**Lösungen zur Klausur „Stochastische Risikomodellierung und statistische Methoden“ (Mai 2013)**

Aufgabe 1 (30 Punkte):

a) Eine Auswertung des Lebensversicherungsbestandes im Kalenderjahr 2012 zeigt folgende Ergebnisse für das Alter 40 (Altersbestimmung nach der Kalenderjahrmethode, d. h. Alter = Kalenderjahr – Geburtsjahr):

| Anzahl Personen | Dauer der Bestandszugehörigkeit in Monaten | Abgangsgrund |
|---|---|---|
| | | |
| 1 | 9 | - |
| 1 | 6 | Tod |
| 1 | 3 | Storno |
| 1 | 10 | Storno |
| 96 | 12 | - |

Ermitteln Sie die rohe Sterbewahrscheinlichkeit und die rohe Stornowahrscheinlichkeit für das Alter 40 nach der Verweildauermethode.

b) Aufgrund weiterer Auswertungen des Bestandes für die Jahre 2009 bis 2012 sowie für weitere Alter liegen Ihnen folgende rohe Sterbewahrscheinlichkeiten vor:

| | Alter im Jahr 2009 | | | | |
|---|---|---|---|---|---|
| | 30 | 31 | 32 | 33 | 34 |
| 2009 | 0,01000 | 0,02000 | 0,03000 | 0,04000 | 0,05000 |
| 2010 | 0,01980 | 0,02973 | 0,03968 | 0,04960 | 0,05964 |
| 2011 | 0,02946 | 0,03936 | 0,04930 | 0,05928 | 0,06930 |
| 2012 | 0,03905 | 0,04896 | 0,05893 | 0,06896 | 0,07904 |

d.h. die Tabelle enthält Wahrscheinlichkeiten $q_x^G$ für Alter $x = 30,...,37$ und die Geburtsjahre $G = 1975,...,1979$. Beispielsweise gilt $q_{37}^{1975} = 0,07904$. Leiten Sie die Sterbewahrscheinlichkeit eines 34-Jährigen im Jahr 2013 nach

i) dem traditionellen Modell
ii) dem Kohortenmodell

her. Schätzen Sie die erforderlichen Trendfaktoren mittels geeigneter Mittelwertbildung.

c) Zeigen Sie, dass bei einem Rententarif folgender Ansatz zur Berücksichtigung des Schwankungsrisikos sinnvoll ist: Hierbei bezeichnet α die Wahrscheinlichkeit, dass unter Berücksichtigung des altersunabhängigen Sicherheitsabschlags $s^\alpha$ (Ansatz $\left(1-s^\alpha\right)\cdot q_x$) mehr Todesfälle rechnungsmäßig erwartet werden als tatsächlich zu erwarten sind:

$$s^\alpha = u_{1-\alpha} \frac{\sqrt{\sum_x q_x\left(1-q_x\right)L_x^M}}{\sum_x q_x L_x^M},$$

Hierbei bezeichnen

| | |
|---|---|
| α | die Wahrscheinlichkeit, dass unter Berücksichtigung eines für alle Alter gleichen Sicherheitsabschlags $s^{\alpha}$ (Ansatz $\left(1-s^{\alpha}\right)\cdot q_x$) mehr Todesfälle rechnungsmäßig erwartet werden als tatsächlich zu erwarten sind |
| $L_x^M$ | die Lebenden des Alters x des Modellbestandes |
| $u_{1-\alpha}$ | das $(1-\alpha)$-Quantil der Standard-Normalverteilung |
| $s^{\alpha}\in\mathbb{R}^{+}$ | Schwankungsabschlag auf die Sterbewahrscheinlichkeit, unabhängig vom Alter *x*. |

d) Nehmen Sie an, dass zur Erzeugung der Generationentafeln aus einer Basistafel ein altersunabhängiger Faktor

$$\frac{q_x^{G+1}}{q_x^G}=a<1$$

verwendet wird. Zur Berücksichtigung des Schwankungsrisikos wird gemäß c) ein altersunabhängiger Schwankungsabschlag für eine Generation *G* bestimmt und für alle Generationentafeln einheitlich in Ansatz gebracht.

Zeigen Sie, dass bei unveränderter Alterszusammensetzung des Bestandes dann die Sicherheit für Generationentafeln mit einem späteren Geburtsjahr abnimmt.

<u>Hinweis</u>: Vergleichen Sie die für ein bestimmtes Sicherheitsniveau $1-\alpha$ erforderlichen Abschläge für die Generationen *G* und $G+1$.

