglich.

(i) Der lineare Prädiktor ist

$$-0.915764 - 0.020167 \cdot 25 - 0.141249 = -1.561188$$

Damit: $P(Y = 1) = \exp(-1.561188)/(1 + \exp(-1.561188)) = 0.17$.

(j) Z.B. logarithmisch (log(Alter)); quadratisch, also Alter und Alter^2; kubisch, also Alter, Alter^2 und Alter^3, oder additive Modelle (GAMs) bei denen die funktionale Form aus den Daten geschätzt wird.

(k) Beim AIC gilt: „Kleiner ist besser“. Da $17488 < 17503$, wird das ursprüngliche Modell mit metrischem Alter (linear im Modell) bevorzugt.

(l) Test: Likelihood-Quotienten-Test. Mit den Bezeichnungen der Koeffizienten wie in Teilaufgabe (f), lauten die Hypothesen: $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ versus H_1 : mindestens einer dieser 3 Koeffizienten ist ungleich 0.

Ausgabe: Die χ^2-Teststatistik hat den Wert 77.416, bei 3 Freiheitsgraden, da die Zahl der Parameter von `Zahlungsart` gerade 3 beträgt und damit den Unterschied in der Anzahl der Parameter des Modells mit Zahlungsart (H_1) im Vergleich zum Modell ohne Zahlungsart (H_0) angibt. Der Test liefert zum Niveau $\alpha = 0.05$ ein signifikantes Ergebnis (da der p-Wert kleiner als α ist), d.h. H_0 wird verworfen und somit ist die `Zahlungsart` eine wichtige Variable zur Erklärung der Wahrscheinlichkeit für Storno.

Aufgabe 4. *[24 Punkte]* Zeitreihenanalyse

(a) *[3 Punkte]* Nennen Sie drei Ziele der Zeitreihenanalyse.

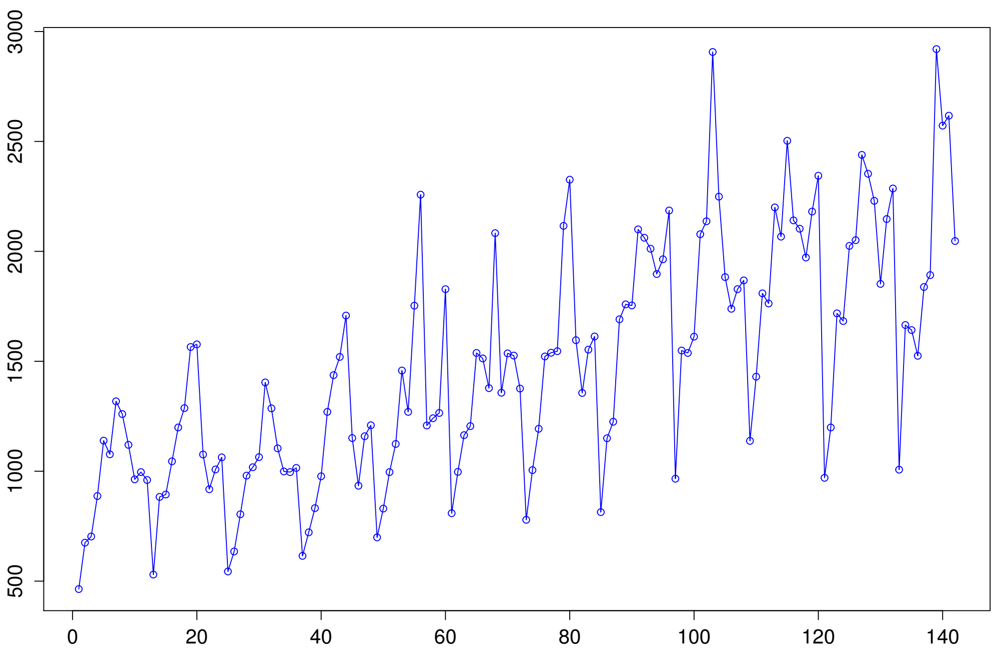
(b) *[6 Punkte]* Entscheiden Sie, ob die folgenden Aussagen richtig oder falsch sind. *Für jede richtige Antwort gibt es zwei Punkte, für jede falsche Antwort gibt es 0 Punkte.*

