Schriftliche Prüfung im Grundwissen

Angewandte Stochastik

Klausur mit Lösungen

gemäß Prüfungsordnung 4.1
der Deutschen Aktuarvereinigung e. V.

am 14. Mai 2021

Hinweise:

- Zwischen dem persönlichen Login zum Download der Prüfungsaufgaben und dem Abschluss des Uploads der Lösungen ist jeglicher Kontakt zu anderen Personen (mit Ausnahme des Support-Teams) bezüglich der Prüfungsaufgaben untersagt.
- Abgesehen davon gibt es bei dieser Prüfung keine Beschränkung bei der Verwendung von Hilfsmitteln, wie z.B. Skript, Übungsaufgaben, Notizen. Außerdem ist die Nutzung eines nicht programmierbaren Taschenrechners 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]*

Überprüfen Sie die von einer Pensionskasse angewendeten biometrischen Rechnungsgrundlagen auf ihre Angemessenheit. Hierbei verwenden Sie zwei Ansätze, indem Sie auf die Anzahl der verstorbenen Rentner und auf die Rentensumme der verstorbenen Rentner abstellen. Der Bestand der Pensionskasse für männliche Rentner der Alter 60 bis 69 stellt sich wie in der folgenden Tabelle dar:

Rentnerbestand (männlich)
Rechnungszins 1%

Alter Anf. d. Jahres	Sterbe-wkeit	Anzahlen Bestand		Anzahl Tote		tatsächliche Rentensummen		Rentensummen der Sterbefälle Ende des Jahres		Deckungsrückstellung für 1€ Rente (Rentenbarwert)	
x	q_x	$n_{x,0}$ Anf. d. Jahres	$n_{x,1}$ Ende d. Jahres	tats. $n_{x,t}$	erw. $n_{x,t}$	$R_{x,0}$ Anf. d. Jahres	$R_{x,1}$ Ende d. Jahres	tats. $R_{x,tats}$	erw. $R_{x,erw}$	$V_{x,0}$ Anf. d. Jahres	$V_{x,1}$ Ende d. Jahres
60	0,0084	1000	992	8	8,4	863.100	856.776	6.324	7.250	22,27	21,67
61	0,0090	1000	990	10	9,0	795.300	788.678	6.622	7.158	21,67	21,07
62	0,0096	1000	991	9	9,6	823.100	815.846	7.254	7.902	21,07	20,46
63	0,0103	1000	988	12	10,3	896.500	888.355	8.145	9.234	20,46	19,86
64	0,0109	1000	990	10	10,9	1.002.000	991.669	10.331	10.922	19,86	19,26
65	0,0116	1000	988	12	11,6	798.400	787.541	10.859	9.261	19,26	18,66
66	0,0122	1000	987	13	12,2	689.700	681.679	8.021	8.414	18,66	18,06
67	0,0129	1000	988	12	12,9	753.000	743.480	9.520	9.714	18,06	17,45
68	0,0135	1000	985	15	13,5	936.100	923.969	12.131	12.637	17,45	16,85
69	0,0142	1000	987	13	14,2	1.021.500	1.007.993	13.507	14.505	16,85	16,24

(a) [18 Punkte] Überprüfen Sie die Anwendbarkeit der Rechnungsgrundlagen

(i) anhand der Anzahl der verstorbenen Rentner,

(ii) anhand der mit der Rentensumme gewichteten Anzahl der verstorbenen Rentner

jeweils mit Hilfe des Iterationstests zum Niveau $\alpha = 0,05$.

(b) [4 Punkte] Entscheiden Sie auf Grundlage der Ergebnisse in (a), ob die Rechnungsgrundlagen beibehalten werden können, und begründen Sie Ihre Entscheidung.

(c) [8 Punkte] Berechnen Sie für das Alter x=63 das Risikoergebnis auf Grundlage der Rentensummen.

Hinweis: Die Dichte $f(t)$ der Binomial-Verteilung lautet
für p=0,5 und n=9, 10, 11, 12 (Werte in %):

t	0	1	2	3	4	5	6	7	8	9	10	11	12
n=9	0,20	1,76	7,03	16,41	24,61	24,61	16,41	7,03	1,76	0,20			
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

Lösung

(a) (i) Überprüfung anhand der Anzahl der Todesfälle

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

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

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

Bei Gültigkeit der Hypothese $\mathbf{H_0}$ kann man folgern, dass die Vorzeichen der Differenzen zwischen den beobachteten und den rechnungsmäßigen Wahrscheinlichkeiten (Häufigkeiten) für zwei aufeinander folgende Alter „mal positiv und mal negativ“ sind. Ist die Anzahl der Vorzeichenwechsel klein, d.h. es kommen wenige Vorzeichenwechsel vor, so deutet dies auf Unterschiede zwischen den rechnungsmäßigen und den tatsächlichen Sterbewahrscheinlichkeiten (Sterbehäufigkeiten) hin.

Teststatistik:

$$T_A = \sum_{x=61}^{69} 1_{\{Sign(n_{x-1,erw}-n_{x-1,tats}) \neq Sign(n_{x,erw}-n_{x,tats})\}}$$

Es ist $n = 9$ und unter der Gültigkeit der Hypothese $\mathbf{H_0}$ ist T_A binomialverteilt mit Wahrscheinlichkeit $\frac{1}{2}$ sowohl für einen Vorzeichenwechsel als auch für keinen Vorzeichenwechsel, d.h. es gilt

$$T_A \sim \mathcal{B}(9, \frac{1}{2})$$

Zu vorgegebenem Signifikanzniveau $\alpha = 0,05$ wird ein Schwellenwert n_α so bestimmt, so dass die Hypothese abgelehnt wird, wenn der Wert der Teststatistik n_α nicht unterschreitet (einseitiger Test).

n_α wird bestimmt aus

$$P(T_A \leq n_\alpha) = \sum_{j=0}^{n_\alpha} \binom{9}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{9-j} = \left(\frac{1}{2}\right)^9 \sum_{j=0}^{n_\alpha} \binom{9}{j} \leq \alpha$$

Wegen

$$f(0) + f(1) = 0,0010 + 0,0176 = 0,0196 < \alpha$$

und

$$f(0) + f(1) + f(2) = 0,0010 + 0,0176 + 0,0703 = 0,0899 > \alpha$$

ist $n_\alpha = 1$.

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

Aufgrund der angegebenen Anzahlen von erwarteten und tatsächlichen Todesfällen ist $T_A = 8$, so dass für $\alpha = 5\%$ 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) Überprüfung anhand der Rentensummen der Todesfälle

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

$\mathbf{H_0}$: Die mit den rechnungsmäßigen Sterbewahrscheinlichkeiten gewichteten Rentensummen stimmen mit den Rentensummen der tatsächlich Verstorbenen überein.

$\mathbf{H_1}$: Die mit den rechnungsmäßigen Sterbewahrscheinlichkeiten gewichteten Rentensummen weichen von den Rentensummen der tatsächlich Verstorbenen ab.