<u>Lösung:</u>

a) Es bezeichne

| | |
|---|---|
| B | Beobachtungszeitraum, |
| $T_x(B)$ | Menge der Personen, die im Alter x während des Beobachtungszeitraums B wegen Tod ausgeschieden sind |
| $S_x(B)$ | Menge der Personen, die im Alter x während des Beobachtungszeitraums B wegen Storno ausgeschieden sind |
| $L_x(B)$ | Menge der Lebenden, die im Beobachtungszeitraum (auch) im Alter x sind |
| $d_{x,i}\in\left[0,1\right]$ | die Verweildauer der Person i aus $L_x(B)$ im Beobachtungszeitraum B, solange die Person im Alter x ist |

Für die Aufgabe gilt

B = [1.1.2012, 31.12.2012]
x = 40
i = 1, ..., 100

i) Ermittlung der rohen Sterbewahrscheinlichkeit $q_x$ nach der Verweildauermethode

$$q_x = \frac{|T_x(B)|}{\sum_{i \in L_x(B)} d_{x,i}} = \frac{|T_x(B)|}{|T_x(B)| + \sum_{i \in L_x(B) \setminus T_x(B)} d_{x,i}}$$

d. h. die Verweildauer derjenigen Personen, die aufgrund der zu untersuchenden Ursache ausscheiden (hier Tod), wird auf 1 gesetzt.

$$q_{40} = \frac{1}{1+\left(96+\frac{9+3+10}{12}\right)} = \frac{1}{98,833} = 0,0101$$

ii) Ermittlung der rohen Stornowahrscheinlichkeit $s_x$ nach der Verweildauermethode

$$s_x = \frac{|S_x(B)|}{\sum_{i \in L_x(B)} d_{x,i}} = \frac{|S_x(B)|}{|S_x(B)| + \sum_{i \in L_x(B) \setminus S_x(B)} d_{x,i}}$$

$$s_{40} = \frac{2}{2+\left(96+\frac{9+6}{12}\right)} = \frac{2}{99,25} = 0,0206$$

b) Gebucht ist $q_{34}^{1979}$ bzw. $q_{34,2013}$.

i)

Die Veränderung der Sterbewahrscheinlichkeit ist im traditionellen Modell nur altersabhängig nach folgendem Ansatz:

$$\frac{q_x^{G+1}}{q_x^G} = \frac{q_{x,t+1}}{q_{x,t}} = \exp(-F(x))$$

x = 34, d. h. es ist F(34) zu ermitteln bzw. zu schätzen. Dies geschieht auf Basis der $q_{x,t}$ für t = 2009, ..., 2012

| G | t | x | $q_{x,t}$ | $\frac{q_{x,t+1}}{q_{x,t}}$ | F(x) |
|---|---|---|---|---|---|
| 1975 | 2009 | 34 | 0,05 | - | - |
| 1976 | 2010 | 34 | 0,04965 | 0,993 | 0,00702 |
| 1977 | 2011 | 34 | 0,04930 | 0,993 | 0,00702 |
| 1978 | 2012 | 34 | 0,04896 | 0,993 | 0,00702 |

Ergebnis (geometrisches Mittel als Schätzer):

$$
\begin{aligned}
q_{34,2013} &= \exp\left(-F\left(34\right)\right)\cdot q_{34,2012} \\
&= \sqrt[3]{0,993^3}\cdot 0,04896 \\
&= 0,993\cdot 0,04896 \\
&= \underline{\underline{0,04861}}
\end{aligned}
$$

ii)
Die Veränderung der Sterblichkeit im Kohortenmodell ist von der Differenz zwischen Jahr und Alter abhängig nach folgendem Ansatz

$$
\frac{q_x^{G+1}}{q_x^G} = \frac{q_{x,t+1}}{q_{x,t}} = \exp\left(-H\left(t+1-x\right)\right)
$$

Es gilt x = 34 und t+1 = 2013, d. h. wegen t + 1 - x = 1979 ist H(1979) zu ermitteln bzw. zu schätzen, d. h. es sind die Kombinationen t+1 und x zu betrachten mit t+1-x = 1979

| G | t | x | $q_{x,t}$ | $q_{x,t+1}$ | $\frac{q_{x,t+1}}{q_{x,t}}$ | H(t+1-x) |
|---|---|---|---|---|---|---|
| 1979 | 2009 | 31 | 0,02000 | 0,01980 | 0,990 | 0,01005 |
| 1979 | 2010 | 32 | 0,02973 | 0,02946 | 0,991 | 0,00904 |
| 1979 | 2011 | 33 | 0,03936 | 0,03905 | 0,992 | 0,00803 |