A Die Autokorrelationsfunktion $\rho(t, t+l), l = 0, 1, \ldots$ eines AR(p) Prozesses ist stets monoton fallend, d.h. $\rho(t, t+l_1) \geq \rho(t, t+l_2)$ für $l_1 < l_2$.

B ARMA Modelle werden verwendet um die bedingte Varianz zu modellieren, wohingegen ARCH Modelle geeignet sind den bedingten Erwartungswert zu modellieren.

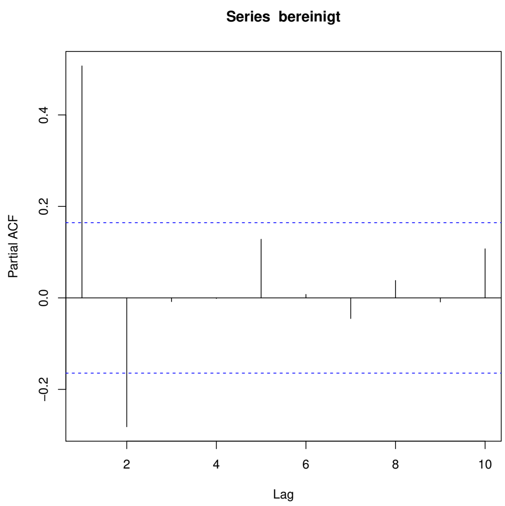
C Bei Annahme einer starren Saisonfigur ist die Summe der Saisonkomponenten Null.

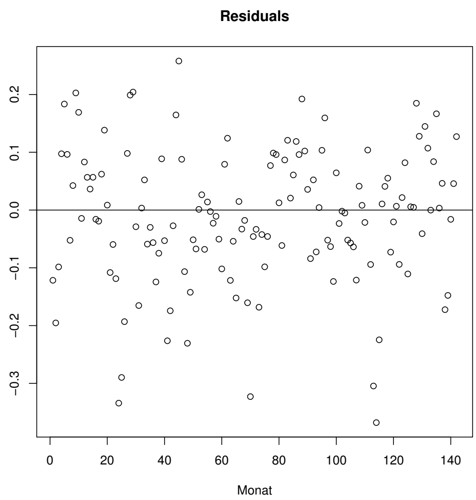
(c) *[3 Punkte]* Die folgende Grafik enthält die *monatlichen* Verkaufszahlen für Rotwein in Australien (Januar 1980 - Oktober 1991). Charakterisieren Sie den Verlauf der Zeitreihe!

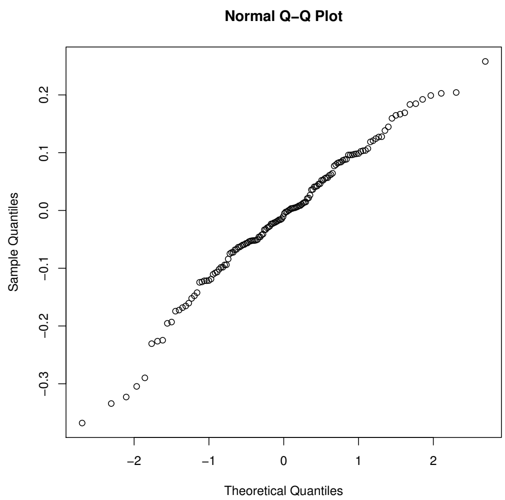


(d) *[4 Punkte]* Warum ist es nicht sinnvoll die Autokorrelationsfunktion bzw. die partielle Autokorrelationsfunktion für die originale Zeitreihe zu berechnen? Welche Transformationen bzw. Bereinigungen schlagen Sie vor?

