Aufgabe 1 (30 Punkte)

Sie haben die Sterblichkeit in einem Bestand von Rentenversicherungen (jährlich vorschüssige Rentenzahlung) überprüft und dabei nicht nur den gesamten Bestand sondern auch den Teilbestand der Versicherten mit einer jährlichen Rente von 20.000 € und höher separat ausgewertet. Für die Alter 70-74 ergeben sich die in der folgenden Tabelle aufgeführten Anzahlen.

Alter	rechnungsmäßige Sterbewahr-scheinlichkeit (2. Ordnung)	Bestand am Anfang des Jahres		tatsächliche Todesfälle im Laufe des Jahres		erwartete Todesfälle im Laufe des Jahres		Differenz tats. - erw. Todesfälle	
		gesamt	davon mit Rente ≥ 20.000	gesamt	davon mit Rente ≥ 20.000	gesamt	davon mit Rente ≥ 20.000	gesamt	davon mit Rente ≥ 20.000
x	q_x	B_G	B_T	T_G	T_T	E_G	E_T		
70	0,0188	10.000	1.200	179	15	188,0	22,6	-9,0	-7,6
71	0,0209	10.000	1.200	213	17	209,0	25,1	4,0	-8,1
72	0,0232	10.000	1.200	218	25	232,0	27,8	-14,0	-2,8
73	0,0257	10.000	1.200	237	17	257,0	30,8	-20,0	-13,8
74	0,0284	10.000	1.200	292	31	284,0	34,1	8,0	-3,1
	Summe	50.000	6.000	1.139	105	1.170,0	140,4	-31,0	-35,4

Tabelle 1

a) Zeigen Sie anhand des χ^2-Tests auf dem 5 % Niveau, dass die Sterbetafel für den gesamten Bestand B_G wohl angemessen ist jedoch für den Teilbestand B_T der Rentenempfänger mit einer Jahresrente von 20.000 € und höher nicht angemessen ist.
(Hinweis: Die Teststatistik beträgt für den Bestand B_G T = 3,134 und für den Teilbestand B_T T = 11,917; zur χ^2-Verteilung wird auf Tabelle 2 verwiesen.)

b) Für einen 71-jährigen Rentner mit einer Jahresrente von 20.000€ betrage die Deckungsrückstellung V_{71} auf Basis der bisher verwendeten Rechnungsgrundlagen zweiter Ordnung sowie bei einem Rechnungszins von 1,25 % 200.000 €. Geben Sie die Deckungsrückstellung V_{70} an.

c) Nehmen Sie an, dass die tatsächliche Sterblichkeit im Teilbestand B_T 75 % der rechnungsmäßigen Sterblichkeit beträgt. Ermitteln Sie den erwarteten Risikoverlust für die 70-jährigen Rentner im Teilbestand B_T.

d) Für Verträge mit einer jährlichen Rente von 20.000 € und mehr soll eine modifizierte Sterbetafel verwendet werden, die ein Sterblichkeitsniveau von 75 % im Vergleich zu der bisher verwendeten Sterbetafel beinhaltet (d.h. $\tilde{q}_x = 0,75 \cdot q_x$ mit $\tilde{q}_x =$ Sterbewahrscheinlichkeit der neuen Tafel 2. Ordnung). Zeigen Sie, dass die altersunabhängigen Sicherheitsabschläge $\tilde{s}^\alpha$ auf die Sterbewahrscheinlichkeiten $\tilde{q}_x$ bei gleichem Sicherheitsniveau höher sein müssen als die Sicherheitszuschläge s^α der bisher verwendeten Tafel. Bei der Ermittlung der Sicherheitsabschläge sollen nur die Anzahlen der Personen betrachtet werden.

Lösungshinweis

Freiheitsgrade	90%	92,5%	95%	97,5%	99%	99,5%
1	2,70554	3,17005	3,84146	5,02389	6,63490	7,87944
2	4,60517	5,18053	5,99146	7,37776	9,21034	10,59663
3	6,25139	6,90464	7,81473	9,34840	11,34487	12,83816
4	7,77944	8,49628	9,48773	11,14329	13,27670	14,86026
5	9,23636	10,00831	11,07050	12,83250	15,08627	16,74960
6	10,64464	11,46595	12,59159	14,44938	16,81189	18,54758
7	12,01704	12,88343	14,06714	16,01276	18,47531	20,27774
8	13,36157	14,26974	15,50731	17,53455	20,09024	21,95495
9	14,68366	15,63094	16,91898	19,02277	21,66599	23,58935
10	15,98718	16,97137	18,30704	20,48318	23,20925	25,18818

Tabelle 2: Quantile der χ^2-Verteilung

Lösungsvorschlag Aufgabe 1

a) **10 Punkte**

Formulierung des Testproblems:

Nullhypothese $\mathbf{H_0}$: Die tatsächlichen und die rechnungsmäßig unterstellten Sterbewahrscheinlichkeiten stimmen überein.

Alternativhypothese $\mathbf{H_1}$: Die tatsächlichen und die rechnungsmäßig unterstellten Sterbewahrscheinlichkeiten sind verschieden.

