Blätter DGVFM
DOI 10.1007/s11857-009-0092-4

ACTUARIAL EXAMS

Bericht zur Prüfung im Mai 2009 über stochastische Risikomodellierung und statistische Methoden (Grundwissen)

Richard Herrmann · Dietmar Pfeifer · Viktor Sandor · Gerald Sußmann

Received: 9 September 2009 / Accepted: 9 September 2009

Am 9. Mai 2009 fand die DAV-Prüfung im Fach Stochastische Risikomodellierung und statistische Methoden (Grundwissen) statt. Von den 216 Teilnehmerinnen und Teilnehmern haben 180 bestanden. Die Klausur wurde als bestanden gewertet, wenn mindestens 40 von 120 möglichen Punkten erreicht wurden.

Aufgabe 1 (32 Punkte)

Zwei Lebensversicherungsunternehmen A und B fusionieren. Im Rahmen der Fusion werden die Bestände der gemischten Versicherungen zu einer Risikogemeinschaft zusammengeführt, da sie (angeblich) gleiche Risiken darstellen. Als Aktuar werden Sie um Antworten zu folgenden Fragen gebeten:

a) Sind die biometrischen Risiken in beiden Beständen gleich?
Als Unterlagen haben Sie für beide Bestände und die letzten 5 Jahre (2004–2008) folgende Informationen zur Verfügung:

1. Versichertenbestand mit individuellen Angaben (Geschlecht, Geburtsdatum, Versicherungsbeginn, Versicherungssumme, Deckungskapital) zum Bilanzstichtag (31.12.),
2. Liste der Abgänge im jeweiligen Geschäftsjahr mit den Angaben unter 1. und zusätzlich nur den Abgangsgrund Tod oder Ablauf der Versicherung (Storno wird vernachlässigt).

(i) Stellen Sie eine Methode dar, mit der Sie die Unterschiede im Risikoverhalten der beiden Bestände analysieren. Verwenden Sie zur Altersbestimmung die Kalenderjahrsmethode (Alter x = Kalenderjahr − Geburtsjahr).
(ii) Welche zusätzlichen Informationen würden Sie benötigen, um ein genaueres Verfahren anwenden zu können. Stellen Sie dieses dar und vergleichen Sie es mit Ihrem Verfahren unter (i).

Viktor Sandor, Rosenheim, Deutschland, E-Mail: sandor@fh-rosenheim.de

Published online: 10 November 2009

(iii) Welche Methoden kennen Sie, um Unterschiede zwischen den beiden Beständen zu erkennen? Wenden Sie eine dieser Methoden auf das folgende Beispiel an.

Alter x	q_x^A	q_x^B
30	1,0%	0,5%
31	1,2%	2,0%
32	0,5%	1,0%
33	1,0%	1,1%
34	1,5%	1,4%

b) Welche Einsparungen ergeben sich beim Sicherheitskapital?
Nehmen Sie an, dass das Schwankungsrisiko in beiden Beständen durch ein entsprechendes Sicherheitskapital berücksichtigt wird. Das 95%-Sicherheitskapital (Value at Risk) zur Abdeckung des Schwankungsrisikos ist in beiden VU vorhanden und beträgt für VU A 1 Mio € und für VU B 5 Mio €.
Nehmen Sie an, dass das Sicherheitskapital in beiden Beständen mit Hilfe der Normalverteilung errechnet wurde. Welches Sicherheitskapital ergibt sich dann nach dem Zusammenlegen der Bestände? Die Standardabweichungen sind 1,5 Mio € und 2 Mio €; das 95%-Quantil der Standardnormalverteilung beträgt $u_{0,95} = 1{,}645$.

Lösung

Zu a) (i)

Ermittlung der relativen Sterbehäufigkeiten aus den Beständen der letzten 5 Jahre.
Da es sich um einen offenen Personenbestand handelt, ist nur die Verweildauermethode geeignet, um die Sterbehäufigkeiten exakt zu ermitteln. Zur Anwendung der Verweildauermethode fehlen jedoch in den Bestandsdaten der Todeszeitpunkt und das Ablaufdatum. Daher kommen die folgenden Methoden in Betracht:
(nur 1 Methode wird verlangt)

Geburtsjahrmethode

Aufgrund der Kalenderjahrmethode zur Altersbestimmung wird jedem Versicherten unabhängig vom unterjährigen exakten Todesdatum in Abhängigkeit vom Geburtsjahr das Alter $x = \text{Kalenderjahr} - \text{Geburtsjahr}$ zugeordnet.