(e) *[6 Punkte]* Für die bereinigte Zeitreihe wurde die partielle Autokorrelationsfunktion, ein QQ Plot bzw. die Residuen berechnet und visualisiert. Interpretieren Sie die Grafiken.







(f) *[2 Punkte]* Sie sollen für die bereinigte Zeitreihe ein Modell schätzen. Machen Sie einen Vorschlag!

Lösung

(a) *[3 Punkte]*

(i) Beschreibung der Zeitreihe grafisch und mit einfachen statistischen Verfahren, z.B. Hervorheben von Trends durch Glättung mittels linearer KQ–Schätzung, gleitender Durchschnitte oder lokal gewichteter Regression (LOESS).

(ii) Modellierung der Zeitreihe durch (meist stationäre) stochastische Prozesse.

(iii) Prognose zukünftiger Werte, sowie Quantifizierung deren Unsicherheit durch Prognoseintervalle.

(b) *[6 Punkte]*

A falsch
B falsch
C richtig

(c) *[3 Punkte]* Die Zeitreihe weist einen linearen Trend auf. Eine saisonale Komponente ist klar erkennbar. Die Varianz wächst über die Zeit.

(d) *[4 Punkte]* Die Zeitreihe ist nicht stationär, die Werte beider Funktionen werden sehr hoch sein. Es empfiehlt sich die Daten zu bereinigen um den Trend bzw. sie saisonale Komponente. Dies kann man entweder über ein Regressionsmodell machen oder über gleitende Durchschnitte.

(e) *[6 Punkte]* Interpretation der Grafiken

(i) Die partielle Autokorrelationsfunktion weist bis (ca.) lag = 2 Werte auf, die gegen Unkorreliertheit sprechen.

(ii) Die Residuen sind symmetrisch um die Nulllinie verteilt ohne erkennbare Muster

(iii) Der QQ-Plot weist keine Abweichungen von Normalverteilung auf

(f) *[2 Punkte]* Aufgrund der Werte der (partiellen) Autokorrelationsfunktion empfiehlt sich ein AR(2) Modell für die transformierte Zeitreihe.

Aufgabe 5. *[Credibility-Theorie, 30 Punkte]*

(a) *[6 Punkte]* Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind, und geben Sie im Fall falscher Aussagen eine kurze Begründung.

(i) Bei einem Bayes'schen Credibility-Modell wird die a-priori-Verteilung des Strukturparameters mit Hilfe von n Beobachtungen des Strukturparameters in die a-posteriori-Verteilung überführt.

(ii) Von konjugierten Verteilungen spricht man, wenn die a-posteriori-Verteilung des Strukturparameters zur selben Verteilungsklasse wie die a-priori-Verteilung gehört.

(iii) Das Bühlmann-Straub-Modell kann man anwenden, ohne eine konkrete a-priori-Verteilung für den Strukturparameter zu spezifizieren.

Schreiben Sie Ihre Antworten auf die Lösungsblätter, die Sie abgeben. Für jede richtige Antwort gibt es zwei Punkte, für jede falsche Antwort 0 Punkte.

(b) *[24 Punkte]* In einem Bayes'schen Credibility-Modell sei die Schadenhöhe X Pareto-verteilt mit Dichte

$$f_{X|\Theta=\vartheta}(x) = \vartheta x^{-(\vartheta+1)} = \vartheta x^{-1} \cdot \exp(-\vartheta \cdot \ln x)$$

für $x \geq 1$. Dabei ist ϑ die Realisierung eines Strukturparameters Θ mit Dichte

$$f_\Theta(\vartheta) = 12 \cdot (\vartheta - 1)\vartheta^{-5}$$

für $\vartheta > 1$. Folgende fünf Beobachtungen der Schadenhöhe liegen vor:

$$x_1 = 1, \qquad x_2 = 2, \qquad x_3 = 5, \qquad x_4 = 10, \qquad x_5 = 15.$$

Berechnen Sie den Wert der zugehörigen (allgemeinen) Credibility-Prämie. Gehen Sie dazu in folgenden Teilschritten vor:

(i) *[5 Punkte]* Berechnen Sie $H(\vartheta) = E(X|\Theta = \vartheta)$, den Erwartungswert der Schadenhöhe bei gegebenem Strukturparameter $\Theta = \vartheta$.