Bezeichnungen

L_x sei die Anzahl der Lebenden des Alters x,

$$Z_{x_j} := \begin{cases} 1, & \text{wenn die } j-\text{te Person des Alters } x \text{ ausscheidet } (j = 0,1,\ldots,L_x) \\ 0, & \text{sonst} \end{cases}.$$

Dann ist die Zufallsvariable Z_{x_j} binomialverteilt mit Erwartungswert $\mathrm{E}(Z_{x_j}) = q_x$.

Außerdem bezeichne

$Z_x := \sum_{j=1}^{L_x} Z_{x_j}$ die Anzahl der beobachteten Ausgeschiedenen des Alters x,

$E_x := L_x \cdot q_x$ die Anzahl der rechnungsmäßig erwarteten Ausgeschiedenen des Alters x.

Dann ist Z_x näherungsweise poissonverteilt mit Parameter E_x, d.h. es gilt $\mathrm{E}(Z_x)=\mathrm{Var}(Z_x)= E_x$.

χ^2-Test

<u>Teststatistik:</u>

$$T = \sum_{x=x_1}^{x_n} \frac{(Z_x - E_x)^2}{E_x} = \sum_{x=x_1}^{x_n} \chi_x^2 \quad \text{mit} \quad \chi_x := \frac{Z_x - E_x}{\sqrt{E_x}}.$$

Unter der Nullhypothese gilt näherungsweise $\chi_x \sim N(0,1)$. Damit ist die Teststatistik näherungsweise χ^2-verteilt mit n Freiheitsgraden.

Zu einem vorgegebenem Signifikanzniveau α wird das $(1-\alpha)$-Quantil der χ^2-Verteilung mit n Freiheitsgraden bestimmt und die Nullhypothese abgelehnt, wenn der Wert der Teststatistik dieses überschreitet, d.h. falls $T > \chi^2_{n;1-\alpha}$ ist.

Durchführung des Tests für den gesamten Bestand B_G:

	gesamter Bestand B_G			
Alter	tats. Todesfälle	erwartete Todesfälle		
x	Z_x	E_x	$(Z_x - E_x)^2$	$(Z_x - E_x)^2 / E_x$
70	179,0	188,0	81,0	0,431
71	213,0	209,0	16,0	0,077
72	218,0	232,0	196,0	0,845
73	237,0	257,0	400,0	1,556
74	292,0	284,0	64,0	0,225
			Teststatistik:	**3,134**

Das 95%-Quantil $\chi^2_{5;0,95}$ der χ^2-Verteilung mit 5 Freiheitsgraden beträgt $\chi^2_{5;0,95} = 11,07$. Damit lehnt der Test für den gesamten Bestand die Hypothese nicht ab.

Durchführung des Tests für den Teilbestand B_T:

	Teilbestand B_T			
Alter	tats. Todesfälle	erwartete Todesfälle		
x	Z_x	E_x	$(Z_x - E_x)^2$	$(Z_x - E_x)^2 / E_x$
70	15,0	22,6	57,8	2,556
71	17,0	25,1	65,6	2,614
72	25,0	27,8	7,8	0,282
73	17,0	30,8	190,4	6,183
74	31,0	34,1	9,6	0,282
			Teststatistik:	**11,917**

Für den Teilbestand B_T gilt $T > \chi^2_{5;0,95} = 11,07$. Damit wird die Nullhypothese abgelehnt; die rechnungsmäßigen Sterbewahrscheinlichkeiten unterscheiden sich von den im Teilbestand vorhandenen Sterbewahrscheinlichkeiten.

b) **6 Punkte**

Nach der Rekursionsformel gilt für die Deckungsrückstellung einer Rentenversicherung in der Rentenbezugszeit

$$V_{70} = R + v \cdot V_{71} \cdot (1 - q_{70})$$

mit

V_{70}	Deckungsrückstellung am Anfang des Versicherungsjahres im Alter 70
V_{71}	Deckungsrückstellung am Ende des Versicherungsjahres im Alter 71
R	Jahresrente vorschüssig, fällig am Beginn des Versicherungsjahres
v	$= \dfrac{1}{1+z}$ Diskontfaktor (z = Zins)
q_{70}	Sterbewahrscheinlichkeit des Versicherten im Alter 70

Demzufolge beträgt die Deckungsrückstellung im Alter 70

$$
\begin{aligned}
V_{70} &= R + v \cdot V_{71} \cdot (1 - q_{70}) \\
&= 20.000 + \frac{1}{1+0,0125} \cdot 200.000 \cdot (1 - 0,0188) \\
&= 213.817
\end{aligned}
$$

c) **6 Punkte**

Wenn für die tatsächliche Sterblichkeit im Teilbestand B_T gilt $\tilde{q}_x = 0,75 \cdot q_x$, so wird im Folgejahr ein Risikoverlust entstehen, da weniger Rentner versterben als rechnungsmäßig angenommen wird. Demzufolge wird die im Alter 70 gebildete rechnungsmäßige Deckungsrückstellung nicht ausreichen, um die tatsächlich am Ende des Jahres zu bilanzierende Deckungsrückstellung zu finanzieren. Der Risikoverlust *A* (bezogen auf den Anfang des Jahres) ist demzufolge die Differenz zwischen der Deckungsrückstellung $\tilde{V}_{70}$ gebildet nach der Rekursionsformel unter Berücksichtigung der Sterbewahrscheinlichkeit $\tilde{q}_{70}$ im Folgejahr und der Deckungsrückstellung V_{70} unter Berücksichtigung der rechnungsmäßigen Sterbewahrscheinlichkeit q_{70}, wobei die Deckungsrückstellung V_{71} am Ende des Jahres in beiden Fällen mit den rechnungsmäßigen Sterbewahrscheinlichkeiten $q_x, x \geq 71$ gebildet wird.