Bei Gültigkeit der Hypothese kann man folgern, dass die Vorzeichen der Differenzen zwischen den beobachteten und den rechnungsmäßigen Rentensummen für zwei aufeinander folgende Alter „mal positiv und mal negativ“ sind. Ist die Anzahl der Vorzeichenwechsel klein, d.h. es kommen wenige Vorzeichenwechsel vor, so deutet dies auf Unterschiede zwischen den rechnungsmäßigen und den tatsächlichen Rentensummen der verstorbenen Rentner hin. Die Teststatistik lautet jetzt

$$T_R = \sum_{x=61}^{69} 1_{\{Sign(R_{x-1,erw} - R_{x-1,tats}) \neq Sign(R_{x,erw} - R_{x,tats})\}}$$

Die Verteilung von T_R ist wie die von T_A: $T_R \sim \mathcal{B}(9, \frac{1}{2})$, so dass ebenfalls $n_\alpha = 1$ gilt mit unverändertem Ablehnungsbereich A.

Aufgrund der angegebenen Rentensummen von erwarteten und tatsächlichen Todesfällen ist $T_R = 2$, so dass für $\alpha = 5\%$ T_R 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.

(b) Die Rechnungsgrundlagen können beibehalten werden. Der Iterationstest zeigt sowohl auf Grundlage der Anzahlen als auch auf Grundlage der Rentensummen keinen signifikanten Unterschied zwischen rechnungsmäßigen und beobachteten Sterblichkeiten bzw. bei einer Gewichtung der Sterblichkeiten mit den Rentensummen.

(c) Berechnung des Risikoergebnisses im Alter x=63

Die Berechnung des Risikoergebnisses wird mit Hilfe der Rekursionsformel vorgenommen:

$$V_0 = R + vV_1(1-q)$$

mit

$$\begin{aligned} V_0 &= \text{Deckungsrückstellung am Anfang des Jahres} \\ V_1 &= \text{Deckungsrückstellung am Ende des Jahres} \\ R &= \text{Rente, jährlich vorschüssig} \\ v &= \frac{1}{1+z} = \text{ Diskontfaktor (z=Zins = 1\%)} \\ q &= \text{Sterbewahrscheinlichkeit des Versicherten} \end{aligned}$$

Mit der Bezeichnungsweise aus der Aufgabenstellung gilt

$$\begin{aligned} R &= R_{x,0} \\ V_0 &= R_{x,0}V_{x,0} \\ V_1 &= R_{x,1}V_{x,1} \\ q &= q_x = \text{ Sterbewahrscheinlichkeit im Alter x} \\ R_{x,1} &= R_{x,0}(1-q_x) = \text{ erwartete Rente am Ende des Jahres.} \end{aligned}$$

Dann lautet die Rekursionsformel

$$R_{x,0}V_{x,0} = R_{x,0} + vR_{x,0}V_{x,1}(1-q_x)$$

und damit die erwartete Deckungsrückstellung am Ende des Jahres

$$R_{x,1}V_{x,1} = \frac{1}{v}\left(R_{x,0}V_{x,0} - R_{x,0}\right)$$

Mit den Werten aus der Aufgabenstellung wird die erwartete Deckungsrückstellung berechnet:

$$R_{x,1}V_{x,1} = 1,01\,(896.500\,(20,46 - 1)) = 17.620.349$$

Die tatsächliche Deckungsrückstellung beträgt am Jahresende:

$$(R_{x,0} - R_{x,tats})\,V_{x,1} = (896.500 - 8.145)\,19,86 = 17.642.730$$

Das Risikoergebnis beträgt dann (17.620.349 - 17.642.730) = -22.381 (Verlust)

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

Gegeben seien Daten x_i, $i = 1, \dots, n$ mit den Kennzahlen

$$\sum_{i=1}^{n} x_i = 22,8, \quad \sum_{i=1}^{n} x_i^2 = 74,3$$

und der empirischen Verteilungsfunktion in Abb. 1.

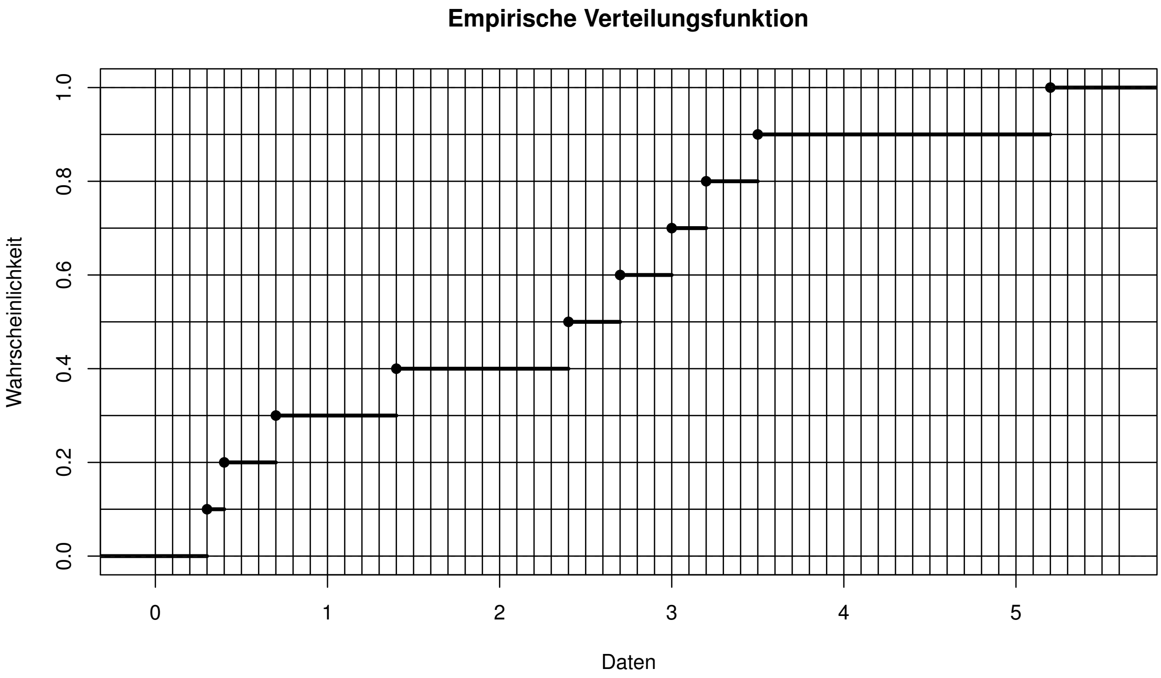


Abbildung 1: Empirische Verteilungsfunktion

(a) *[8 Punkte]* Erstellen Sie den einfachen Boxplot der Daten.

(b) *[4 Punkte]* Bestimmen Sie Mittelwert und empirische Varianz der Daten.

(c) *[12 Punkte]* In der Abbildung 2 ist der Normal-Q-Q-Plot gegeben.

(i) *[4 Punkte]* Erläutern Sie die Werte auf den Achsen. Welche Verteilung liegt dem Plot zugrunde?

(ii) *[6 Punkte]* Bestimmen Sie näherungsweise Achsenabschnitt und Steigung der eingezeichneten Geraden. Wie kann man diese beiden Parameter interpretieren?

(iii) *[2 Punkte]* Interpretieren Sie den Normal-Q-Q-Plot hinsichtlich einer Verteilungsannahme.