Bezeichne

L	eine Personengesamtheit, unter Risiko stehend
T	Tote aus L
$q = \frac{\lvert T\rvert}{\lvert L\rvert}$	relative Sterbehäufigkeit = rohe Sterbewahrscheinlichkeit
$L' = L\backslash T =$	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.

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.

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

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

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)

Als exakte Methode steht nur die Verweildauermethode zur Verfügung. Zur Anwendung der Verweildauermethode sind jedoch zusätzliche Angaben erforderlich

- Todesdatum
- Datum des Versicherungsablaufs

Nach Vorliegen dieser Angaben ist die Verweildauermethode anwendbar.
Unter Vorliegen dieser Informationen kann das Alter x als das bürgerliche Alter bestimmt werden.
Bezeichne $d_{x,i}$ $(0 \le d_{x,i} \le 1)$ die Verweildauer der Person i im Alter x in der Personengesamtheit. Die rohen Sterbewahrscheinlichkeiten werden definiert durch

$$q_x^v = \frac{|T_x\,(B)|}{\sum\limits_{i \in L(B)} d_{x,i}} = \frac{|T_x\,(B)|}{|T_x\,(B)| + \sum\limits_{i \in L'(B)} d_{x,i}}\,.$$

$d_{x,i}$ gibt den Zeitraum an, in dem die Person i im Alter x unter Risiko steht, d. h.

$d_{x,i} \leq 1$ falls die Versicherung abläuft oder bei unterjährigem Versicherungsbeginn

$d_{x,i} = 1$ falls der Tod im Alter x eintritt oder die Versicherung im ganzen Jahr unter Risiko stand.

(iii)

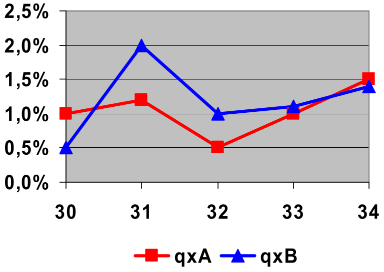
Die Unterschiede zwischen den beiden Beständen können durch den Vergleich der empirischen Sterbewahrscheinlichkeiten mithilfe eines Testverfahrens ermittelt werden.
Dazu wird die Nullhypothese H_0 bzw. die Alternativhypothese H_1 formuliert:
H_0: Die Sterbewahrscheinlichkeiten stimmen in beiden Versichertenbeständen überein.
H_1: Die Sterbewahrscheinlichkeiten in beiden Versichertenbeständen sind verschieden.
Mögliche Testverfahren:

- Vorzeichentest
- Iterationstest

(Es genügt, nur ein Testverfahren darzustellen)

Beispiel:

x	q_x^A	q_x^B
30	1,0%	0,5%
31	1,2%	2,0%
32	0,5%	1,0%
33	1,0%	1,1%
34	1,5%	1,4%



Vorzeichentest

Geht man davon aus, dass in beiden Beständen die Sterbewahrscheinlichkeiten übereinstimmen, dann kann man annehmen, dass bei den Differenzen zwischen Sterbehäufigkeiten beider Bestände gleich viele positive wie negative Vorzeichen auftreten.
Teststatistik:

$$T = \sum_{x=30}^{34} 1_{\left\{q_x^A > q_x^B\right\}}$$

D. h., man zählt die positiven Vorzeichen, die sich bei den Differenzen ergeben.

Unter Gültigkeit der Nullhypothese H_0 ist T binomialverteilt mit Wahrscheinlichkeit $1/2$ für jedes Vorzeichen, d. h. es gilt

$$T \sim B\left(5, \frac{1}{2}\right).$$

Zu einem vorgegebenen Signifikanzniveau α werden dann zwei Schwellenwerte n_α und $5 - n_\alpha$ bestimmt, so dass die Nullhypothese H_0 abgelehnt wird, wenn der Wert der Teststatistik n_α unterschreitet bzw. $5 - n_\alpha$ überschreitet.
Dabei wird n_α wegen der Symmetrie der Binomialverteilung bestimmt aus

$$2 \cdot \mathrm{P}\left(T < n_\alpha\right) = 2 \cdot \sum_{j=0}^{n_\alpha - 1} \binom{5}{j} \cdot \left(\frac{1}{2}\right)^j \cdot \left(\frac{1}{2}\right)^{5-j} \overset{!}{\leq} \alpha\,.$$