$$
\begin{aligned}
A &= \tilde{V}_{70} - V_{70} \\
&= R + v \cdot V_{71} \cdot (1 - \tilde{q}_{70}) - \left[R + v \cdot V_{71} \cdot (1 - q_{70}) \right] \\
&= v \cdot V_{71} \cdot \left[1 - \tilde{q}_{70} - 1 + q_{70} \right] \\
&= v \cdot V_{71} \cdot \left[q_{70} - \tilde{q}_{70} \right] \\
&= v \cdot V_{71} \cdot 0,25 \cdot q_{70}
\end{aligned}
$$

Bezogen auf eine Rente von 20.000 beträgt die Deckungsrückstellung im Alter 71 200.000, d.h. für eine Jahresrente von 1 beträgt sie 10.

Bezeichne R_S die Summe der Jahresrenten der 70-Jährigen so beträgt der erwartete Risikoverlust A_S für den Teilbestand B_T

$$
\begin{aligned}
A_S &= R_S \cdot 10 \cdot 0,25 \cdot 0,0188 \\
&= 0,047 \cdot R_S
\end{aligned}
$$

D.h. A_S beläuft sich auf 4,7% der Summe der Jahresrenten.

d) **8 Punkte**

Bezeichne zu der in Teilaufgabe a) eingeführten Notation

$u_{1-\alpha}$ das (1-α)-Quantil der Standard-Normalverteilung

$s^\alpha \in \mathbb{R}$ Schwankungsabschlag auf die Sterbewahrscheinlichkeiten unabhängig vom Alter

Der übliche Ansatz bei einer Rentenversicherung lautet:

$$\mathrm{P}\left(\sum_x Z_x \geq \sum_x (q_x - s^\alpha q_x) L_x\right) \overset{!}{=} 1-\alpha\,,$$

d.h., die Anzahl der Toten soll mit Wahrscheinlichkeit 1-α mindestens so groß sein wie die Anzahl der Toten, die man bei den mit einem altersunabhängigen Faktor s^α >0 multiplizierten Sterbewahrscheinlichkeiten erwartet.

Wegen des Zentralen Grenzwertsatzes kann angenommen werden, dass die Gesamtzahl der Toten $Z = \sum_x Z_x$ näherungsweise normalverteilt ist mit Erwartungswert $\mathrm{E}(Z) = \sum_x q_x L_x$ und Varianz $\mathrm{Var}(Z) = \sum_x q_x (1-q_x) L_x$.

Damit ergibt sich aus obiger Bedingung die äquivalente Darstellung

$$\mathrm{P}\left(Z \geq (1-s^\alpha)\cdot \mathrm{E}(Z)\right) \overset{!}{=} 1-\alpha$$

$$\Leftrightarrow \quad \mathrm{P}\left(\frac{Z-\mathrm{E}(Z)}{\sqrt{\mathrm{Var}(Z)}} \geq -s^\alpha \frac{\mathrm{E}(Z)}{\sqrt{\mathrm{Var}(Z)}}\right) \overset{!}{=} 1-\alpha$$

$$\Rightarrow \quad s^\alpha = u_{1-\alpha} \frac{\sqrt{\mathrm{Var}(Z)}}{\mathrm{E}(Z)} = u_{1-\alpha} \frac{\sqrt{\sum_x q_x (1-q_x) L_x}}{\sum_x q_x L_x}$$

Entsprechend bezeichne

$$\tilde{s}^\alpha = u_{1-\alpha} \frac{\sqrt{\sum_x \tilde{q}_x (1-\tilde{q}_x) L_x}}{\sum_x \tilde{q}_x L_x}$$

Wegen $\tilde{q}_x = a \cdot q_x$ mit 0 < a < 1 gilt

$$\begin{aligned}
\frac{\tilde{s}^\alpha}{s^\alpha} &= \frac{u_{1-\alpha}}{u_{1-\alpha}} \frac{\sum_x q_x L_x}{\sum_x \tilde{q}_x L_x} \frac{\sqrt{\sum_x \tilde{q}_x \left(1-\tilde{q}_x\right) L_x}}{\sqrt{\sum_x q_x \left(1-q_x\right) L_x}} \\
&= \frac{\sum_x q_x L_x}{a \sum_x q_x L_x} \frac{\sqrt{\sum_x a q_x \left(1-a q_x\right) L_x}}{\sqrt{\sum_x q_x \left(1-q_x\right) L_x}} \\
&= \frac{1}{\sqrt{a}} \cdot \frac{\sqrt{\sum_x q_x \left(1-a q_x\right) L_x}}{\sqrt{\sum_x q_x \left(1-q_x\right) L_x}}
\end{aligned}$$