Ergebnis (geometrisches Mittel als Schätzer):

$$
\begin{aligned}
q_{34,2013} &= \exp\left(-H\left(1979\right)\right)\cdot q_{34,2012} \\
&= \sqrt[3]{0,99\cdot 0,991\cdot 0,992}\cdot 0,04896 \\
&= 0,991\cdot 0,04896 \\
&= \underline{\underline{0,04852}}
\end{aligned}
$$

c) Bezeichne $T_x$ die Zufallsvariable der im Alter x Gestorbenen. Wegen des Zentralen Grenzwertsatzes können wir annehmen, dass die Gesamtzahl der Toten $Z := \sum_x T_x$ näherungsweise normalverteilt ist mit Erwartungswert $E(Z) = \sum_x q_x L_x$ und Varianz $Var(Z) = \sum_x q_x\left(1-q_x\right)L_x$. Es muss gelten:

$$
\begin{aligned}
& P\left(Z \geq \sum_x \left(q_x - s^\alpha q_x\right)L_x\right) \stackrel{!}{=} 1-\alpha \\
\Leftrightarrow\quad & P\left(Z \geq \left(1-s^\alpha\right)E\left(Z\right)\right) \stackrel{!}{=} 1-\alpha \\
\Leftrightarrow\quad & P\left(\frac{Z-E\left(Z\right)}{\sqrt{Var\left(Z\right)}} \geq -s^\alpha \frac{E\left(Z\right)}{\sqrt{Var\left(Z\right)}}\right) \stackrel{!}{=} 1-\alpha \\
\Rightarrow\quad & s^\alpha = u_{1-\alpha}\frac{\sqrt{Var\left(Z\right)}}{E\left(Z\right)} = u_{1-\alpha}\frac{\sqrt{\sum_x q_x\left(1-q_x\right)L_x}}{\sum_x q_x L_x}
\end{aligned}
$$

d) Bezeichne für eine Generationentafel zum Geburtsjahr G ${}^{G}s^{\alpha}$ den für das Sicherheitsniveau 1-α erforderlichen Sicherheitsabschlag. Dann gilt nach c)

$$ {}^{G}s^{\alpha} = u_{1-\alpha} \cdot \frac{\sqrt{\sum_x q_x^G \left(1-q_x^G\right) L_x}}{\sum_x q_x^G L_x} $$

bzw.

$$ {}^{G+1}s^{\alpha} = u_{1-\alpha} \cdot \frac{\sqrt{\sum_x q_x^{G+1} \left(1-q_x^{G+1}\right) L_x}}{\sum_x q_x^{G+1} L_x} $$

Wegen $q_x^{G+1} = a \cdot q_x^G$ mit $0 < a < 1$ gilt

$$ \frac{{}^{G+1}s^{\alpha}}{{}^{G}s^{\alpha}} = \frac{u_{1-\alpha}}{u_{1-\alpha}} \frac{\sum_x q_x^G L_x}{\sum_x q_x^{G+1} L_x} \frac{\sqrt{\sum_x q_x^{G+1} \left(1-q_x^{G+1}\right) L_x}}{\sqrt{\sum_x q_x^G \left(1-q_x^G\right) L_x}} $$

$$ = \frac{\sum_x q_x^G L_x}{a \sum_x q_x^G L_x} \frac{\sqrt{\sum_x a q_x^G \left(1-a q_x^G\right) L_x}}{\sqrt{\sum_x q_x^G \left(1-q_x^G\right) L_x}} $$

$$ = \frac{1}{\sqrt{a}} \cdot \frac{\sqrt{\sum_x q_x^G \left(1-a q_x^G\right) L_x}}{\sqrt{\sum_x q_x^G \left(1-q_x^G\right) L_x}} $$

Wegen $0 < a < 1$ ist $1 - a\,q_x^G > 1 - q_x^G$ und es gilt

$$ \frac{{}^{G+1}s^{\alpha}}{{}^{G}s^{\alpha}} > 1 $$

Der erforderliche Abschlag für das spätere Geburtsjahr G +1 ist größer als der für das Geburtsjahr G. Bei einem einheitlichen Abschlag für alle Geburtsjahre würde die Sicherheit also bei einem späteren Geburtsjahr abnehmen.

**Lösungen zur Klausur „Stochastische Risikomodellierung und statistische Methoden“ (Mai 2013)**

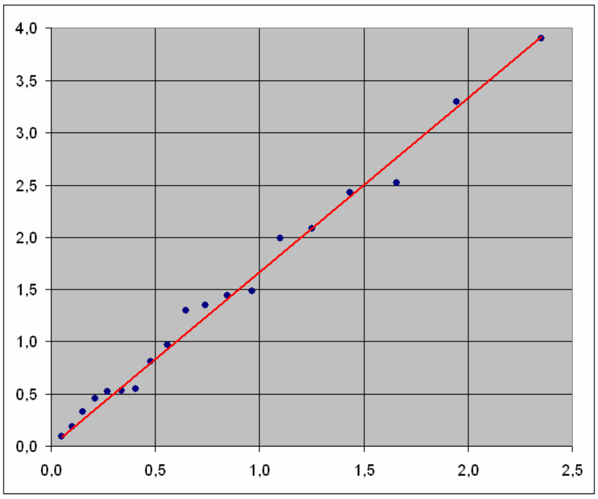
## Aufgabe 2 (25 Punkte):