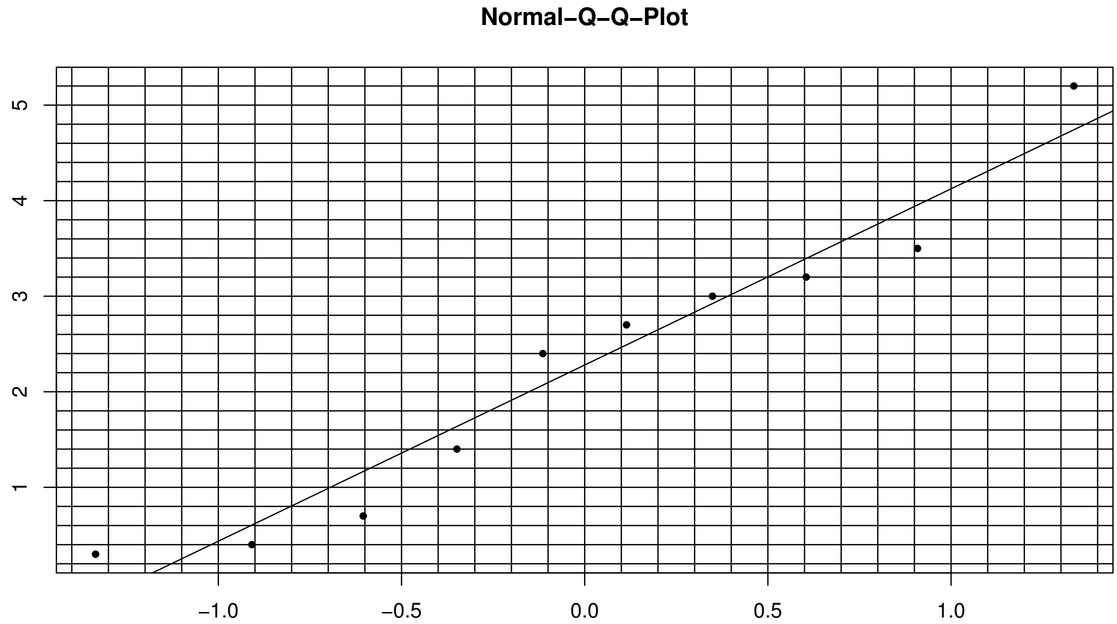


Abbildung 2: QQ-Plot

(d) *[6 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 Gegeben sei die empirische Verteilungsfunktion. Dann kann man daraus ein Histogramm erstellen.

B Gegeben sei die empirische Verteilungsfunktion. Dann kann man daraus einen Boxplot erstellen.

C Gegeben sei die empirische Verteilungsfunktion. Dann kann man daraus einen Q-Q Plot erstellen.

Lösung

(a) *[8 Punkte=4 für die Werte, 4 für den Plot]* Die fünf benötigten Werte kann man ablesen. Maximum und Minimum sind

$$x_{max} = 5,2 \text{ und } x_{min} = 0,3.$$

Das obere und untere Quartil ergeben sich beispielsweise aus

$$x_p = x_{(\lfloor np \rfloor + 1)} \text{ falls } np \notin \mathbb{N},$$

da $n = 10$ gilt, also

$$x_{0,25} = x_{(3)} = 0,7 \text{ und } x_{0,75} = x_{(8)} = 3,2.$$

Für den Median wählen wir

$$\frac{1}{2}\left(x_{(5)} + x_{(6)}\right) = \frac{1}{2}(2,4 + 2,7) = 2,55.$$

Damit ergibt sich der Boxplot in Abb. 3.

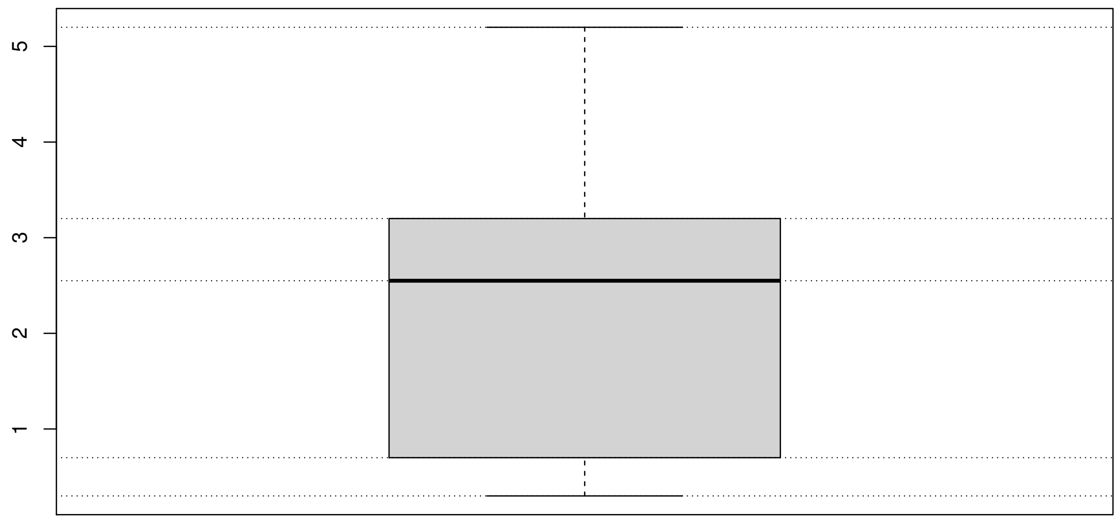


Abbildung 3: Boxplot

(b) *[4 Punkte]* Mit den gegebenen Werten folgt mit $n = 10$ für Mittelwert $\overline{x}$ und empirischer Varianz s^2

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{22,8}{10} = 2,28$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) = \frac{74,3}{9} - \frac{10 \cdot 2,28^2}{9} = 2,48$$

(c) *[12 Punkte]*

(i) *[4 Punkte]* Auf der x-Achse sind die Quantile $u_{\frac{k}{11}}$, $k = 1,...,10$ aufgetragen

-1,34 -0,91 -0,60 -0,35 -0,11 0,11 0,35 0,60 0,91 1,34

auf der y-Achse die der Größe nach geordneten Schäden. Es wird angenommen, dass eine $N(\mu, \sigma^2)$-Verteilung vorliegt.

(ii) *[6 Punkte]* Die Gerade schneidet die y-Achse bei etwa $a = 2,3$ und geht durch die Punkte $\left(-\frac{8}{10}, \frac{8}{10}\right)$ und $\left(\frac{5}{10}, \frac{32}{10}\right)$. Es ergibt sich die Steigung $b = \frac{2,4}{1,3} = 1,85$. Wenn die Verteilungsannahme stimmt, dann liegen die eingezeichneten Punkte nahe der Gerade $y = \mu + \sigma x$, also könnte man Erwartungswert und Standardabweichung mit $2,3$ und $1,85$ schätzen. Diese Werte entsprechen grob den Schätzwerten in (b).

(iii) *[2 Punkte]* Die Annahme der Normalverteilung wird durch den QQ-Plot unterstützt, da die Anpassungsgerade gut die Punkte approximiert.

(d) *[6 Punkte]* Eine Begründung war nicht erforderlich. Dennoch wird sie hier gebeben, damit die Antworten nachvollzogen werden können. Alle drei Behauptungen sind wahr, da man aus der empirischen Verteilungsfunktion alle Einzelschäden ablesen kann.

(A) Wahr.

(B) Wahr.

(C) Wahr.