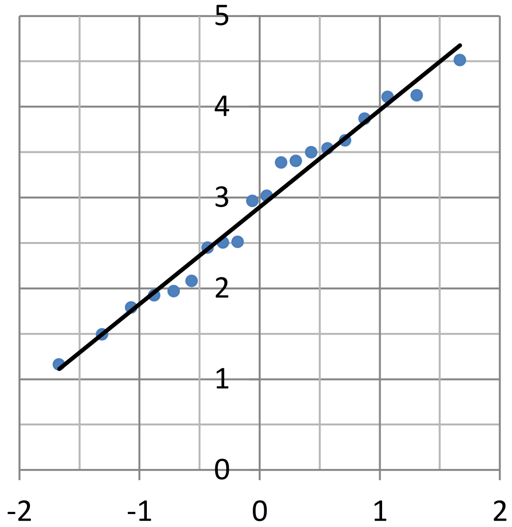
Wegen 0 < a < 1 ist $1-a\,q_x > 1-q_x$ und $\frac{1}{\sqrt{a}} > 1$, so dass gilt

$$\frac{\tilde{s}^{\alpha}}{s^{\alpha}} > 1$$

Um das gleiche Sicherheitsniveau in den modifizierten Rechnungsgrundlagen zu erreichen ist es also erforderlich, die Sicherheitsabschläge ebenfalls zu modifizieren i.e. zu erhöhen. Darüber hinaus wäre noch zu prüfen, ob z.B. aufgrund des kleineren Bestandes weitere Modifikationen erforderlich sind.

Aufgabe 2 (30 Punkte)

a) Der Aktuar eines Sachversicherer hat die Daten für den Gesamtschaden in Mio. € der letzten 20 Jahre $x_1,\dots,x_{20}$ in Sparte A grafisch aufbereitet. Zu sehen ist ein Q-Q-Plot und die zugehörige Ausgleichsgerade der Werte $\ln(x_{(k)})$ gegen die Quantile der Standard-Normalverteilung.



Wie hoch waren der höchste und der niedrigste Gesamtschaden in den letzten 20 Jahren in Sparte A?

b) Auf welche Verteilung für die Zufallsvariable X = *Jahresgesamtschaden Sparte A in Mio. €* überprüft der Aktuar mit diesem Q-Q-Plot? Warum rechtfertigt der Q-Q-Plot diese Annahme?

c) Bestimmen Sie mit Hilfe der Grafik und auf Basis von Teil b) Schätzer für die beiden Verteilungsparameter der vermuteten Verteilung von X. Erläutern Sie Ihr Vorgehen. Bei den Angaben reicht eine Nachkommastelle.

d) Berechnen Sie den Value at Risk von X zum Niveau α = 1% auf Basis der vorigen Ergebnisse. Welche für das Risikomanagement wichtige anschauliche Bedeutung hat dieser Wert?
Hinweise: Es gilt $u_{1-\alpha} = 2{,}326$. Wenn Sie Teil c) nicht gelöst haben, verwenden Sie hier und im Folgenden $X \sim LN(3;0{,}8^2)$.

e) Auf ähnliche Weise hat der Aktuar für die Zufallsvariable Y = *Jahresgesamtschaden Sparte B in Mio. €* eine Verteilung mit der Dichte

$$f(y) = 1{,}2 \cdot (1+y)^{-2{,}2}$$

($y > 0$) ermittelt. Berechnen Sie den Value at Risk von Y zum Niveau α = 1%.

f) Simulieren Sie je 2 Werte x_1,x_2 und y_1,y_2 von X und Y. Verwenden Sie dazu die folgenden Zufallszahlen $u_1,\dots,u_4$, die Realisationen von unabhängigen und auf [0,1] gleichverteilten Zufallsvariablen $U_1,\dots,U_4$ sind. Verwenden Sie die ersten beiden Zahlen für x_1,x_2 und die restlichen für y_1,y_2. Es reicht die Angabe von je drei Nachkommastellen. Verwenden Sie bei der Simulation von x_1,x_2 NICHT die Inversionsmethode.

0,131	0,367	0,439	0,878

Lösung:

a) Der kleinste Wert auf der y-Achse ist 1,16, der größte 4,51. Da es sich um logarithmierte Daten handelt, sind exp(1,16) = 3,19 Mio. € bzw. exp(4,51) = 90,92 Mio. € der kleinste bzw. größte Jahresgesamtschaden. **(3 P)**

b) Da die logarithmierten Werte gegen eine Standard-Normalverteilung getestet werden, prüft der Aktuar, ob ln(X) eine Normalverteilung besitzt, d.h. ob X logarithmisch normalverteilt ist. Die Annahme ist gerechtfertigt, da die Punkte des Q-Q-Plots nahezu auf einer Geraden liegen. **(3 P)**

c) Aus der Grafik liest man ab: Steigung der Regressionsgerade = 1,1; y-Achsenabschnitt der Regressionsgerade = 2,9. Da die Normalverteilung eine Lage-Skalen-Familie {μ + σZ} bildet mit Z ~ N(0, 1), folgt ln(X) ~ 2,9 + 1,1 · Z ~ N(2,9; $1{,}1^2$) mit den Schätzwerten μ^ = 2,9 und σ^ = 1,1. Daher ist X ~ LN(2,9; $1{,}1^2$). **(4 P)**

d) Für LN(μ,σ^2)-verteilte Zufallsvariablen X ist $\mathrm{VaR}_\alpha(X) = \exp(\mu + u_{1-\alpha}\sigma)$ (Skript Kapitel 1.4). Daher gilt hier mit dem Hinweis

$$\mathrm{VaR}_{1\%}(X) = \exp(2{,}9 + 2{,}326 \cdot 1{,}1) = \exp(5{,}4586) = 234{,}77.$$