Bei einer Gebäudeversicherung wurden in den letzten 31 Jahren folgende Ereigniszeitpunkte für Stürme mit Großschäden verzeichnet:

| Sturm Nr. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Datum | 13.6.1981 | 2.1.1982 | 4.5.1985 | 11.5.1987 | 23.9.1988 | 16.9.1989 | 9.1.1991 | 30.7.1993 | 15.8.1998 | 27.7.2002 |

| Sturm Nr. | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Datum | 26.11.2002 | 2.2.2003 | 19.7.2005 | 4.1.2006 | 15.7.2006 | 29.12.2007 | 2.2.2008 | 13.3.2010 | 12.9.2011 | 6.7.2012 |

Der Aktuar des Unternehmens möchte die Vermutung prüfen, dass die Ereigniszeitpunkte durch einen homogenen Poisson-Prozess modelliert werden können. Dazu rechnet er die Kalenderdaten in eine zeitstetige Skala um (Beginn: 1.1.1981, Zeiteinheit: 1 Kalenderjahr) und trägt die Zwischenereigniszeiten ( = Verweildauern) in folgendem Q-Q-Plot auf.




a) Was wird bei korrektem Vorgehen auf den beiden Achsen abgetragen? Warum gibt es hier nur 19 Datenpunkte?

b) Was für ein Regressionsmodell ist hier sinnvoll?

c) Warum kann auf Grund dieser Graphik die Hypothese eines homogenen Poisson-Prozesses akzeptiert werden?

d) Ermitteln Sie aus der Graphik einen Schätzer für den Poisson-Parameter $\lambda$.

e) Schätzen Sie damit die Wahrscheinlichkeit dafür ein, dass innerhalb von zwei Kalenderjahren kein Sturmereignis mit einem Großschaden eintritt.

Lösung: Der homogene Poisson-Prozess ist dadurch charakterisiert, dass seine Zwischenereigniszeiten ( = Verweildauern) stochastisch unabhängige, $\mathcal{E}(\lambda)$-exponentialverteilte Zufallsvariablen sind.

a) Die $\mathcal{E}(\lambda)$-Verteilung gehört zu einer (reinen) Skalenfamilie mit $\sigma = \frac{1}{\lambda}$. Auf der Abszisse werden daher die Quantile der $\mathcal{E}(1)$-Verteilung abgetragen, bei $n$ Daten sind das gerade die Größen $-\ln(1-u_k) = -\ln\left(1-\frac{k}{n+1}\right) = \ln\left(\frac{n+1}{n+1-k}\right)$ für $k = 1,\cdots,n$. Auf der Ordinate werden die der Größe nach sortierten beobachteten Zwischenereigniszeiten $x_{(k)}$ abgetragen.

   Es liegen zwar 20 beobachtete Zeitpunkte $T_1,\cdots,T_{20}$ vor, aber es gibt nur 19 Zwischenereigniszeiten $X_k = T_{k+1} - T_k$, $k = 1,\cdots,19$.

b) Eine lineare Regression durch den Nullpunkt, da es sich um eine (reine) Skalenfamilie handelt.

c) Nach visueller Beurteilung ist die Anpassung an eine Gerade (durch den Nullpunkt) hinreichend gut.

d) Als Schätzer $\hat{\sigma}$ für den Parameter $\sigma$ verwendet man die Steigung der Regressionsgeraden; der Graphik ist zu entnehmen, dass die Gerade durch den Punkt (1,5 | 2,5) verläuft, so dass sich hier $\hat{\sigma} = \frac{2,5}{1,5} = \frac{5}{3} = 1,\overline{6}$ und damit $\hat{\lambda} = \frac{1}{\hat{\sigma}} = \frac{3}{5} = 0,6$ ergibt.

e) Das betrachtete Ereignis hat die Poisson-Wahrscheinlichkeit $e^{-2\lambda} = 0,3012,$ das entspricht der Wahrscheinlichkeit dafür, dass der (homogene) Poisson-Prozess in einem Intervall der Länge 2 keinen Zuwachs hat (d.h. in dieser Zeitspanne keine Sturmereignisse eintreten).

**Lösungen zur Klausur „Stochastische Risikomodellierung und statistische Methoden“ (Mai 2013)**

Aufgabe 3 (30 Punkte):

$(\Omega, \mathcal{A}, \mathbb{P})$ sei der zugrundeliegende Wahrscheinlichkeitsraum. Für einen Bestand mit Volumen $v \in \mathbb{N}$ wird die Schadenzahl $N$ mit der Zufallsvariablen $N : \Omega \to \mathbb{N}_0$ und der Frequenz $Z : \Omega \to \left\{ \frac{k}{v} \middle| k \in \mathbb{N}_0 \right\}$, $Z := \frac{N}{v}$ modelliert. Dabei wird $N \sim \mathcal{P}(\lambda v)$ mit $\lambda > 0$ als Poisson-verteilt angenommen. Beobachtet werden unabhängige Realisierungen von $Z_1, \dots, Z_n$ (Frequenzen) und Volumina $v_1, \dots, v_n > 0$.

a) i) Bestimmen Sie die gemeinsame Wahrscheinlichkeitsfunktion (also die gemeinsame Zähldichte) von $(Z_1, \dots, Z_n)$ und die Fisherinformation $I_n(\lambda)$ von $(Z_1, \dots, Z_n)$.