Aufgabe 3. *[36 Punkte]* Induktive Statistik

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

Mit einem Poisson-Regressionsmodell soll untersucht werden, wie die *Anzahl Y von Schadensereignissen* bei Schiffspolicen innerhalb eines bestimmten Zeitintervalls von den Kovariablen Schiffstyp und Alter (in Jahren) eines Schiffes abhängt. Der Schiffstyp liegt in drei Kategorien (Typ 1, Typ 2, Typ 3) vor. Ein Modell liefert folgendes Ergebnis:

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.496032   0.259222  -5.771 7.87e-09 ***
Typ2        -0.380139   0.163049     ?     0.0197 *
Typ3         0.634329   0.140825   4.504 6.66e-06 ***
Alter        0.083575   0.009818   8.513  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Null deviance: 358.17  on 199  degrees of freedom
Residual deviance: 227.80  on 196  degrees of freedom
AIC: 601.49
```

(a) *[4 Punkte]* Beantworten Sie folgende Fragen (je richtige Antwort 2 Punkte).

(i) Begründen Sie kurz, warum die Poisson-Verteilung eine geeignete Verteilung für Y sein könnte.

(ii) Die Anzahl der Policen im Regressionsmodell sei n. Wie groß ist n gemäß Modell? (*kurze Begründung*)

(b) *[4 Punkte]* Geben Sie die natürliche Linkfunktion und die Responsefunktion des Modells (ausführlich mit dem linearen Prädiktor) an. Verwenden Sie für den linearen Prädiktor die Parametrisierung des oben geschätzen Modells mit den Koeffizienten der Ausgabe auf 2 Nachkommastellen gerundet. Stellen Sie die Likelihood als Funktion der Erwartungswerte λ_i der Zufallsvariablen Y_i, $i = 1, ..., n$ dar.

(c) *[4 Punkte]* Erläutern Sie die Berechnung des z-Werts und den genauen Zusammenhang zum p-Wert (p-value). Wie lautet der z-Wert zum Koeffizienten von Schiffstyp 2?

(d) *[8 Punkte]* Interpretation der Koeffizienten und des Modells (je richtiger Antwort 2 Punkte).

(i) Interpretieren Sie den geschätzten Koeffizienten für das Merkmal `Alter`. Die Interpretation soll dabei auf der Skala von Y (also auf Basis der Responsefunktion) erfolgen.

(ii) Interpretieren Sie die geschätzten Koeffizienten für das Merkmal Typ. Die Interpretation soll dabei auf der Skala von Y (also auf Basis der Responsefunktion) erfolgen.

(iii) Geben Sie den gesamten kodierten Vektor der Kovariablen für ein 30 Jahre altes Schiff vom Typ 2 an.

(iv) Welche Schadensanzahl schätzt man für ein 20 Jahre altes Schiff vom Typ 3?

(e) *[6 Punkte]* Alternative Modellspezifikationen (je richtiger Antwort 2 Punkte)

(i) Erläutern Sie den Zusammenhang zwischen Likelihood und AIC und wann diese jeweils für den Vergleich unterschiedlicher Modelle verwendet werden können.

(ii) Das Merkmal `Alter` ist als metrisches Merkmal linear in das Modell aufgenommen worden. Nennen Sie zwei weitere Möglichkeiten für die Aufnahme des Alters in das Modell.

(iii) Das Merkmal `Alter` werde jetzt kategorisiert. Das Modell mit kategorisiertem Alter und Schiffstyp liefert ein Akaike Informationskriterium (AIC) von 649. Beurteilen Sie, welches der beiden Modelle Sie bevorzugen: das obige Modell oder das Modell mit kategorisiertem Alter.

(f) *[10 Punkte]* In obigem Modell soll für beide Kovariablen mittels eines statistischen Tests überprüft werden, ob sie einen signifikanten Einfluss auf die Anzahl der Schadensereignisse haben. Geben Sie für beide Kovariablen die entsprechenden Hypothesen H_0 und H_1 an (*formal und in Worten*). Folgende Ausgabe zeigt die Ergebnisse der Tests:

```
LR        Chisq Df Pr(>Chisq)
Typ      63.680  2  1.486e-14 ***
Alter    78.433  1  < 2.2e-16 ***
```

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

Lösung Aufgabe 3

(a) *[4 Punkte]*

(i) Poisson-Verteilung, da Y eine Zählvariable ist.

(ii) $n = 200$. Begründung: `Null deviance` bzw. `Residual deviance` sind auf Basis des Modells nur mit Intercept (1 Parameter) bzw. des angegebenen Modells (hier: 4 Parameter incl. Intercept) berechnet. n errechnet sich als angegebene Freiheitsgrade (`degrees of freedom`) plus Anzahl der Freiheitsgrade (also 199+1=196+4=200).

(b) *[4 Punkte]* Die natürliche Linkfunktion ist der log-Link. Sei λ der (bedingte) Erwartungswert von Y.

$$\log(\lambda) = x_i^T \beta = -1.50 - 0.38 * Typ2 + 0.63 * Typ3 + 0.08 * Alter,$$

Die Responsefunktion ist die Umkehrfunktion

$$\lambda = \exp(x_i^T \beta) = \exp(-1.50 - 0.38 * Typ2 + 0.63 * Typ3 + 0.08 * Alter).$$

Die Likelihood lautet ($i = 1, \ldots, 200$):

$$L = \prod_{i=1}^{200} \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i)$$

(c) *[4 Punkte]* Es gilt: Z-Wert = Schätzung/Standardfehler und damit für Typ 2:

$$z = -0.380139/0.163049 = -2.33$$

Der (zweiseitige) p-Wert ist dabei zweimal die Wahrscheinlichkeit, dass eine standardnormalverteilte Zufallsgröße Z größer ist als der Betrag des z-Werts, also $2 * P(Z > |z|)$.

(d) *[8 Punkte]*

(i) Erhöht sich das Alter um ein 1 Jahr, erhöht sich (ceterus paribus) die erwartete (oder mittlere) Anzahl an Schadensereignissen um den Faktor $\exp(0.083575) = 1.087$, also um etwa 9%.

(ii) Schiffstyp 2 hat (ceterus paribus) im Vergleich zu Typ 1 (der Referenzkategorie) eine um den Faktor $\exp(-0.380139) = 0.68$ geringere erwartete Anzahl von Schadensereignissen (32% weniger). Schiffstyp 3 hat (ceterus paribus) im Vergleich zu Typ 1 (der Referenzkategorie) eine um den Faktor $\exp(0.634329) = 1.89$ höhere erwartete Anzahl von Schadensereignissen (89% mehr).

(iii) $x_i = (1, 1, 0, 30)$. Die erste „1“ steht für den Intercept, die folgenden beiden Dummyvariablen „(1,0)“ kodieren den Schiffstyp 2 und „30“ steht für ein 30 Jahre altes Schiff.

(iv) Man setzt die entsprechenden Ausprägungen der Variablen in die Responsefunktion ein:

$\lambda_i = \exp(-1.496032 + 0.634329 + 0.083575 * 20) = \exp(0.809797) = 2.25$ ist dann die erwartete Anzahl. Der Intercept ist -1.496032, 0.634329 ist der Koeffizient der Dummyvariablen, die den Schiffstyp 3 kodiert (hier also „1“, da Schiffstyp 3 vorliegt) und 0.083575 ist der Koeffizient für Alter.

(e) *[6 Punkte]*

(i) AIC = -2 Loglikelihood + 2 $*$ Anzahl der geschätzten Parameter. Die Loglikelihood von verschiedenen Modellen kann nur im genesteten Fall verglichen werden, d.h. wenn das „kleinere“ Modell (zum Beispiel die entsprechenden Kovariablen) im größeren Modell enthalten ist. Das AIC kann allgemein für Modellvergleiche verwendet werden. In allen Fällen muss die Datenbasis die gleiche sein (gleicher Datensatz, gleicher Stichprobenumfang).

(ii) 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.

(iii) Es gilt: 601.49 = AIC (Alter metrisch) < AIC (Alter kategorisiert) = 649. Kleineres AIC wird bevorzugt, d.h. man wählt das Modell mit metrischem Alter.

(f) *[10 Punkte]* Hypothesen:

Schiffstyp in Worten:

H_0 : Schiffstyp hat keinen Einfluss vs. H_1 : Schiffstyp hat Einfluss.

Schiffstyp formal:

$H_0 : \beta_{Typ2} = \beta_{Typ3} = 0$ vs. H_1 : mindestens ein β von Typ ist ungleich 0

Alter in Worten:

H_0 : Alter hat keinen Einfluss vs. H_1 : Alter hat Einfluss.

Alter formal:

$H_0 : \beta_{Alter} = 0$ vs. $H_1 : \beta_{Alter} \neq 0$.

Durchgeführter Test: Likelihood-Quotiententest (likelihood ratio test).

Felder der Ausgabe:

Spalte LR: Variablen

Spalte Chisq: Teststatistiken des LQ-Tests basierend auf der χ^2-Verteilung

Spalte df: Freiheitsgrade, $df = 2$ bei Typ, da 2 Koeffizienten zu Typ gehören, $df = 1$ bei metrischem Alter.

Spalte Pr(>Chisq): p-Werte, d.h. Wahrscheinlichkeit dass eine χ^2-verteilte Zufallsvariable mit df Freiheitsgraden größer ist als die jeweilige Teststatistik (63.680 bzw. 78.433).

Testentscheidung: H_0 wird für Typ und Alter abgelehnt, d.h. beide Variablen haben einen (signifikanten) Einfluss (auf die erwartete Anzahl von Schadensereignissen), da die p-Werte jeweils kleiner als das vorgegebene $\alpha = 0.05$ sind.

Aufgabe 4. *[24 Punkte]* Zeitreihenanalyse

(a) *[3 Punkte]* Nennen Sie drei Annahmen für Zeitreihenmodelle.

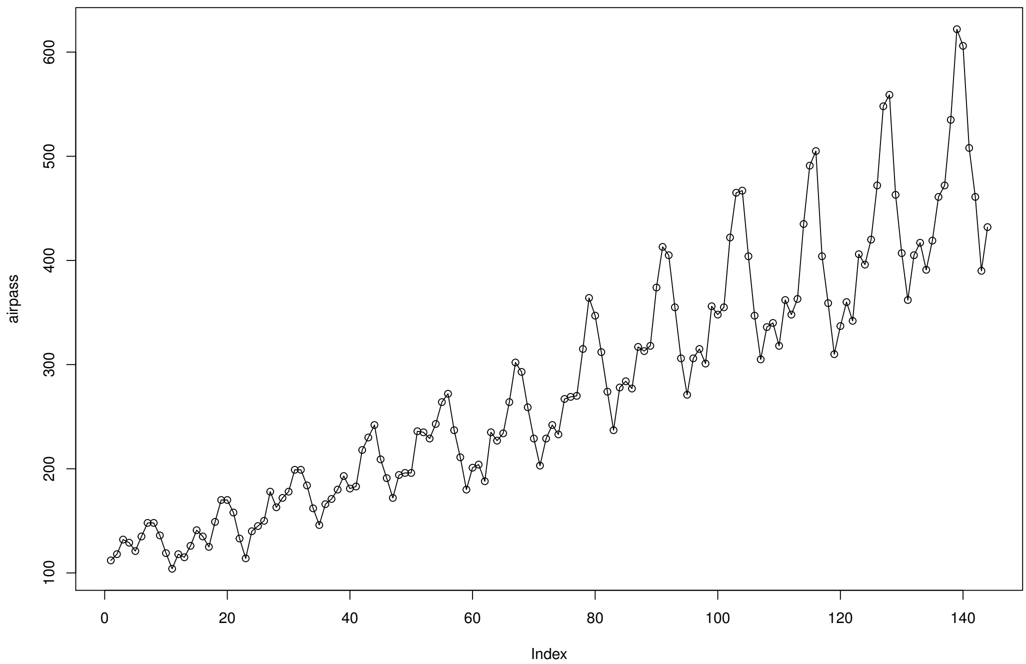
(b) *[6 Punkte]* Entscheiden Sie, ob die folgenden Aussagen richtig oder falsch sind.
Für jede richtige [falsche] Antwort gibt es zwei [Null] Punkte.

A Bei AR(p) Modellen kann man zukünftige Werte schätzen, nicht aber deren Unsicherheit.

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

C Möchte man eine Zeitreihe um den quadratischen Trend bereinigen, kann man dies mittels Differenzen erster Ordnung erreichen.

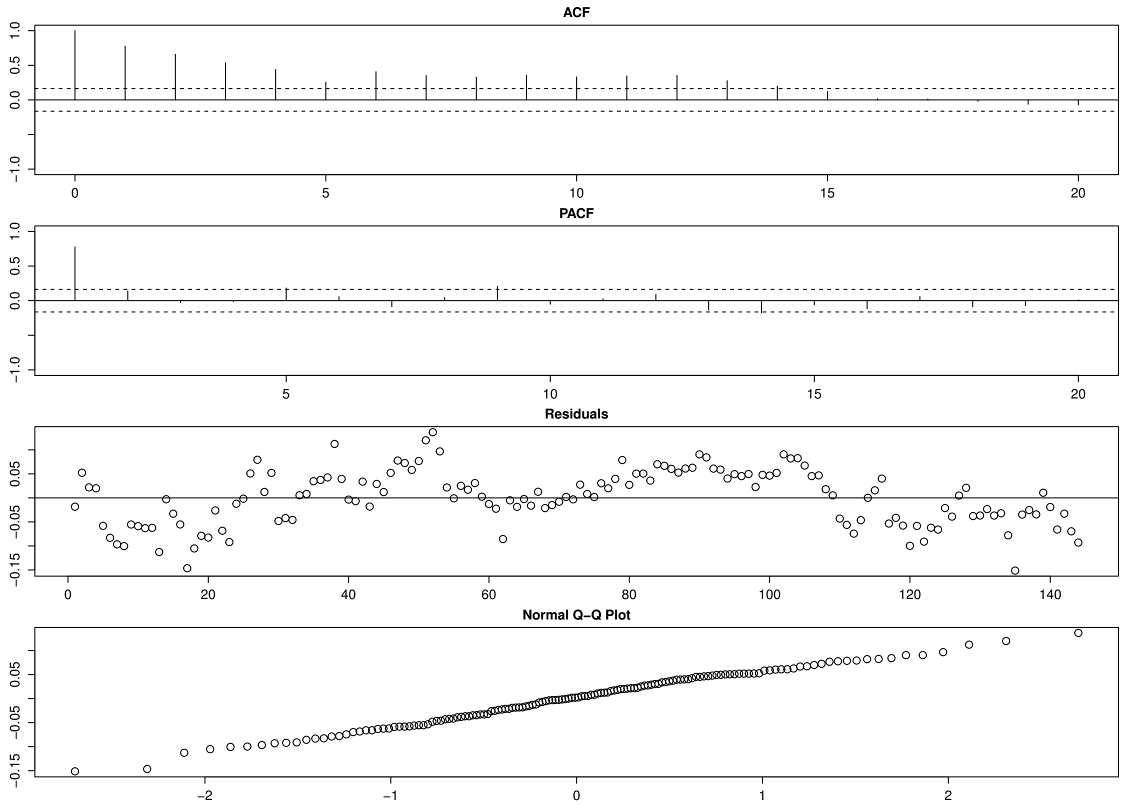
(c) (i) *[3 Punkte]* Die folgende Grafik enthält die *monatlichen* Passagierzahlen eines internationalen Flughafens (Januar 1949 - Dezember 1960). Nennen Sie drei Charakteristika, die geeignet sind den Verlauf der Zeitreihe zu beschreiben!