Das 99%-Quantil des Jahresgesamtschadens beträgt also 234,77 Mio. €. Dieser Betrag wird in 99% aller Fälle vom Jahresgesamtschaden nicht überschritten. **(3 P)**

e) Wir bestimmen zunächst die Verteilungsfunktion:

$$F(y) = \int_0^y 1{,}2 \cdot (1+t)^{-2{,}2} dt = -\ [(1+t)^{-1{,}2}]_0^y = 1 - (1+y)^{-1{,}2}$$

Da es sich um eine stetige streng monotone Funktion handelt, ermittelt man den VaR mit Hilfe der Umkehrfunktion:

$$\mathrm{VaR}_\alpha(Y) = F^{-1}(1-\alpha) = (1-(1-\alpha))^{-1/1{,}2} - 1 = \alpha^{-5/6} - 1.$$

Hier also $\mathrm{VaR}_{1\%}(Y) = 0{,}01^{-5/6} - 1 = 45{,}42$. In Sparte B wird der Schaden von 45,42 Mio. € in 99% der Fälle nicht überschritten. **(8 P)**

f) Da ln(X) ~ N(2,9; $1{,}1^2$), ist zunächst eine Normalverteilung mit diesen Parametern zu simulieren. Diese wiederum kann wegen (ln(X) – 2,9) / 1,1 ~ N(0,1) durch eine Standard-Normalverteilung simuliert werden. Wir verwenden dafür die Box-Muller-Methode (Skript, Kapitel 10.4):

$$\widetilde{t_1} = \sqrt{-2\ln(u_1)} \cdot \cos(2\pi u_2) = \sqrt{-2\ln(0{,}131)} \cdot \cos(2\pi \cdot 0{,}367) = -1{,}352$$
$$\widetilde{t_2} = \sqrt{-2\ln(u_1)} \cdot \sin(2\pi u_2) = \sqrt{-2\ln(0{,}131)} \cdot \sin(2\pi \cdot 0{,}367) = 1{,}496$$

Für ln(X) ergibt dies die simulierten Werte

$$\widetilde{x_1} = 1{,}1 \cdot (-1{,}352) + 2{,}9 = 1{,}413 \quad \text{und} \quad \widetilde{x_2} = 1{,}1 \cdot 1{,}496 + 2{,}9 = 4{,}546$$

Daraus erhalten wir als gesuchte Zufallszahlen

$$x_1 = e^{1{,}413} = 4{,}108 \quad \text{und} \quad x_2 = e^{4{,}546} = 94{,}255$$

Für Y verwenden wir die Inversionsmethode (Skript, Kapitel 10.2), F^{-1} wurde bereits in e) berechnet:

$y_1 = F^{-1}(u_3) = 0{,}439^{-5/6} - 1 = 0{,}986$ $\qquad$ $y_2 = F^{-1}(u_4) = 0{,}878^{-5/6} - 1 = 0{,}115$

(9 P)

Aufgabe 3 (30 Punkte):

Zur Bearbeitung der Aufgabe sind die folgenden Quantile hilfreich:

χ^2*-Verteilung*: $\chi^2_{1;0,95} = 3,84,\ \chi^2_{2;0,95} = 5,99,\ \chi^2_{3;0,95} = 7,82,\ \chi^2_{19;0,95} = 30,14,\ \chi^2_{19;0,95} = 31,41.$

Standardnormal-Verteilung: $u_{0,9} = 1,28,\ u_{0,95} = 1,64,\ u_{0,975} = 1,96$

Sei $N : \Omega \to \mathbb{N}_0$ die Zufallsvariable „Anzahl Schäden eines Versicherungsnehmers“ und $M = \min(1, N)$ die Zufallsvariable, die angibt ob Schäden eingetreten sind.

a) Begründen Sie, dass M Bernoulli verteilt B(p)=B(1,p) ist mit $p = P(N \geq 1)$.

b) Bestimmen Sie die Informationsmatrix $I\,(p)$ von M .

c) Seien $M_1, \ldots, M_K$ unabhängig und identisch verteilt wie M. Wir betrachten den Schätzer

$\hat{p} = \frac{1}{K}\sum_{k=1}^{K} M_k$ für p. Sie können ohne Beweis davon ausgehen, dass $\hat{p}$ ein ML-Schätzer ist.