Iterationstest

Wenn man davon ausgeht, dass in beiden Beständen die Sterbewahrscheinlichkeiten übereinstimmen, dann kann man folgern, dass die beobachteten Wahrscheinlichkeiten des einen Bestandes „mal größer und mal kleiner" als die des anderen sind, d. h. dass also viele Vorzeichenwechsel auftreten.
Teststatistik:

$$T = \sum_{x=31}^{34} 1_{\left\{Sign\left(q_x^A - q_x^B\right) \neq Sign\left(q_{x-1}^A - q_{x-1}^B\right)\right\}}$$

Wie beim Vorzeichentest bildet man auch beim Iterationstest zunächst die Differenzen zwischen den Sterbehäufigkeiten desselben Alters. Im Anschluss zählt man die aufgetretenen Vorzeichenwechsel.
Unter der Gültigkeit der Nullhypothese H_0 ist T binomialverteilt mit Wahrscheinlichkeit $1/2$ für jeden Vorzeichenwechsel, d. h. es gilt

$$T \sim B\left(4, \frac{1}{2}\right).$$

Zu vorgegebenem Signifikanzniveau α wird ein Schwellenwert n_α so bestimmt, dass die Nullhypothese abgelehnt wird, wenn der Wert der Teststatistik n_α unterschreitet.
Dabei wird n_α bestimmt aus

$$\mathrm{P}(T < n_\alpha) = \sum_{j=0}^{n_\alpha - 1} \binom{4}{j} \cdot \left(\frac{1}{2}\right)^j \cdot \left(\frac{1}{2}\right)^{4-j} \overset{!}{\leq} \alpha.$$

Zu b)

Bezeichne S^A den Schaden im Bestand A und S^B den im Bestand B.
Dann gilt näherungsweise $S^A \sim \mathcal{N}(\mu_A, \sigma_A^2)$ und $S^B \sim \mathcal{N}(\mu_B, \sigma_B^2)$.
Für den Value at Risk gilt dann mit Φ als Verteilungsfunktion der Standard-Normalverteilung

$$\begin{aligned} VaR^{0,05}\left(S^A\right) &= \mu_A + \sigma_A \cdot \Phi^{-1}\left(1 - 0{,}05\right) \\ &= \mu_A + \sigma_A \cdot u_{0,95} \\ &= 1 \text{ Mio} \,€ \end{aligned}$$

und

$$\begin{aligned}VaR^{0,05}\left(S^B\right) &= \mu_B + \sigma_B \cdot \Phi^{-1}(1-0{,}05)\\ &= \mu_B + \sigma_B \cdot u_{0,95}\\ &= 5 \text{ Mio €}\end{aligned}$$

Für die Summe S der zwei unabhängigen normalverteilten Zufallsvariablen S^A und S^B gilt

$$S = S^A + S^B \sim \mathcal{N}\left(\mu_A + \mu_B, \sigma_A^2 + \sigma_B^2\right)$$

Für das erforderliche Sicherheitskapital der fusionierten Bestände gilt dann:

$$\begin{aligned}VaR^{0,05}(S) &= \mu_A + \mu_B + u_{0,95}\sqrt{\sigma_A^2 + \sigma_B^2}\\ &= 1 - \sigma_A u_{0,95} + 5 - \sigma_B u_{0,95} + u_{0,95}\sqrt{\sigma_A^2 + \sigma_B^2}\\ &= 6 + u_{0,95}\left\{\sqrt{\sigma_A^2 + \sigma_B^2} - (\sigma_A + \sigma_B)\right\}\\ &= 6 + u_{0,95}\left\{\sqrt{1{,}5^2 + 2^2} - (1{,}5 + 2)\right\}\\ &= 6 + 1{,}645\,\{2{,}5 - 3{,}5\}\\ &= 6 - 1{,}645\\ &= 4{,}355 \text{ Mio €}\end{aligned}$$

Die Einsparung beim Sicherheitskapital beträgt 1,645 Mio €.