(ii) *[1 Punkte]* Ist die Zeitreihe stationär? Begründen Sie Ihre Aussage! (Ein Grund ist hinreichend)

(iii) *[6 Punkte]* Welche Transformationen bzw. Bereinigungen schlagen Sie vor um eine stationäre Zeitreihe zu erreichen? (mit kurzer Begründung) [jeweils ein Vorschlag entsprechend Ihrer Antwort zu (c), d.h. insgesamt 3 Vorschläge]

(iv) *[3 Punkte]* Für die bereinigte Zeitreihe wurde die Autokorrelationsfunktion bzw. die partielle Autokorrelationsfunktion, die Residuen bzw. ein QQ Plot (der Residuen) berechnet und visualisiert. Interpretieren Sie drei Grafiken Ihrer Wahl (jeweils 1 Punkt).



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

Lösung

(a) *[3 Punkte]*

(i) Die Werte sind zu äquidistanten - gleichabständigen Zeitpunkten vorhanden.

(ii) Die Werte sind Realisierungen einer metrisch skalierten Variablen.

(iii) Die Werte sind korreliert (abhängig würde auch akzeptiert werden).

(b) *[6 Punkte]*

A falsch (Stochastische Prozesse lassen sich durch ihre Verteilung bzw. Momente beschreiben)
B richtig (ARCH ist Akronym für AutoRegressive Conditional Heteroscedasticity)
C richtig (Das Verwenden von Differenzen ist allgemeines Trendbereinigungsverfahren)

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

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

(iii) *[6 Punkte]* Transformationen

i. Trendbereinigung durch Differenzen erster Ordnung oder Schätzung einer polynomialen Funktion.

ii. Gleitende Durchschnitte zur Bereinigung der Saisonalität.

iii. log-Transformation mit dem Ziel die Varianz zu stabilisieren.

(iv) *[3 Punkte]* Interpretation

i. Autokorrelationsfunktion: Die positiven Werte fallen bis lag = 15 langsam
ab (unter leichter Schwankung).

ii. Partielle Autokorrelationsfunktion: Für lag = 1 ist ist der Wert positiv und relevant.

iii. Residuen-Plot: Die Residuen streuen um den Wert Null; am Anfang bzw. am Ende (im Mittelteil) der Zeitreihe sind die Mehrheit der Werte negativ (positiv).

iv. QQ-Plot: Die Beobachtungswerte zweier Merkmale, deren Verteilung man vergleichen will, werden jeweils der Größe nach geordnet. Diese geordneten Daten werden zu Wertepaaren zusammengefasst und in einem Koordinatensystem abgetragen. Ergeben die Punkte (annähernd) eine Gerade, kann man vermuten, dass den beiden Merkmalen die gleiche Verteilung zu Grunde liegt, was hier der Fall ist (und lt. Angabe handelt es sich um eine Normalverteilung).

(v) *[2 Punkte]* Aufgrund der Werte der (partiellen) Autokorrelationsfunktion empfiehlt sich ein AR(p) Modell für die transformierte Zeitreihe mit $p = 1$.

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

(a) *[4 Punkte]* Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind, und geben Sie im Fall falscher Aussagen eine kurze Begründung.

(a1) In die linearisierte Credibility-Prämie gehen die beobachteten Schadenhöhen $x_1, \dots, x_n$ mit ihrem Mittelwert $\frac{1}{n}\sum_{i=1}^{n} x_i$ ein.

(a2) Im Bayes'schen Credibility-Modell sind die Schadenhöhen stochastisch unabhängig vom Strukturparameter.

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) *[11 Punkte]* Die Zufallsvariable X beschreibe die jährliche Schadenanzahl eines Versicherungsnehmers. X soll in einem Bayes'schen Credibility-Modell so modelliert werden, als ob es sich um eine Schadenhöhe handeln würde. X folge dabei einer Poissonverteilung mit Zähldichte

$$f_{X|\Theta=\vartheta}(x) = \frac{\vartheta^x \cdot \exp(-\vartheta)}{x!}$$

für $x \in \{0,1,2,\dots\}$. ϑ sei die Realisierung eines Strukturparameters Θ, welcher auf dem Intervall $(0, b)$ gleichverteilt sei, wobei $b > 0$ noch geeignet gewählt werden muss. Folgende zehn Beobachtungen x_i von X liegen vor:

i	1	2	3	4	5	6	7	8	9	10
x_i	2	3	3	4	2	0	1	4	1	4

Berechnen Sie den Wert der zugehörigen <u>linearisierte</u> Credibility-Prämie. Gehen Sie dazu in folgenden Teilschritten vor:

(b1) *[2 Punkte]* Geben Sie die Funktion $H(\vartheta) = E(X|\Theta = \vartheta)$ an und berechnen Sie daraus den Erwartungswert $E(X) = E(H(\Theta))$ in Abhängigkeit des Parameters b.

(b2) *[2 Punkte]* Aus Marktbeobachtungen sei bekannt, dass $E(X) = 1{,}5$. Welche Wahl für b ergibt sich hieraus?

(b3) *[7 Punkte]* Berechnen Sie damit den Wert der <u>linearisierten</u> Credibility-Prämie H^{**}.

(c) *[15 Punkte]* Berechnen Sie für das Modell und die Beobachtungen aus (b) den Wert der allgemeinen Credibility-Prämie H^*. Gehen Sie dazu in folgenden Teilschritten vor:

(c1) *[7 Punkte]* Rechnen Sie zunächst nach, dass sich als a-posteriori-Dichte

$$f_{\Theta|X_1=x_1,\ldots,X_{10}=x_{10}}(\vartheta) = \frac{10^{25}}{24!} \cdot \frac{\vartheta^{24} \cdot \exp(-10\,\vartheta)}{F_{(25;\,10)}(3)}$$

für $\vartheta \in (0,3)$ ergibt. $F_{(25;10)}$ ist darin die Verteilungsfunktion einer Gamma-Verteilung $\Gamma(25;10)$.

(Hinweis: Es ist $\int_0^b \vartheta^m \cdot \exp(-n\vartheta)\,d\vartheta = \frac{m!}{n^{m+1}} F_{(1+m;\,n)}(b)$, Beweis nicht erforderlich. Falls Sie (b) nicht gelöst haben, können Sie $b = 3$ annehmen.)

(c2) *[2 Punkte]* Sind die Verteilung von X und die a-priori-Verteilung von Θ konjungierte Verteilungen? Begründen Sie Ihre Antwort.

(c3) *[6 Punkte]* Berechnen Sie den Wert der allgemeinen Credibility-Prämie H^*.