(ii) *[6 Punkte]* Rechnen Sie nach, dass die gemeinsame Dichte der Schäden und des Strukturparameters durch

$$f(\vartheta, x_1, \dots, x_5) = \frac{12}{\exp(a)} \cdot (\vartheta - 1) \cdot \exp(-\vartheta \cdot a)$$

gegeben ist, wobei hier und im Folgenden $a := \sum_{i=1}^{5} \ln x_i$ ist.

(iii) *[6 Punkte]* Ermitteln Sie die a-posteriori Dichte $f_{\Theta|X_1=x_1,\dots,X_5=x_5}(\vartheta)$ des Strukturparameters Θ.

(Hinweis: $\int_1^\infty (\vartheta - 1) \cdot \exp(-\vartheta \cdot a)\, d\vartheta = \frac{1}{a^2 \exp(a)}$ für $a > 0$, Beweis nicht erforderlich)

(iv) *[7 Punkte]* Berechnen Sie den Wert der (allgemeinen) Credibility-Prämie H^* für die gegebenen Schadenbeobachtungen.

(Hinweis: $\int_1^\infty \vartheta \cdot \exp(-\vartheta \cdot a)\, d\vartheta = \frac{1+a}{a^2 \exp(a)}$ für $a > 0$, Beweis nicht erforderlich. Falls Sie (i) bzw. (iii) nicht gelöst haben, können Sie $H(\vartheta) = \frac{\vartheta}{\vartheta - 1}$ und $f_{\Theta|X_1=x_1,\dots,X_5=x_5}(\vartheta) = a^2 \exp(a) \cdot (\vartheta - 1) \exp(-\vartheta \cdot a)$ ansetzen)

Lösung:

(a)

(i) falsch: … mit Hilfe von n Schadenbeobachtungen …

(ii) richtig

(iii) richtig

(b)

(i) Der Erwartungswert der Schadenhöhe bei gegebenem Strukturparameter ist

$$H(\vartheta) = E(X|\Theta = \vartheta) = \int x \cdot f_{X|\Theta=\vartheta}(x)\,dx = \int_1^\infty \vartheta x^{-\vartheta} dx = \vartheta \left[\frac{x^{-\vartheta+1}}{-\vartheta+1}\right]_1^\infty = \frac{\vartheta}{\vartheta-1}.$$

(ii) Für die gemeinsame Dichte der Schäden und des Strukturparameters gilt

$$\begin{aligned} f(\vartheta, x_1, \dots, x_5) &= f_\Theta(\vartheta) \cdot \prod_{i=1}^5 f_{X|\Theta=\vartheta}(x_i) \\ &= 12 \cdot (\vartheta - 1)\vartheta^{-5} \cdot \vartheta^5 \left(\prod_{i=1}^5 x_i\right)^{-1} \exp\left(-\vartheta \cdot \sum_{i=1}^5 \ln x_i\right) \\ &= \frac{12}{\prod_{i=1}^5 x_i} \cdot (\vartheta - 1) \cdot \exp\left(-\vartheta \cdot \sum_{i=1}^5 \ln x_i\right) \\ &= \frac{12}{\exp(a)} \cdot (\vartheta - 1) \cdot \exp(-\vartheta \cdot a) \end{aligned}$$

wobei $a := \sum_{i=1}^5 \ln x_i$ gesetzt wird.