Aufgabe 2 (22 Punkte)

Die folgende Tabelle enthält die Jahres-Schäden einer Sachversicherung der letzten 9 Jahre in Mio €:

Jahr	2000	2001	2002	2003	2004	2005	2006	2007	2008
Schaden	11,5	6,5	2,0	15,5	11,0	6,5	12,0	12,5	5,5

Es soll mit Hilfe eines Q-Q-Plots überprüft werden, ob die Daten Realisierungen einer Weibull-verteilten Zufallsvariablen X sind, mit Verteilungsfunktion $F_X(x; a) = 1 - e^{-ax^2}$ für $x \geq 0$, wobei $a > 0$ ein unbekannter Parameter ist.

a) Zeigen Sie, dass die zu prüfende Verteilung einer reinen Skalenfamilie entstammt (Fall $\mu = 0$). Was ist hier der zugehörige Skalenparameter σ?
b) Zeichnen Sie den zugehörigen Q-Q-Plot, indem Sie auf der Ordinatenachse („x-Achse“) die Größen $z_k = Q_Z\left(\frac{k}{10}\right) = F_X^{-1}\left(\frac{k}{10}; 1\right)$ abtragen und auf der Abszisse („y-Achse“) die der Größe nach sortierten Jahresschäden $y_k = x_{(k)}$, $k = 1, \ldots, 9$.Benutzen Sie dafür das vorgegebene Diagramm. Geben Sie die z_k formelmäßig an.

c) Bestimmen Sie die zugehörige *Regressionsgerade durch den Nullpunkt* und tragen Sie diese ebenfalls in das Diagramm ein. (Warum ist dies sinnvoll?) Wie beurteilen Sie das Ergebnis im Hinblick auf die ursprüngliche Fragestellung bzw. den unbekannten Parameter a? [Alternativ dürfen Sie auch eine Ausgleichsgerade durch den Nullpunkt nach *Augenmaß* einzeichnen.]

Hinweis: Für gegebene Datenpaare (z_k, y_k), $k = 1, \ldots, n$ ist die Steigung s der Regressionsgeraden durch den Nullpunkt gegeben durch

$$s = \frac{\sum_{k=1}^{n} z_k y_k}{\sum_{k=1}^{n} z_k^2}.$$

Hilfstabelle (darf zur Rechnung verwendet werden, zwei Dezimalen Genauigkeit genügt):

k	1	2	3	4	5	6	7	8	9	Σ	s
$z_k y_k$	0,65	2,60	3,88	4,65	9,16	11,01	13,17	15,86	23,52		
z_k^2	0,11	0,22	0,36	0,51	0,69	0,92	1,20	1,61	2,30		

Lösung

Zu a)

Sei Z eine Zufallsvariable mit $F_Z = F_X(\cdot\,; 1)$ (Standard-Weibull-Verteilung). Dann besitzt $X = \mu + \sigma Z$ die Verteilungsfunktion

$$F_X(x) = P(\sigma Z \leq x - \mu) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = 1 - \exp\left(\left(\frac{x-\mu}{\sigma}\right)^2\right) = 1 - e^{-ax^2} \quad \text{für} \quad x \geq 0;$$

Koeffizientenvergleich: $\mu = 0$, also reine Skalenfamilie in $\sigma = \frac{1}{\sqrt{a}}$.