(Hinweis: Nutzen Sie nochmals das Integral aus dem Hinweis zu (c1). Ein Statistikprogramm oder Excel liefert für die Gammaverteilung $F_{(25;\,10)}(3) = 0{,}8428$ und $F_{(26;\,10)}(3) = 0{,}7916$.)

Lösung:

(a)

(a1) richtig

(a2) falsch: Je nach Ausprägung des Strukturparameters liegt eine andere Schadenhöhenverteilung vor.

(b)

(b1) Der Erwartungswert der Schadenhöhe bei gegebenem Strukturparameter ist für die Poissonverteilung $H(\vartheta) = E(X|\Theta = \vartheta) = \vartheta$. Somit ist wegen der Gleichverteilung von Θ

$$E(X) = E(H(\Theta)) = E(\Theta) = b/2.$$

(b2) Für $b = 3$ ergibt sich in (b1) der Wert von $E(X) = 1{,}5$.

(b3) Zur Berechnung des Credibility-Faktors benötigt man

$$Var\big(H(\Theta)\big) = Var(\Theta) = b^2/12 = 0{,}75$$

$$E(Var[X|\Theta]) = E(\Theta) = b/2 = 1{,}5$$

wobei wegen der Poisson-Verteilung $Var[X|\Theta] = \Theta$ ist. Als Credibility-Faktor ergibt sich

$$z = \frac{Var\big(H(\Theta)\big)}{\frac{1}{n}E(Var[X|\Theta]) + Var\big(H(\Theta)\big)} = \frac{0{,}75}{\frac{1{,}5}{10} + 0{,}75} = 0{,}8333.$$

Die linearisierte Credibility-Prämie ist somit

$$H^{**} = z\bar{X} + (1 - z)E(X) = 0{,}8333 \cdot \frac{24}{10} + 0{,}1667 \cdot 1{,}5 = 2{,}25.$$

(c)

(c1) Die A-posteriori-Dichte ist

$$f_{\Theta|X_1=x_1,\dots,X_{10}=x_{10}}(\vartheta) = \frac{f_\Theta(\vartheta)\cdot\prod_{i=1}^{10} f_{X|\Theta=\vartheta}(x_i)}{\int f_\Theta(\tilde{\vartheta})\cdot\prod_{i=1}^{10} f_{X|\Theta=\tilde{\vartheta}}(x_i)\, d\tilde{\vartheta}}.$$

Der Zähler ist dabei

$$\frac{1}{b} \cdot \prod_{i=1}^{10} \frac{\vartheta^{x_i} \cdot \exp(-\vartheta)}{x_i!} = \frac{\vartheta^{\left(\sum_{i=1}^{10} x_i\right)} \cdot \exp(-10\,\vartheta)}{b \cdot \prod_{i=1}^{10}(x_i!)} = \frac{\vartheta^{24} \cdot \exp(-10\,\vartheta)}{3 \cdot \prod_{i=1}^{10}(x_i!)}$$

für $\vartheta \in (0;3)$. Mit dem Hinweis der Aufgabe erhält man für den Nenner

$$\int_0^3 \frac{\tilde{\vartheta}^{24} \cdot \exp\left(-10\,\tilde{\vartheta}\right)}{3 \cdot \prod_{i=1}^{10}(x_i!)}\, d\tilde{\vartheta} = \frac{1}{3 \cdot \prod_{i=1}^{10}(x_i!)} \int_0^3 \tilde{\vartheta}^{24} \cdot \exp\left(-10\,\tilde{\vartheta}\right) d\tilde{\vartheta}$$

$$= \frac{1}{3 \cdot \prod_{i=1}^{10}(x_i!)} \cdot \frac{24!}{10^{25}} \cdot F_{(25;\,10)}(3).$$

Insgesamt ergibt sich wie angegeben

$$f_{\Theta|X_1=x_1,\ldots,X_{10}=x_{10}}(\vartheta) = \frac{10^{25}}{24!} \cdot \frac{\vartheta^{24} \cdot \exp(-10\,\vartheta)}{F_{(25;\,10)}(3)}.$$

(c2) Nein, denn sonst hätte die a-posteriori-Verteilung zur selben Verteilungsklasse wie die a-priori-Verteilung gehören müssen (Gleichverteilung), was offensichtlich nicht der Fall ist.

(c3) Die allgemeine Credibility-Prämie berechnet man unter nochmaliger Nutzung des Hinweises gemäß

$$\begin{aligned}
H^* &= \int_0^3 H(\vartheta)\, f_{\Theta|X_1=x_1,\ldots,X_{10}=x_{10}}(\vartheta)\, d\vartheta \\
&= \int_0^3 \vartheta \cdot \frac{10^{25}}{24!} \cdot \frac{\vartheta^{24} \cdot \exp(-10\,\vartheta)}{F_{(25;\,10)}(3)}\, d\vartheta \\
&= \frac{10^{25}}{24! \cdot F_{(25;\,10)}(3)} \int_0^3 \vartheta^{25} \cdot \exp(-10\,\vartheta)\, d\vartheta \\
&= \frac{10^{25}}{24! \cdot F_{(25;\,10)}(3)} \cdot \frac{25!}{10^{26}} F_{(26;\,10)}(3) = \frac{25}{10} \cdot \frac{F_{(26;\,10)}(3)}{F_{(25;\,10)}(3)} \\
&= \frac{25}{10} \cdot \frac{0{,}7916}{0{,}8428} = 2{,}3481.
\end{aligned}$$

Nachrichtlich: Beweis der Hilfsformeln:

Die Verteilungsfunktion der $\Gamma(1+m;\, n)$-Verteilung berechnet sich aus der Dichte gemäß

$$F_{(1+m;\,n)}(b) = \int_0^b \frac{n^{m+1}}{\Gamma(m+1)} \cdot \vartheta^m \cdot \exp(-n\vartheta)\, d\vartheta = \frac{n^{m+1}}{m!} \int_0^b \vartheta^m \cdot \exp(-n\vartheta)\, d\vartheta,$$

woraus sich durch Umstellen die Formel des Hinweises ergibt.

Aufgabe 6. *[Stochastische Prozesse und deren Simulation] [30 Punkte]*

(a) Zur Modellierung einer Berufsunfähigkeitsversicherung betrachten wir die drei Statusgruppen Aktive (A), Invalide (I) und Verstorbene (T). Zu jedem Zeitpunkt ist jede Person in genau einer dieser Gruppen. Wir bezeichnen mit $P(A(t,x) \to I(t+1,x+1))$ die Wahrscheinlichkeit, dass eine zum Zeitpunkt t aktive Person des (ganzzahligen) Alters x nach einem Jahr (also zum Zeitpunkt $t+1$ und mit Alter $x+1$) invalide ist. Analoge Bezeichnungen gelten für die anderen Kombinationen von A, I und T. Der Einfachheit halber seien diese einjährigen Übergangswahrscheinlichkeiten im Folgenden unabhängig vom Alter und dem Zeitpunkt, wir schreiben daher $P(A \to I)$ usw. Aus Beobachtungsdaten ergeben sich die Werte

$$P(A \to I) = 0,2, \quad P(I \to A) = 0,15, \quad P(A \to T) = 0,1, \quad P(I \to T) = 0,2.$$

(i) *[2 Punkte]* Für $t \in \mathbb{N}_0$ sei X_t die Statusgruppe einer Person zum Zeitpunkt t. Den Voraussetzungen zufolge ist $(X_t)_{t\in\mathbb{N}_0}$ eine homogene Markovkette. Geben Sie die Übergangsmatrix Π an. Ordnen Sie dabei die Zustände in der Reihenfolge A, I und T an.

(ii) *[4 Punkte]* In $t = 0$ betritt eine Person die Statusgruppe der Aktiven. Bestimmen Sie den Zustandsvektor der Person nach zwei Jahren.

(iii) *[5 Punkte]* Bestimmen Sie die Eigenwerte von Π.