(iii) Mit dem Hinweis der Aufgabe ist

$$\begin{aligned} \int_1^\infty f_\Theta(\vartheta) \cdot \prod_{i=1}^5 f_{X|\Theta=\vartheta}(x_i)\,d\vartheta &= \int_1^\infty \frac{12}{\exp(a)} \cdot (\vartheta - 1) \cdot \exp(-\vartheta \cdot a)\ d\vartheta \\ &= \frac{12}{\exp(a)} \cdot \frac{1}{a^2 \exp(a)} = \frac{12}{a^2 \exp(2a)}, \end{aligned}$$

so dass

$$\begin{aligned} f_{\Theta|\mathrm{X}_1=x_1,\ldots,\mathrm{X}_5=x_5}(\vartheta) &= \frac{f_\Theta(\vartheta)\cdot\prod_{i=1}^5 f_{X|\Theta=\vartheta}(x_i)}{\int_1^\infty f_\Theta(\tilde\vartheta)\cdot\prod_{i=1}^5 f_{X|\Theta=\tilde\vartheta}(x_i)\,d\tilde\vartheta} \\ &= \frac{a^2\exp(2a)}{12}\cdot\frac{12}{\exp(a)}\cdot(\vartheta-1)\cdot\exp(-\vartheta\cdot a) \\ &= a^2\exp(a)\cdot(\vartheta-1)\exp(-\vartheta\cdot a)\,. \end{aligned}$$

(iv) Die Credibility-Prämie berechnet man unter Nutzung des Hinweises gemäß

$$\begin{aligned} H^* &= \int_1^\infty H(\vartheta)\, f_{\Theta|\mathrm{X}_1=x_1,\ldots,\mathrm{X}_5=x_5}(\vartheta)\,d\vartheta \\ &= \int_1^\infty \frac{\vartheta}{\vartheta-1}\cdot a^2\exp(a)\,(\vartheta-1)\exp(-\vartheta\cdot a)\,d\vartheta \\ &= a^2\exp(a)\cdot\int_1^\infty \vartheta\cdot\exp(-\vartheta\cdot a)\,d\vartheta = a^2\exp(a)\cdot\frac{1+a}{a^2\exp(a)} \\ &= 1+a\,. \end{aligned}$$

Für die gegebenen Schadenbeobachtungen ist $a=\sum_{i=1}^5 \ln x_i = 7{,}3132$, also $H^* = 1+7{,}3132 = 8{,}3132$.

<u>Nachrichtlich: Beweis der Hilfsformeln:</u>

$$\begin{aligned} \int_1^\infty (\vartheta-1)\cdot\exp(-\vartheta\cdot a)\,d\vartheta &= \int_0^\infty z\cdot\exp(-(z+1)\cdot a)\,dz \\ &= \exp(-a)\int_0^\infty z\cdot\exp(-z\cdot a)\,dz = \exp(-a)\frac{\Gamma(2)}{a^2} = \frac{1}{a^2\exp(a)} \end{aligned}$$

und

$$\begin{aligned} \int_1^\infty \vartheta\cdot\exp(-\vartheta\cdot a)\,d\vartheta &= \int_1^\infty(\vartheta-1)\cdot\exp(-\vartheta\cdot a)\,d\vartheta + \int_1^\infty \exp(-\vartheta\cdot a)\,d\vartheta \\ &= \frac{1}{a^2\exp(a)} + \left[\frac{1}{-a}\exp(-\vartheta\cdot a)\right]_1^\infty = \frac{1}{a^2\exp(a)} + \frac{1}{a}\exp(-a) \\ &= \frac{1}{a^2\exp(a)} + \frac{1}{a\exp(a)} = \frac{1+a}{a^2\exp(a)}\,. \end{aligned}$$

Aufgabe 6. *[Monte Carlo Methoden] [30 Punkte]*

(a) *[5 Punkte]* Die Zufallsvariable X habe die Verteilungsfunktion

$$F(x) = \frac{e^x}{1+e^x} \qquad (x \in \mathbb{R}).$$

Geben Sie einen Algorithmus zur Simulation von X an.