Bestimmen Sie die asymptotische Verteilung $\hat{p}$ mit Hilfe der ML-Theorie.

d) Gegeben seien unabhängige Realisierungen $m_1, \ldots, m_K$ von M. In der folgenden Tabelle sind die absoluten Häufigkeiten zusammengefasst:

Ausprägung von *M*	**Absolute Häufigkeit**
0	16
1	4

Bestimmen Sie den ML-Schätzwert $\hat{p}$ und geben Sie ein asymptotisches Konfidenzintervall zum Niveau von 95 % an. Verwenden Sie die asymptotische Verteilung aus c).

e) Überprüfen Sie mit dem Likelihood-Quotienten-Test die Hypothese $p_0 = 0{,}1$ zu einem Signifikanzniveau von zum Niveau $\alpha = 5\ \%$ für die Daten in d).

f) Ein Kollege, vermutet, dass N Poisson verteilt ist mit Parameter $\lambda > 0$ und schlägt vor einen Schätzer $\hat{\lambda}$ von λ direkt aus $\hat{p}$ zu bestimmen. Bestimmen Sie nach dieser Anleitung $\hat{\lambda}$ und die asymptotische Verteilung von $\hat{\lambda}$.

Lösungsvorschläge Aufgabe 3 (a 5, b 5, c 4, d 5, e 6, f 5)

a) Es gilt $M(\omega)=1 \Leftrightarrow \min\left(N(\omega),1\right)=1 \Leftrightarrow N(\omega)\ge 1$ also $\{M=1\}=\{N\ge 1\},\{M=0\}=\{N=0\}$.
Somit sind 0 und 1 die einzigen Realisierungen von M und $p=P\left(\{M=1\}\right)=P\left(\{N\ge 1\}\right)$.

$$p=P\left(\{M=1\}\right)=P\left(\{N\ge 1\}\right).$$

b) Für $m\in\{0,1\}$ ist die Likelihood Funktion von M gegeben durch $L(p)=p^m(1-p)^{1-m}$. Es gilt

$$\ell(p)=\ln L(p)=m\ln p+(1-m)\ln(1-p)$$

$$\ell'(p)=\frac{m}{p}-\frac{1-m}{1-p}$$

$$\ell''(p)=-\frac{m}{p^2}-\frac{1-m}{(1-p)^2}$$

$$E\left(-\ell''(p)\right)=\frac{E(M)}{p^2}+\frac{E(1-M)}{(1-p)^2}=\frac{p}{p^2}+\frac{1-p}{(1-p)^2}=\frac{1}{p}+\frac{1}{1-p}=\frac{1}{p(1-p)}$$

und damit folgt $I(p)=E\left(-\ell''(p)\right)=\dfrac{1}{p(1-p)}$.

c) Es gilt asymptotisch $\sqrt{K}\left(\hat{p}-p\right)\sim N\left(0,\dfrac{1}{I(p)}\right)=N\left(0,p(p-1)\right)$ und somit näherungsweise

$$\hat{p}\sim N\left(p,\frac{p(1-p)}{K}\right)$$

d) Einsetzen ergibt $\hat{p}=\dfrac{4}{20}=0,2.$ Setzt man diesen Wert in c) für p ein, gilt näherungsweise

$\hat{p}\sim N\left(\dfrac{4}{20},\dfrac{0,16}{20}\right)$ und man erhält das asymptotische Konfidenzintervall

$$\left(\hat{p}-u_{0,075}\sqrt{\frac{\hat{p}(1-\hat{p})}{K}},\hat{p}+u_{0,075}\sqrt{\frac{\hat{p}(1-\hat{p})}{K}}\right)=\left(\hat{p}-1,96\sqrt{\frac{0,16}{20}},\hat{p}+1,96\sqrt{\frac{0,16}{20}}\right)=\left(0,025;0,375\right).$$

e) Wegen der Unabhängigkeit gilt $X:=\sum_{k=1}^{K}M_k\sim B(K,p)$. Sei $x:=\sum_{k=1}^{K}m_k$. Damit ist

$L(p)=p^x(1-p)^{K-x}$ eine Likelihood, und die Testgröße ist $\dfrac{L(p_0)}{L(\hat{p})}=\left(\dfrac{p_0}{\hat{p}}\right)^x\left(\dfrac{1-p_0}{1-\hat{p}}\right)^{K-x}$. In d)

ist K=20, x=4. Mit $\hat{p}=0,2$, $p_0=0,1$ ergibt sich dafür $\dfrac{L(0,1)}{L(0,2)}=\left(\dfrac{1}{2}\right)^4\left(\dfrac{9}{8}\right)^{16}\approx 0,41.$ Wegen

$0,41>\exp\left(-\dfrac{\chi^2_{1;0,975}}{2}\right)\approx\exp\left(-\dfrac{3,84}{2}\right)\approx 0,15$ wird die Nullhypothese $H_0: p_0=0,2$ nicht verworfen.

f) Ist N Poisson verteilt mit Parameter $\lambda>0$, dann gilt $P(N\ge 1)=1-P(N=0)=1-e^{-\lambda}=p.$
Löst man nach λ ergibt sich $\lambda=-\ln(1-p)$ also $\hat{\lambda}=-\ln(1-\hat{p})$. Da die Funktion $h:(0,1)\ni p\mapsto -\ln(1-p)\in\mathbb{R}$ streng monoton wächst, ist $\hat{\lambda}$ eine ML-Schätzer für λ. Damit

ergibt sich die asymptotische Verteilung $\sqrt{K}\left(\hat{\lambda}-\lambda\right) \sim N\left(0, \frac{h'(p)^2}{I(p)}\right)$. Mit $h'(p) = \frac{1}{1-p}$, $I(p) = \frac{1}{p(1-p)}$ folgt $\frac{h'(p)^2}{I(p)} = \frac{p}{1-p}$ also $\hat{\lambda} \sim N\left(\lambda, \frac{p}{K(1-p)}\right)$.