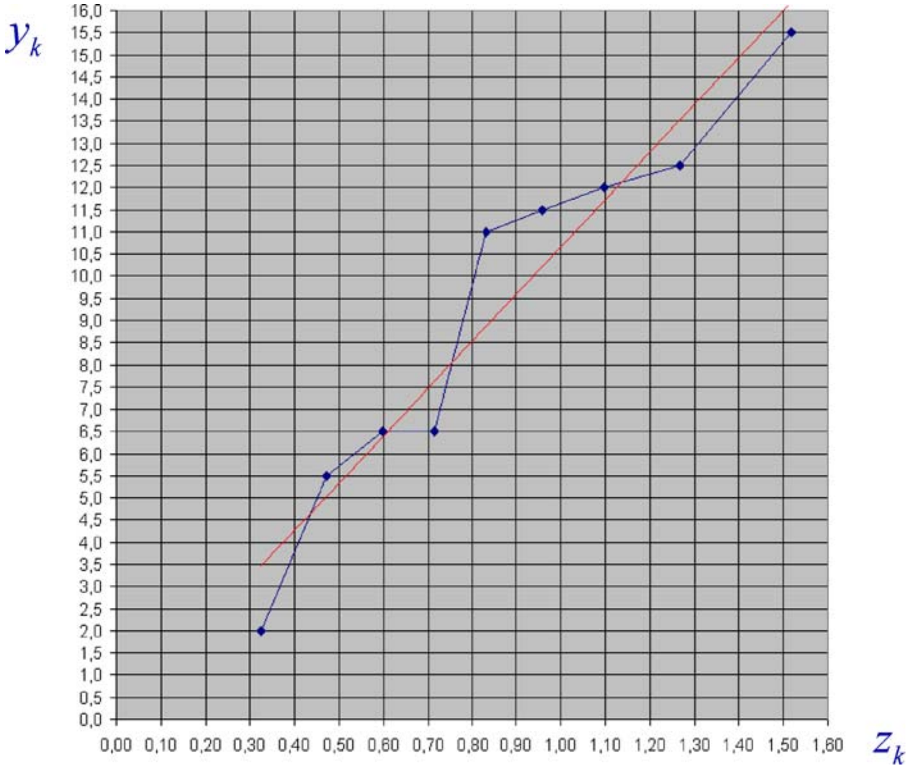
Zu b)

Es gilt $z_k = Q_Z\left(\frac{k}{10}\right) = F_X^{-1}\left(\frac{k}{10}; 1\right) = \sqrt{-\ln\left(1 - \frac{k}{10}\right)}$.

Zu c)

Tabelle:

k	1	2	3	4	5	6	7	8	9	Σ	s
$z_k y_k$	0,65	2,60	3,88	4,65	9,16	11,01	13,17	15,86	23,52	84,49	10,67
z_k^2	0,11	0,22	0,36	0,51	0,69	0,92	1,20	1,61	2,30	7,92	



Beurteilung: akzeptable Anpassung nach Augenmaß;

$$a = \frac{1}{\hat{\sigma}^2} = \frac{1}{s^2} = \frac{1}{10{,}67^2} = 0{,}0937^2 = 0{,}00878$$

Aufgabe 3 (32 Punkte)

Einem Versicherungsunternehmen liegen Schäden $x_1, \ldots, x_{20}$ mit

$$\frac{1}{20}\sum_{i=1}^{20} x_i = 3{,}02 \qquad \frac{1}{19}\sum_{i=1}^{20}(x_i - 3{,}02)^2 = 0{,}70$$

$$\frac{1}{20}\sum_{i=1}^{20} \ln\left(\frac{x_i}{2}\right) = 0{,}38 \qquad \frac{1}{19}\sum_{i=1}^{20}\left[\ln\left(\frac{x_i}{2} - 0{,}38\right)\right]^2 = 0{,}06$$

$$\frac{1}{20}\sum_{i=1}^{20} \mathrm{e}^{\frac{x_i}{2}} = 5{,}02 \qquad \frac{1}{19}\sum_{i=1}^{20}\left(\mathrm{e}^{\frac{x_i}{2}} - 5{,}02\right)^2 = 9{,}39$$

vor, die als Realisationen einer Zufallsvariablen mit Verteilungsfunktion

$$F(x;\vartheta) = \begin{cases} 0 \text{ falls } x \le 2 \\ 1 - \left(\frac{2}{x}\right)^{\vartheta} \text{ falls } x > 2 \end{cases} \tag{*}$$

mit Parameter $\vartheta > 0$ angesehen werden. Man kann davon ausgehen, dass die Schäden unabhängig und identisch verteilt sind.