*Zwischenergebnis:* $I_n(\lambda) = \frac{\sum_{i=1}^{n} v_i}{\lambda}$, $\lambda \mapsto \sum_{i=1}^{n} (-\lambda v_i + z_i v_i \ln \lambda)$ *ist eine Loglikelihood*

ii) Beweisen Sie, dass $\hat{\lambda}_n = \frac{\sum_{i=1}^{n} v_i Z_i}{\sum_{i=1}^{n} v_i}$ ein Maximum-Likelihood-Schätzer von $\lambda$ ist.

Mit obigen Ergebnissen sollen Erwartungswert und Varianz von $N$ für einen Bestand mit Volumen $v$ geschätzt werden.

b) Bestimmen Sie $E(N)$, $Var(N)$ und begründen Sie, dass $\hat{\lambda}_n v$ bzw. $\sqrt{\hat{\lambda}_n v}$ Maximum-Likelihood-Schätzer des Erwartungswerts bzw. der Standardabweichung von $N$ sind.

c) Ohne Beweis wird angenommen, dass Satz 5.3.9 des Folienskripts für $\hat{\lambda}_n$ gilt, dass also $\hat{\lambda}_n$ näherungsweise $\mathcal{N}\left(\lambda, \frac{1}{I_n(\lambda)}\right)$-verteilt ist. Geben Sie die asymptotische Verteilung von $\sqrt{\hat{\lambda}_n v}$ an.

Gegeben seien die folgenden Beobachtungen:

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | $\Sigma$ |
|---|---|---|---|---|---|---|---|
| $z_i$ | $\frac{2}{7}$ | $\frac{1}{3}$ | $\frac{3}{11}$ | 0,1 | $\frac{1}{3}$ | 0,125 | |
| $v_i$ | 14 | 12 | 11 | 10 | 9 | 8 | 64 |
| $v_i z_i$ | 4 | 4 | 3 | 1 | 3 | 1 | 16 |

d) Bestimmen Sie für einen Bestand mit Volumen $v = 16$ mit obigen Daten die Schätzwerte für $E(N)$ und $\sqrt{Var(N)}$.

e) Geben Sie für einen Bestand mit Volumen $v = 16$ mit Hilfe der Maximum-Likelihood-Theorie ein asymptotisches 90%-Schätzintervall für die Standardabweichung von $N$ unter der Verwendung der Daten und d) an.

*Quantile von* $\mathcal{N}(0,1)$: $u_{0,9} = 1{,}28$ $u_{0,95} = 1{,}64$ $u_{0,975} = 1{,}96$ $u_{0,99} = 2{,}33$

<u>Lösung:</u> Zu a) i)

Es gilt für $(z_1,\dots,z_n)\in\left\{\left(\frac{k_1}{v_1},\dots,\frac{k_n}{v_n}\right)\middle| k_1,\dots,k_n\in\mathbb{N}_0\right\}$ (also $z_i v_i\in\mathbb{N}_0,\ i=1,\dots,n$)

$$f(z_1,\dots,z_n,\lambda)=\prod_{i=1}^n e^{-\lambda v_i}\frac{(\lambda v_i)^{z_i v_i}}{(z_i v_i)!}$$

$$\ell_n(\lambda)=\sum_{i=1}^n\Big(-\lambda v_i+z_i v_i\big(\ln(\lambda)+\ln(v_i)\big)-\ln\big((z_i v_i)!\big)\Big)$$

oder kürzer auch

$$\ell_n(\lambda)=\sum_{i=1}^n\left(-\lambda v_i+z_i v_i\ln\lambda\right)$$

$$\frac{d}{d\lambda}\ell_n(\lambda)=\sum_{i=1}^n\left(-v_i+\frac{z_i v_i}{\lambda}\right)$$

$$\frac{d^2}{d\lambda^2}\ell_n(\lambda)=-\sum_{i=1}^n\frac{z_i v_i}{\lambda^2}$$

$$I_n(\lambda)=E\left(-\frac{d^2}{d\lambda^2}\ell_n(\lambda,Z_1,\dots,Z_n)\right)=E\left(\sum_{i=1}^n\frac{Z_i v_i}{\lambda^2}\right)=\sum_{i=1}^n\frac{\lambda v_i}{\lambda^2}=\frac{\sum_{i=1}^n v_i}{\lambda}$$

Zu a) ii)

Aus der dritten Zeile von i) folgt durch Nullsetzen

$$\sum_{i=1}^n v_i=\sum_{i=1}^n\frac{z_i v_i}{\hat\lambda}\Rightarrow\hat\lambda_n=\frac{\sum_{i=1}^n z_i v_i}{\sum_{i=1}^n v_i}.$$

Wegen $\frac{d^2}{d\lambda^2}\ell_n(\lambda)=-\sum_{i=1}^n\frac{z_i v_i}{\lambda^2}<0$ liegt in $\hat\lambda_n$ ein Maximum von $\ell_n$ vor.