(b) *[8 Punkte]* Für eine bivariate Copula C sei der folgende Simulationsalgorithmus gegeben:

(1) Erzeuge zwei unabhängige auf $[0,1]$ gleichverteilte Zufallszahlen u und w.

(2) Setze $v := \frac{u \cdot \sqrt{w}}{1-(1-u)\sqrt{w}}$.

(3) Gib den Vektor $(u,v)^\top$ aus.

Sei $\mathbf{X} = (X_1, X_2)^\top$ ein bivariater Zufallsvektor mit $C_\mathbf{X} = C$ und $X_1 \sim \mathcal{E}(3)$ sowie $X_2 \sim \mathcal{U}(-5,-2)$. Geben Sie für dieses explizite $\mathbf{X}$ den Simulationsalgorithmus an. Berechnen Sie damit unter Verwendung der beiden unabhängigen $\mathcal{U}(0,1)$-Zufallszahlen

$$u = 0,8873 \qquad \text{und} \qquad w = 0,3002$$

einen simulierten Vektor aus der Verteilung $F_\mathbf{X}$.

(c) *[3 Punkte]* Erzeugen Sie auf Basis der beiden unabhängigen standardnormalverteilten Zufallszahlen $w_1 = 0,457$ und $w_2 = 0,327$ eine Simulation $(z_1, z_2)^\top$ einer bivariaten Standardnormalverteilung mit Korrelation $\rho = -0,7$.

(d) *[8 Punkte]* Im folgenden Modell wird der Aktienkurs S_t durch zwei abhängige stochastische Differentialgleichungen beschrieben:

$$dS_t = \mu \cdot S_t\, dt + \sqrt{v_t} \cdot S_t\, dW_t^{(1)}$$
$$dv_t = \kappa \cdot [\theta - v_t]\, dt + \sigma \cdot \sqrt{v_t}\, dW_t^{(2)}$$

Dabei ist $(W_t^{(1)}, W_t^{(2)})^\top$ eine bivariate Normalverteilung. Es seien folgende Parameter gegeben:

$$\mu = 0,01, \quad \kappa = 3, \quad \theta = 0,1, \quad \sigma = 0,001, \quad \rho(W_t^{(1)}, W_t^{(2)}) = -0,7$$

sowie $S_0 = 8.000$ und $v_0 = 0,07$.

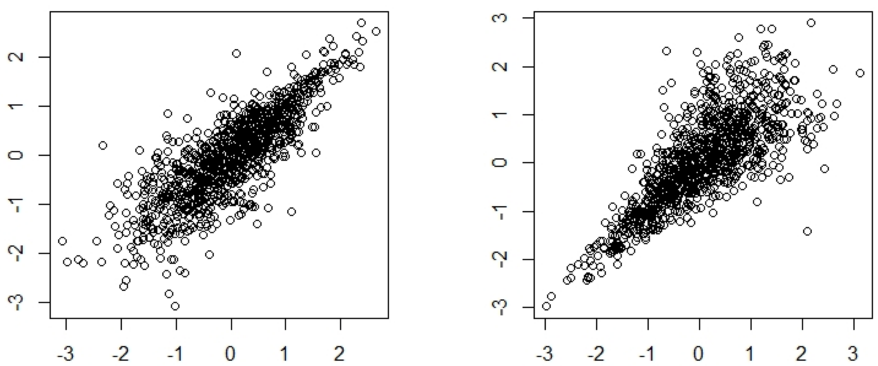
(i) Simulieren Sie die Werte S_1 und v_1 auf Basis der Zufallszahlen $(z_1, z_2)^\top$ aus Teil (c). Wie heißt die von Ihnen verwendete Simulationstechnik? (Wenn Sie (c) nicht bearbeitet haben, verwenden Sie $z_1 = 0,457$ und $z_2 = -0,086$.)