a) Zeigen Sie, dass $\ell : (0, \infty) \to \mathbb{R}$, $\vartheta \mapsto n \ln \vartheta + \vartheta \sum_{i=1}^{n} \ln\left(\frac{2}{x_i}\right) - \sum_{i=1}^{n} \ln x_1$, $n = 20$ eine Log-Likelihood Funktion ist.
b) Bestimmen Sie einen Maximum-Likelihood Schätzer $\hat{\vartheta}_n$ und daraus den Schätzwert zu obigen Daten.
c) Bestimmen Sie die Informationsmatrix.
d) Bestimmen Sie die asymptotische Verteilung von $\hat{\vartheta}_n$, und geben Sie damit für obige Daten ein Konfidenzintervall für ϑ zu einem Niveau von 95% an.
e) Der Parameter ϑ aus (*) kann als Realisierung eines Strukturparameters Θ aufgefasst werden. Die Markterfahrung zeigt, dass für dessen Verteilung gilt: $P^{\Theta} = U[2{,}5;\ 3{,}5]$ (stetige Gleichverteilung über dem Intervall $[2{,}5;\ 3{,}5]$). Bestimmen Sie die Credibility-Prämie für obige Daten.

Sie dürfen für die zu (*) gehörige bedingte Verteilung $P^X(\,\cdot\,|\Theta)$ ohne Rechnung verwenden:

$$\mathrm{Var}(E(X|\Theta)) = 5{,}33, \quad E(\mathrm{Var}(X|\Theta)) = 6{,}75, \quad E(E(X|\Theta)) = 1{,}02$$

Quantile der Standardnormalverteilung:

p	0,80	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
$u_p = \Phi^{-1}(p)$	0,84	1,28	1,64	1,96	2,33	2,58	3,09	3,29

Lösung

Wir bezeichnen mit X die Zufallsvariable eines Schadens, ferner seien $\boldsymbol{x} = (x_1, \ldots, x_n)$ die beobachteten Schäden, $\boldsymbol{X} = (X_1, \ldots, X_n)$ wobei $X_i \stackrel{iid}{\sim} X$.

Zu a)

Die Dichte $f : \mathbb{R} \to [0, \infty)$ von X ist gegeben durch

$$f(x) = \begin{cases} 0 & \text{falls } x \le 2 \\ 2^{\vartheta} \cdot \vartheta x^{-\vartheta-1} & \text{falls } x > 2. \end{cases} \tag{*}$$

Aufgrund der Unabhängigkeit gilt für die Likelihood

$$L(\vartheta; x_1, \ldots, x_n) = 2^{n\vartheta} \vartheta^n \prod_{i=1}^{n} x_i^{-\vartheta-1}$$

und somit für $\ell := \ln L$

$$\begin{aligned} \ell(\vartheta) &= n\vartheta \ln 2 + n \ln \vartheta - (\vartheta + 1) \sum_{i=1}^{n} \ln x_i \\ &= n \ln \vartheta + \vartheta \sum_{i=1}^{n} \ln\left(\frac{2}{x_i}\right) - \sum_{i=1}^{n} \ln x_i \end{aligned}$$

Zu b)

$$\begin{aligned} \ell'(\vartheta) &= \tfrac{n}{\vartheta} + \sum_{i=1}^{n} \ln\left(\tfrac{2}{x_i}\right) \\ \ell''(\vartheta) &= -\tfrac{n}{\vartheta^2} < 0 \end{aligned}$$

Somit liegt in $\frac{n}{\sum_{i=1}^{n} \ln\left(\frac{2}{x_i}\right)}$ ein Maximum der Likelihoodfunktion vor. Damit ist die Zufallsvariable

$$\hat{\vartheta} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{2}{X_i}\right)}$$

ein Maximum Likelihood Schätzer. Im vorliegenden Fall ergibt sich der Schätzwert

$$\hat{\vartheta} = \frac{1}{0{,}38} = 2{,}63$$

Zu c)