Zu b)

Laut Voraussetzung gilt $N\sim\mathcal{P}(\lambda v)$ und somit $E(N)=\lambda v,\ Var(N)=\lambda v$. Die Funktionen $h:(0,\infty)\to(0,\infty),\ h(\lambda)=\lambda v$ und $g:(0,\infty)\to(0,\infty),\ g(\lambda)=\sqrt{\lambda v}$ sind bijektiv und stetig differenzierbar. Damit sind $h(\hat\lambda_n)$ und $g(\hat\lambda_n)$ Schätzer von $h(\lambda)=\lambda v$ und $g(\lambda)=\sqrt{\lambda v}$ .

Zu c)

Verwende die Bezeichnungen von b). Es gilt $g'(\lambda v)=\frac{v}{2\sqrt{\lambda v}}=\frac{\sqrt{v}}{2\sqrt{\lambda}}$ und es folgt, dass asymptotisch

$$g(\hat\lambda_n)\sim\mathcal{N}\left(g(\lambda),\frac{g'(\lambda)^2}{I_n(\lambda)}\right)=\mathcal{N}\left(g(\lambda),\frac{v}{4\sum_{i=1}^n v_i}\right)\text{ wegen }\frac{g'(\lambda)^2}{I_n(\lambda)}=\frac{v}{4\lambda}\cdot\frac{\lambda}{\sum_{i=1}^n v_i}=\frac{v}{4\sum_{i=1}^n v_i}\text{ gilt.}$$

Zu d)

Mit a) und $n = 6$ gilt $\hat{\lambda}_n = \frac{1}{4}$.

Mit c) folgt $\widehat{E(N)} = \hat{\lambda}_n \cdot v = 4$, $\sqrt{\widehat{Var(N)}} = \sqrt{\hat{\lambda}_n \cdot v} = 2$.

Zu e)

Mit d) und c) ergibt sich, indem man die Schätzwerte einsetzt, die Varianz $\frac{v}{4\sum_i^n v_i} = \frac{16}{4 \cdot 64} = \frac{1}{16}$, folglich als 90 % Schätzintervall $\left( g\left(\hat{\lambda}\right) - \frac{u_{0,95}}{4}, g\left(\hat{\lambda}\right) + \frac{u_{0,95}}{4} \right) = \left( 2 - \frac{1,64}{4}, 2 + \frac{1,64}{4} \right) = (1,59; 2,41)$.

**Lösungen zur Klausur „Stochastische Risikomodellierung und statistische Methoden“ (Mai 2013)**

Aufgabe 4 (35 Punkte):

$X_1,\cdots,X_{100}$ seien die unabhängigen Jahresgesamtschäden, welche die Versicherungsnehmer $i=1,\cdots,100$ im Jahr 2012 verursacht haben. Beobachtet wurde das arithmetische Mittel $\hat{\mu}=12{,}95$ und die empirische Standardabweichung $\hat{\sigma}=16{,}07$ dieser Jahresgesamtschäden (Werte in Einheiten von 10 EUR). Aus dem Markt sei bekannt, dass $X_i$ einer $\mathcal{E}(\lambda_i)$-Exponentialverteilung folgt, mit jeweils individuellem, unbekanntem Strukturparameter $\lambda_i \in (0;1]$.

a) Unter der Nullhypothese, dass die Strukturparameter aller Versicherungsnehmer identisch sind $(\lambda_1=\lambda_2=\ldots=\lambda_{100}=:\lambda)$, folgen die Schäden aller Versicherungsnehmer einer $\mathcal{E}(\lambda)$-Verteilung.

- Berechnen Sie den theoretischen Varianzquotienten $v:=\dfrac{Var(X_i)}{\left[E(X_i)\right]^2}$ dieser Verteilung und vergleichen Sie diesen mit dem empirischen Varianzquotienten $\hat{v}:=\dfrac{\hat{\sigma}^2}{\hat{\mu}^2}$ der Daten. Welche Alternativhypothese legt der Vergleich nahe?

- Testen Sie zum Niveau 5%, ob die Nullhypothese identischer Strukturparameter haltbar ist. Dabei können Sie ohne Beweis verwenden, dass der empirische Varianzquotient $\hat{v}$ unter der Nullhypothese mit einer Wahrscheinlichkeit von 95% Werte im Intervall $[0{,}68;1{,}44]$ annimmt.

b) Um der Möglichkeit Rechnung zu tragen, dass sich die unbeobachteten Strukturparameter für die einzelnen Versicherungsnehmer unterscheiden, gehen Sie nun davon aus, dass jeder Strukturparameter $\lambda_i$ Realisierung einer Zufallsvariablen $\Lambda_i:=\dfrac{1}{U_i}$ ist. Dabei ist $U_i$ auf dem Intervall $[1;2\mu-1]$ gleichverteilt mit $\mu=10{,}0.$ Berechnen Sie die lineare Credibility-Prämie für einen Versicherungsnehmer *i*, für den in den letzten drei Jahren (*n* = 3) die Jahresgesamtschäden 12,0 sowie 7,0 und 8,0 beobachtet wurden.