(ii) Simulieren Sie den Wert S_2 auf Basis von Teil (i). Verwenden Sie dazu die weitere standardnormalverteilte Zufallszahl $z_3 = 0,188$.

(e) *[6 Punkte]* Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind. Schreiben Sie Ihre Antworten auf die Lösungsblätter, die sie abgeben. Für jede richtige Antwort gibt es 1,5 Punkte, für jede falsche Antwort 0 Punkte.

(i) Ist F die Verteilungsfunktion einer lognormalverteilten Zufallsvariablen und $U \sim \mathcal{U}(0,1)$, dann gilt $P(F^{-1}(U) \leq t) = F(t)$.

(ii) Es wurden 1.000 Zufallspaare aus einer bivariaten Verteilung mit normalverteilten Rändern erzeugt und geplottet. In der linken Abbildung wurde eine Clayton-Copula verwendet, in der rechten eine Gumbel-Copula.



(iii) Sind $u_1, \ldots, u_n$ Simulationen aus der Verteilung $\mathcal{U}(0,1)$, dann sind $1-2u_1, \ldots, 1-2u_n$ Simulationen aus der Verteilung $\mathcal{U}(-1,0)$.

(iv) Verdoppelt man die Anzahl der Simulationen zur Schätzung einer stochastischen Kenngröße, dann halbiert sich die Länge des zugehörigen Konfidenzintervalls.

Lösung

(a) Hier kann die Inversionsmethode verwendet werden:

$$\frac{e^x}{1+e^x} = u \quad \Leftrightarrow \quad x = \ln\left(\frac{u}{1-u}\right)$$

Der Algorithmus lautet:

(1) Erzeuge eine unabhängige auf $[0,1]$ gleichverteilte Zufallszahl u.

(2) Gib $x := \ln\left(\frac{u}{1-u}\right)$ aus.

(b) Nach dem Satz von Sklar gilt: Sind univariate Verteilungen F_i, $i = 1,\dots,d$, und eine Copula C gegeben, dann ist das Folgende ein Algorithmus zur Simulation der Verteilung $C(F_1(x_1),\dots,F_d(x_d))$:

(1) Erzeuge einen Vektor $(u_1,\dots,u_d)^\top$ aus der Verteilung C.

(2) Gib den Vektor

$$(F_1^{\leftarrow}(u_1),\dots,F_d^{\leftarrow}(u_d))^\top$$

aus.

Laut Aufgabenstellung ist

$$F_1^{\leftarrow}(u_1) = -\frac{1}{3}\cdot\ln(u_1), \qquad F_2^{\leftarrow}(u_2) = 3u_2 - 5.$$

Also lautet der gewünschte Algorithmus

(1) Erzeuge zwei unabhängige auf $[0,1]$ gleichverteilte Zufallszahlen u und w.

(2) Setze $v := \frac{u\cdot\sqrt{w}}{1-(1-u)\sqrt{w}}$.

(3) Gib $(x,y)^\top$ aus mit

$$x = -\frac{1}{3}\ln(u), \qquad y = 3v - 5.$$

Setzt man die vorgegebenen Werte für u und w ein, erhält man $v = 0,5182$ und damit $x = 0,0399$ und $y = -3,4455$.

(c) Man erhält aus zwei unabhängigen standardnormalverteilten Zufallszahlen z_1, z_2 solche mit Korrelation ρ durch

$$z_1 := w_1, \quad z_2 := \rho\cdot w_1 + \sqrt{1-\rho^2}\cdot w_2.$$

Mit den gegebenen Werten erhält man $z_1 = 0,457$ und $z_2 = -0,086$.