Aufgabe 4 (30 Punkte):

In einem Lebensversicherungskollektiv, das zu Jahresbeginn n stochastisch unabhängige Versicherungsnehmer umfasst, bezeichne X_i den Status der Police des i–ten Versicherungsnehmers am Ende des Jahres:

$$X_i := \begin{cases} 0 & \text{falls Versicherungsnehmer } i \text{ seine Police nicht storniert hat,} \\ 1 & \text{falls Versicherungsnehmer } i \text{ seine Police storniert hat.} \end{cases}$$

Die Stornowahrscheinlichkeit sei für alle Versicherungsnehmer des Kollektivs $p \in [0; 1]$.

a) Welcher Verteilung folgt die Anzahl der Storni im Kollektiv, $N := \sum_{i=1}^{n} X_i$?

b) Sie möchten die Stornoquote $Y := \frac{N}{n}$ durch ein verallgemeinertes lineares Modell mit Dichte

$$f(y) = exp\left\{\frac{w}{\psi}\left(y\vartheta - b(\vartheta)\right) + c(y, \psi/w)\right\}$$

beschreiben.

b₁) Welche Funktion $b(\cdot)$ sollten Sie angesichts des Ergebnisses von a) wählen, und wie sollten Sie w und ψ wählen (*Angabe der Lösung genügt*)?

b₂) In welchem Zusammenhang stehen $p = E(Y)$ und ϑ (*mit Herleitung*)?

b₃) Ermitteln Sie die zugehörige kanonische Linkfunktion $g(\cdot)$ (*mit Herleitung*).

c) Das Kollektiv kann mittels des diskreten Merkmals „Tarifvariante“ (A/B) und des stetigen Merkmals „Versicherungssumme pro Police (in 10.000 EUR)“ unterteilt werden. Dabei haben Sie folgende Beobachtungen gemacht:

Beobachtung i	Tarifvariante t_i	Versicherungs-summe pro Police (in 10.000 EUR) vs_i	Anzahl Versicherungs-nehmer n_i	Stornoquote y_i
1	A	2	900	11,9%
2	A	5	500	37,8%
3	B	1	100	1,1%
4	B	5	500	7,6%

c₁) Geben Sie eine 4 x 3-Designmatrix **X** für diese Beobachtungen an und begründen Sie deren Aufbau.

c₂) Erstellen Sie einen Mittelwertplot, bei dem Sie $g(y_i)$ über vs_i abtragen und dabei jeweils einen Graphen für Tarifvariante A und einen für Tarifvariante B erzeugen. Dabei können Sie $g(y_1) = -2{,}00$, $g(y_2) = -0{,}50$, $g(y_3) = -4{,}50$ und $g(y_4) = -2{,}50$ verwenden. (*x-Achse: Wertebereich [0; 6] in Teilschritten von 1 / y-Achse: Wertebereich [-5; 0] in Teilschritten von 0,5 / ein Teilschritt entspricht 1 cm*).

c₃) Diskutieren Sie auf Basis Ihres visuellen Eindrucks die Eignung der kanonischen Linkfunktion für die vorliegenden Daten und ob Interaktionen zwischen den Tarifmerkmalen modelliert werden sollten.

c_4) Aufgrund der besonderen Datenlage können Sie hier auf die Maximum-Likelihood-Schätzung der Regressionsparameter $(\beta_1, \beta_2, \beta_3)$ verzichten und die Schätzwerte direkt aus dem Mittelwertplot ablesen. Geben Sie die Schätzwerte $\left(\widehat{\beta_1}, \widehat{\beta_2}, \widehat{\beta_3}\right)$ an und zeichnen Sie in das Schaubild ein, wo Sie diese abgelesen haben.

d) Sie möchten nun das Stornorisiko in einem Kollektiv von 1.000 Versicherungsnehmern aus Tarifvariante B mit einer Versicherungssumme von 20.000 EUR pro Police einschätzen.

d_1) Welchen Erwartungswert und welche Standardabweichung weist die Stornoquote in diesem Kollektiv auf? Runden Sie dabei auf 0,01%.

d_2) Wie hoch kann auf dieser Basis die Stornoquote im 99,5%-Quantil werden?
(*Hier können Sie davon ausgehen, dass die Stornoquote approximativ normalverteilt ist mit den Parametern aus d_1). Das 99,5%-Quantil der Standardnormalverteilung beträgt 2,58.*)

Lösung:

a) Die Anzahl der Storni im Kollektiv folgt einer Binomialverteilung $N \sim Bin(n,p)$.

1 Punkt

b) b_1) Die Stornoquote Y ist wie $\frac{1}{n} Bin(n,p)$ verteilt. Gemäß Formelsammlung des Skripts wählt man $b(\vartheta) = \ln(1+\exp(\vartheta))$. Als Gewichte setzt man die Kollektivgröße an, $w \coloneqq n$, als Dispersionsparameter $\psi \coloneqq 1$.

2 Punkte

b_2) Für GLMs gilt $p = E(Y) = \mu = b'(\vartheta) = \frac{\exp(\vartheta)}{1+\exp(\vartheta)}$.