*<u>Hinweis:</u> Sie können (ohne Beweis) verwenden, dass die lineare Credibility-Prämie durch* $H_i^{**}:=z_n\bar{X}_i+(1-z_n)E(X_i)$ *definiert ist, mit* $z_n=\dfrac{Var(E[X_i\,|\,\Lambda_i])}{\frac{1}{n}E(Var[X_i\,|\,\Lambda_i])+Var(E[X_i\,|\,\Lambda_i])}$. *Außerdem gilt* $E(U_i)=\mu$ *und* $Var(U_i)=\dfrac{(\mu-1)^2}{3}$.

c) Für die Versicherungsnehmer (VN) 1 und 2 liege eine Schadenhistorie von insgesamt drei Jahren vor:

| Jahr $t$ | Schaden $X_1(t)$ von VN 1 | Schaden $X_2(t)$ von VN 2 |
|---|---|---|
| 0 (entspricht 2010) | 2,0 | 6,0 |
| 1 (entspricht 2011) | 7,0 | 9,0 |
| 2 (entspricht 2012) | 8,0 | 13,0 |

Sie möchten ein verallgemeinertes lineares Modell mit

$$f_i(x_i) = \exp\left(\frac{1}{\psi}(x_i\vartheta_i - b(\vartheta_i)) + c(x_i, \psi)\right)$$

an diese Schadenbeobachtungen $X_i(t)$ anpassen.

- Zeigen Sie, dass sich für $b(\vartheta_i) = -\ln(-\vartheta_i)$, $\psi = 1$ und $c = 0$ die $\mathcal{E}(-\vartheta_i)$-Verteilung ergibt.
- Die Exponentialverteilung ist ein Spezialfall der Gamma-Verteilung. Geben Sie mit diesem Wissen die zur Exponentialverteilung zugehörige kanonische Link-Funktion $g(\mu)$ und die Varianzfunktion $V(\mu)$ an.

d) Ein Statistikprogramm liefert folgendes Schätzergebnis für das verallgemeinerte lineare Modell aus c) mit der Versicherungsnehmernummer als diskreter und dem Jahr *t* als stetiger Kovariate:

<table>
<tr><td colspan="3"><u>Modellspezifikation</u><br>Verteilung: Exponential<br>Linkfunktion: kanonisch</td></tr>
<tr><td><u>Beobachtungsdaten</u></td><td><u>Designmatrix</u></td><td><u>Schätzergebnis</u><br><u>Regressionsparameter</u></td></tr>
<tr><td>$\begin{pmatrix} 2 \\ 7 \\ 8 \\ 6 \\ 9 \\ 13 \end{pmatrix}$</td><td>$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$</td><td>$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0,2481 \\ -0,0623 \\ -0,0589 \end{pmatrix}$</td></tr>
</table>

- Ermitteln Sie für Versicherungsnehmer 2 die Nettoprämie $E\left(X_2(2)\right)$ im Jahr 2 und berechnen Sie daraus eine Schätzung für den unbeobachteten Strukturparameter $\lambda_2$.
- Berechnen Sie die Varianz $Var\left(X_2(2)\right)$ des zugehörigen Schadens.

e) Das Statistikprogramm liefert nun zusätzlich eine Schätzung $\hat{\psi} = 0,1358$ für den Parameter $\psi$. Geben Sie damit eine neue Schätzung für $E\left(X_2(2)\right)$ und $Var\left(X_2(2)\right)$ an und begründen Sie diese.

<u>Lösung:</u>

a) Für die $\mathcal{E}(\lambda)$-Verteilung gilt $E(X_i)=\frac{1}{\lambda}$ und $Var(X_i)=\frac{1}{\lambda^2}$, so dass der theoretische Varianzquotient $v=\frac{Var(X_i)}{[E(X_i)]^2}=1$ beträgt.

Dem gegenüber steht ein empirischer Varianzquotient von $\hat{v}=\frac{16{,}07^2}{12{,}95^2}=1{,}54>1.$

Unter der Nullhypothese $\lambda_1 = \lambda_2 = \ldots = \lambda_{100}$ ist der empirische Varianzquotient ein Schätzer für den theoretischen Varianzquotienten. Aufgrund der Abweichung zwischen empirischem und theoretischem Varianzquotienten liegt die Alternativhypothese nahe, dass nicht alle Strukturparameter $\lambda_i$ identisch sind.