(d) Wir verwenden das Euler-Verfahren. Die diskretisierten Versionen der SDGL lauten

$$\hat{S}_{t+\Delta t} = \hat{S}_t + \mu \cdot \hat{S}_t \cdot \Delta t + \sqrt{\hat{v}_t} \cdot \hat{S}_t \cdot \sqrt{\Delta t} \cdot z_1$$
$$\hat{v}_{t+\Delta t} = \kappa \cdot [\theta - \hat{v}_t] \cdot \Delta t + \sigma \cdot \sqrt{\hat{v}_t} \cdot \sqrt{\Delta t} \cdot z_2,$$

wobei z_1, z_2 standardnormalverteilte Zufallszahlen mit Korrelation $\rho(W_t^{(1)}, W_t^{(2)})$ sind.

(i) Für $t = 0$ und $\Delta t = 1$ ergibt sich

$$\hat{S}_1 = \hat{S}_0 + \mu \cdot \hat{S}_0 + \sqrt{\hat{v}_0} \cdot \hat{S}_0 \cdot z_1$$
$$\hat{v}_1 = \kappa \cdot [\theta - \hat{v}_0] + \sigma \cdot \sqrt{\hat{v}_0} \cdot z_2.$$

Setzt man die angegebenen Parameter und die Werte aus (c) ein (mit $\hat{S}_0 = S_0$ und $\hat{v}_0 = v_0$), erhält man $\hat{S}_1 = 9.542$ und $\hat{v}_1 = 0,16$.

(ii) Analog erhält man für $t = 1$

$$\hat{S}_2 = \hat{S}_1 + \mu \cdot \hat{S}_1 + \sqrt{\hat{v}_1} \cdot \hat{S}_1 \cdot z_3 = 10.355.$$

Es soll noch angemerkt werden, dass die Diskretisierung durchaus auch zu negativen Werten für $\hat{S}_t$ führen kann.

Alternative Lösung: Da der Prozess für S_t bei konstanter Volatilität $\sqrt{v}$ eine explizite Lösung der Form

$$S_{t+\Delta t} = S_t \cdot \exp\Big((\mu - \tfrac{1}{2}v) \cdot \Delta t + \sqrt{v} \cdot \sqrt{\Delta t} \cdot Z\Big)$$

mit $Z \sim \mathcal{N}(0,1)$ hat, ist auch Folgendes eine geeignete Simulationsmethode:

(i) Für $t = 0$ und $\Delta t = 1$ bilde

$$\hat{S}_1 = \hat{S}_0 \cdot \exp\Big((\mu - \tfrac{1}{2}\hat{v}_0) \cdot \Delta t + \sqrt{\hat{v}_0} \cdot \sqrt{\Delta t} \cdot z_1\Big) = 8.805$$
$$\hat{v}_1 = \kappa \cdot [\theta - \hat{v}_0] + \sigma \cdot \sqrt{\hat{v}_0} \cdot z_2 = 0,16.$$

(ii) Analog erhält man für $t = 1$

$$\hat{S}_2 = \hat{S}_1 \cdot \exp\Big((\mu - \tfrac{1}{2}\hat{v}_1) \cdot \Delta t + \sqrt{\hat{v}_1} \cdot \sqrt{\Delta t} \cdot z_3\Big) = 8.851.$$

(e) Lösungen mit kurzer Begründung:

(i) Richtig, das ist die Grundlage der Inversionsmethode. Da F stetig und streng monoton wachsend ist, gilt $F^{\leftarrow} = F^{-1}$.

(ii) Falsch. Da eine Clayton-Copula nur eine untere Tailabhängigkeit besitzt, eine Gumbel-Copula aber nur eine obere, erwartet man für das Modell mit Gumbel-Copula mehr Punkte in der rechten oberen Ecke und für das mit der Clayton-Copula mehr in der linken unteren Ecke.

(iii) Falsch, das sind $\mathcal{U}(-1, 1)$-Zufallszahlen.

(iv) Falsch, es wird um den Faktor $1/\sqrt{2}$ kleiner.