2 Punkte

b_3) Die kanonische Linkfunktion ergibt sich durch $g(\mu) = {b'}^{-1}(\mu)$. Dazu löst man $b'(\vartheta) = \mu$ nach ϑ auf. Über den Zwischenschritt $\exp(\vartheta) = \frac{\mu}{1-\mu}$ ergibt sich $g(\mu) = \ln \frac{\mu}{1-\mu}$ (logit).

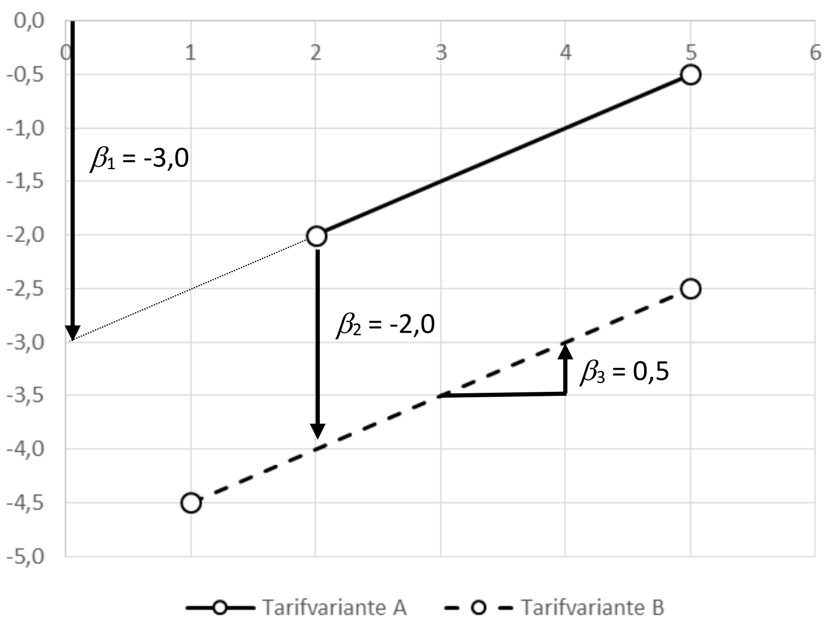
3 Punkte

c) c_1) Als Designmatrix $\mathbf{X}$ kann man $\mathbf{X} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 5 \\ 1 & 1 & 1 \\ 1 & 1 & 5 \end{pmatrix}$ wählen.

Die erste Spalte erzeugt den Achsenabschnitt. Die zweite Spalte kodiert, ob es sich bei der entsprechenden Beobachtung um eine Beobachtung aus Tarifvariante B handelt (diskretes Merkmal). Die dritte Spalte enthält die Ausprägungen des stetigen Merkmals Versicherungssumme.

4 Punkte

c_2) Für den Mittelwertplot verwendet man $g(y_1) = -2{,}00$, $g(y_2) = -0{,}50$, $g(y_3) = -4{,}50$, $g(y_4) = -2{,}50$ und erhält



5 Punkte

c_3) Die Linkfunktion g ist vom Grundsatz her dann geeignet, wenn sie im Mittelwertplot auf einen linearen Zusammenhang führt (Anmerkung: Streng genommen kann die Güte der Linkfunktion im vorliegenden Fall allerdings nicht abschließend beurteilt werden, da pro Tarifvariante lediglich 2 Datenpunkte vorliegen, die man in jedem Fall durch eine Gerade verbinden kann).

Aus der Parallelität der Graphen ergibt sich, dass keine Interaktionen zwischen den Tarifmerkmalen erkennbar sind.

3 Punkte

c_4) Aus dem Achsenabschnitt des Graphen für Tarifvariante A ergibt sich $\widehat{\beta_1} = -3{,}0$. Aus dem Abstand zwischen den Graphen für Tarifvariante A und B ergibt sich $\widehat{\beta_2} = -2{,}0$. Die Steigung der Graphen ergibt $\widehat{\beta_3} = 0{,}5$.

4 Punkte

d) d_1) Für das betrachtete Kollektiv ergibt sich als linearer Prädiktor $-3{,}0 - 2{,}0 + 0{,}5 \cdot 2 = -4{,}0$.

Die erwartete Stornoquote beträgt dem somit $\mu = \frac{\exp(-4{,}0)}{1+\exp(-4{,}0)} = 1{,}80\%$.

Die Varianz der Stornoquote berechnet sich aus der Varianzfunktion $V(\mu) = \mu(1-\mu)$ durch $\frac{\psi}{w}V(\mu) = \frac{1}{n}\mu(1-\mu)$. Hier ergibt sich die Varianz $\sigma^2 = \frac{1}{1000}1{,}80\%(1 - 1{,}80\%) = 0{,}000018$ und die Standardabweichung $\sigma = \sqrt{0{,}000018} = 0{,}42\%$.

4 Punkte

d_2) Approximiert man die Stornoquote durch eine Normalverteilung, so ergibt sich das 99,5%-Quantil als $\mu + \phi^{-1}(99{,}5\%) \cdot \sigma = 1{,}80\% + 2{,}58 \cdot 0{,}42\% = 2{,}88\%$.

2 Punkte