Aufgrund des Hinweises ist $\hat{v}$ eine Teststatistik und $[0{,}68;\, 1{,}44]^c=(-\infty;\, 0{,}68)\cup(1{,}44;\, \infty)$ ein Verwerfungsbereich für einen Test der Nullhypothese zum Niveau 5%. Da $\hat{v}=1{,}54$ im Verwerfungsbereich liegt, wird die Nullhypothese identischer Strukturparameter verworfen.

b) Es gilt

$$\overline{X}_i=\frac{12{,}0+7{,}0+8{,}0}{3}=9{,}0,\; Var(E[X_i\,|\,\Lambda_i])=Var\left(\frac{1}{\Lambda_i}\right)=Var(U_i)=\frac{1}{3}(\mu-1)^2=81/3=27,$$

$$E(Var[X_i\,|\,\Lambda_i])=E\left(\frac{1}{\Lambda_i^2}\right)=E\left(U_i^2\right)=Var(U_i)+\{E(U_i)\}^2=\frac{1}{3}(\mu-1)^2+\mu^2 = 27+100=127 \text{ und}$$

$$E(X_i)=E\left[E(X_i|\Lambda_i)\right]=E\left(\frac{1}{\Lambda_i}\right)=E(U_i)=\mu=10.$$

(Bemerkung: Aus diesen Werten ergibt sich als Varianzquotient

$$v:=\frac{Var(X_i)}{[E(X_i)]^2}=\frac{E(Var[X_i\,|\,\Lambda_i])+Var(E[X_i\,|\,\Lambda_i])}{\mu^2}=\frac{127+27}{100}=1{,}54,$$ also der in a) berechnete Wert, so dass $\mu=10{,}0$ eine sinnvolle Wahl ist.)

Für den Credibility-Faktor ergibt sich damit $z_n=\frac{27}{27+127/3}=0{,}3894.$

Mit diesen Werten berechnet sich die lineare Credibility-Prämie zu

$$H_i^{**}=z_n\overline{X}_i+(1-z_n)E(X_i)=0{,}3894\cdot 9{,}0+0{,}6106\cdot 10{,}0=9{,}61.$$

c) Einsetzen von $b(\vartheta_i)=-\ln(-\vartheta_i)$, $\psi=1$ und c = 0 ergibt die Dichte

$$f_i(x_i)=\exp\left(x_i\vartheta_i+\ln(-\vartheta_i))\right)=-\vartheta_i\exp\left(x_i\vartheta_i\right)=-\vartheta_i\exp\left(-x_i(-\vartheta_i)\right)$$

der $\mathcal{E}(-\vartheta_i)$-Verteilung.

Für allgemeines $\psi$ ergibt sich mit $b(\vartheta) = -\ln(-\vartheta)$ die Gamma-Verteilung, so dass die kanonische Linkfunktion $g(\mu) = \frac{1}{\mu}$ ist und die Varianzfunktion $V(\mu) = \mu^2$.

d) Der zugehörige lineare Prädiktor ist $(1; 1; 2) \cdot (\beta_1, \beta_2, \beta_3)^T = 0{,}2481 - 0{,}0623 - 2 \cdot 0{,}0589 = 0{,}0680$. Mit der kanonischen Linkfunktion berechnet sich daraus der Erwartungswert $E(X_2(2)) = 1/0{,}0680 = 14{,}7059$. Zwischen dem Strukturparameter $\lambda_2$ und dem Erwartungswert $E(X_2(2))$ besteht im Rahmen der Exponentialverteilung der Zusammenhang $E(X_2(2)) = \frac{1}{\lambda_2}$, so dass $\lambda_2 = 0{,}0680$.

Die Varianz ergibt sich mit der Varianzfunktion (oder direkt mit der Exponentialverteilung) gemäß $Var(X_2(2)) = V(E(X_2(2))) = \{E(X_2(2))\}^2 = 14{,}7059^2 = 216{,}2635$.

e) Der konkrete Wert von $\psi$ hat keinen Einfluss auf die Maximum-Likelihood-Parameterschätzung für die Regressionsparameter $(\beta_1, \beta_2, \beta_3)$, so dass die Schätzung für den Erwartungswert unverändert $E(X_2(2)) = 14{,}7059$ beträgt. In die Berechnung der Varianz geht $\psi$ gemäß $Var(X_2(2)) = \psi \cdot V(E(X_2(2)))$ ein, so dass eine verbesserte Schätzung der Varianz $Var(X_2(2)) = 0{,}1358 \cdot V(E(X_2(2))) = 0{,}1358 \cdot 216{,}2635 = 29{,}3686$ beträgt.

Verteilung der Punktzahlen auf die Unteraufgaben:

| | |
|---|---|
| 1 a) | 8 |
| 1b) | 10 |
| 1c) | 6 |
| 1d) | 6 |
| 2a) | 10 |
| 2b) | 3 |
| 2c) | 2 |
| 2d) | 5 |
| 2e) | 5 |
| 3a) | 10 |
| 3b) | 5 |
| 3c) | 5 |
| 3d) | 5 |
| 3e) | 5 |
| 4a) | 8 |
| 4b) | 10 |
| 4c) | 6 |
| 4d) | 8 |
| 4e) | 3 |