Hinweis: Verwenden Sie die Spaltenvektoren $(4,3,0)^\top$ und $(-1,1,0)^\top$ als Hilfsmittel.

(iv) *[5 Punkte]* Untersuchen Sie, ob die Markovkette $(X_t)_{t\in\mathbb{N}_0}$ einen stationären Zustand besitzt. Falls ja, geben Sie diesen an.

(b) In der Situation von Teil (a) lautet der Zustandsvektor zum Zeitpunkt $t = 2$ einer in $t = 0$ in die Aktivengruppe eingetretenen Person

$$\mathbf{p}(2) = (0,52; 0,27; 0,21).$$

(i) *[4 Punkte]* Geben Sie einen expliziten Simulationsalgorithmus an, der gemäß dieser Verteilung einen Wert aus der Menge $\{A, I, T\}$ ausgibt.

(ii) *[4 Punkte]* Wir betrachten nun 1000 aktive Personen in $t = 0$ und wollen die Aufteilung der Gruppe auf A, I und T in $t = 2$ simulieren, also einen Vektor $\mathbf{s} = (s_A, s_I, s_T)$ erzeugen, der angibt, wie viele der Personen in welchem Zustand sind. Geben Sie einen expliziten Algorithmus dafür an. Verwenden Sie dazu den Algorithmus aus Teil (bi).

(c) *[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 2 Punkte, für jede falsche Antwort 0 Punkte.

(i) Ist $(X_n)_{n\in\mathbb{N}_0}$ eine endliche homogene Markovkette mit Übergangsmatrix Π und $\boldsymbol{u}\cdot\Pi = \boldsymbol{u}$ für einen Zustandsvektor $\boldsymbol{u}$, dann ist $\boldsymbol{u}$ ein stationärer Zustand (im Sinne einer asymptotischen Verteilung) von $(X_n)_{n\in\mathbb{N}_0}$.

(ii) Ist $(X_t)_{t\geq 0}$ ein stochastischer Prozess mit $X_0 \equiv 0$ sowie stationären Zuwächsen und $X_t \sim \mathcal{E}(t)$ für alle $t > 0$, dann gilt für $s < t$: $X_t - X_s \sim \mathcal{E}(t-s)$.

(iii) Für den stochastischen Prozess $X_t = t^2 + t - W_t^2$ gilt $\mathrm{d}X_t = 2t\,\mathrm{d}t - 2W_t\,\mathrm{d}W_t$.

Lösungsvorschläge

(ai) Die fehlenden Angaben (hier in rot) ergeben sich aus der Tatsache, dass sich die Werte in jeder Zeile der Matrix zu Eins aufaddieren; die letzte Zeile ist klar, da man den Status T nicht mehr verlassen kann:

$$\Pi = \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,15 & 0,65 & 0,2 \\ 0 & 0 & 1 \end{pmatrix}$$

(aii) Der Startzustand ist $\mathbf{p}(0) = (1,0,0)$. Dann ist

$$\mathbf{p}(2) = \mathbf{p}(0) \cdot \Pi^2 = (1,0,0) \cdot \begin{pmatrix} 0,52 & 0,27 & 0,21 \\ 0,2025 & 0,4525 & 0,345 \\ 0 & 0 & 1 \end{pmatrix} = (0,52; 0,27; 0,21).$$

(aiii) $\lambda = 1$ ist immer ein Eigenwert. Weiterhin gilt

$$\begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,15 & 0,65 & 0,2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 3,4 \\ 2,55 \\ 0 \end{pmatrix} = 0,85 \cdot \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}$$

sowie

$$\begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,15 & 0,65 & 0,2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -0,5 \\ 0,5 \\ 0 \end{pmatrix} = 0,5 \cdot \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Also sind $\lambda = 0,85$ und $\lambda = 0,5$ die weiteren Eigenwerte.

(aiv) Da es genau einen Eigenwert λ mit $|\lambda| = 1$ gibt und $\lambda = 1$ die algebraische Vielfachheit Eins hat (denn es gibt drei verschiedene Eigenwerte), existiert nach dem Konvergenzsatz für endliche homogene Markovketten ein stationärer Zustand $\mathbf{p}^*$. Dies ist genau der Zustandsvektor, der die Gleichung $\mathbf{p}^* \cdot \Pi = \mathbf{p}^*$ löst. Da

$$(p_1^*, p_2^*, p_3^*) \cdot \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,15 & 0,65 & 0,2 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= (0,7p_1^* + 0,15p_2^*, 0,2p_1^* + 0,65p_2^*, 0,1p_1^* + 0,2p_2^* + p_3^*)$$

lautet das zu lösende Gleichungssystem

$$\begin{aligned} 0,7p_1^* + 0,15p_2^* &= p_1^* \\ 0,2p_1^* + 0,65p_2^* &= p_2^* \\ 0,1p_1^* + 0,2p_2^* + p_3^* &= p_3^* \end{aligned}$$

mit der Lösungsmenge $\{(0,0,x) : x \in \mathbb{R}\}$. Der stationäre Zustand lautet also $(0,0,1)$.
Da es sich bei T um einen absorbierenden Zustand handelt, kann man $(0,0,1)$ auch direkt (also durch Einsetzen in das Gleichungssystem) als den stationären Zustand identifizieren.

(bi) Ohne Einschränkung habe die Zufallsvariable – wir nennen sie Z – die Werte 1,2 und 3. Aus der Inversionsmethode erhält man:

(1) Erzeuge eine $\mathcal{U}(0,1)$-Zahl u.

(2) Falls $0 < u \leq 0,52$, setze $Z = 1$; falls $0,52 < u \leq 0,79$ $(= 0,52 + 0,27)$, setze $Z = 2$; sonst setze $Z = 3$.

(bii) Wir simulieren 1000 Mal die Zufallsgröße Z und addieren die erhaltenen Ergebnisse (wobei 1=A, 2=I und 3=T):

(1) Erzeuge unabhängige Zahlen $z_1, \dots, z_{1000}$ wie in (bi).

(2) Setze $\mathbf{s}_0 := (0,0,0)$. Für $i = 1, \dots, 1000$ setze

$$\mathbf{s}_i := \mathbf{s}_{i-1} + \begin{cases} (1,0,0), & \text{falls } z_i = 1 \\ (0,1,0), & \text{falls } z_i = 2 \\ (0,0,1), & \text{falls } z_i = 3. \end{cases}$$

Der Vektor $\mathbf{s}_{1000}$ enthält die simulierten Anzahlen der drei Statusgruppen in $t = 2$.

(ci) Falsch: Die Existenz einer Lösung des Gleichungssystems führt nicht automatisch zu einem stationären Zustand. Vielmehr müssen noch z.B. die Voraussetzungen eines entsprechenden Konvergenzsatzes erfüllt sein.

(cii) Richtig: Die Stationarität der Zuwächse bedeutet nämlich

$$X_t - X_s \sim X_{t-s} - X_0 = X_{t-s}.$$

(ciii) Richtig: Denn es gilt $\int_0^t 2u \, \mathrm{d}u = t^2$ und $\int_0^t W_u \mathrm{d}W_u = \frac{1}{2}(W_t^2 - t)$.