Laut Definition gilt

$$\begin{aligned} i_n(\vartheta) &= -E(\ell''(-\vartheta, \boldsymbol{X})) = E\left(\tfrac{n}{\vartheta^2}\right) = \tfrac{n}{\vartheta^2} \text{bzw.} \\ i_n(\vartheta) &= n \cdot i(\vartheta) \quad \text{mit} \quad i(\vartheta) = \tfrac{1}{\vartheta^2}. \end{aligned}$$

Zu d)

Bezeichne mit $\hat{\vartheta}_n$ den ML-Schätzer aus (b). Dann folgt

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{i(\vartheta)}\right) = \mathcal{N}\left(0, \vartheta^2\right)$$

Damit ergibt sich für große $n \in \mathbb{N}$

$$\hat{\vartheta}_n \sim \mathcal{N}\left(\vartheta, \frac{\vartheta^2}{n}\right).$$

Als Konfidenzintervall für ϑ erhalten wir

$$\left(\hat{\vartheta}_n - \frac{u_{1-\alpha/2}}{\sqrt{n}\sqrt{i(\hat{\vartheta}_n)}}, \hat{\vartheta}_n + \frac{u_{1-\alpha/2}}{\sqrt{n}\sqrt{i(\hat{\vartheta}_n)}}\right).$$

wobei $\alpha = 0{,}05$, $\hat{\vartheta}_n = 2{,}63$, $i(\hat{\vartheta}_n) = \frac{1}{\vartheta_n^2} = \frac{1}{2{,}63^2}$, $n = 20$ und somit ist

$$\left(2{,}63 - \frac{1{,}96 \cdot 2{,}63}{\sqrt{20}}, 2{,}63 + \frac{1{,}96 \cdot 2{,}63}{\sqrt{20}}\right) = (1{,}48;\, 3{,}78)$$

das gesuchte Konfidenzintervall.

Zu e)

Für die Credibility-Prämie H^{**} gilt

$$H^{**} = z\overline{X} + (1-z)H$$

$$\text{mit} \quad z = \frac{\mathrm{Var}(E(X|\Theta))}{\frac{1}{n}E(\mathrm{Var}(X|\Theta)) + \mathrm{Var}(E(X|\Theta))}.$$

Hier folgt nun aus den Angaben

$$z = \frac{5,33}{\frac{1}{20}\cdot 6,75 + 5,33} = 0,94$$
$$H^{**} = 0,94 \cdot 3,02 + 0,06 \cdot 1,02 = 2,90$$

Die Credibility-Prämie beträgt also 2,90.

Aufgabe 4 (34 Punkte)

Betrachtet werde ein Kollektiv von n x-jährigen zu Beginn eines Betrachtungsjahres lebenden Männern. Sei K die Zufallsgröße der Anzahl Sterbefälle dieses Kollektivs (Annahmen: Alle Angehörigen des Kollektivs besitzen die gleiche Sterbewahrscheinlichkeit q_x, die einzelnen Sterbefälle sind von einander unabhängig).

a) Welche Verteilung aus der Exponentialfamilie eignet sich zur Beschreibung der Wahrscheinlichkeitsverteilung von K? Man begründe dies!
Man stelle $P(K = k)$ in Abhängigkeit von n und q_x dar!
b) Sei Y die relative Häufigkeit der Sterbefälle. Man stelle Y in Abhängigkeit der o. a. Größen dar und geben den Erwartungswert von Y mit den o. a. Größen an.
c) Ziel ist, die Wahrscheinlichkeitsverteilung $P(Y = y)$ in der im Kontext der Verallgemeinerten Linearen Modelle üblichen Darstellung zu schreiben:
$$P(Y = y) = \exp\left[\frac{w}{\Psi}(\vartheta \cdot y - b(\vartheta)) + c\left(y; \frac{\Psi}{w}\right)\right] \tag{1}$$
Man gebe die kanonische Link-Funktion der o. a. Verteilung an.
d) Welcher der o.a. Parameter der in a) angegebenen Verteilung kann hier zur Darstellung von ϑ eingesetzt werden? Warum?
e) Man ersetze in der Formel aus Aufgabenteil a) diesen Parameter durch ϑ!
f) Welche Zusammenhänge zwischen Erwartungswert von Y und ϑ führen auf die Funktion $b(\vartheta)$ in (1)?
g) Man schreibe die in e) gefundene Darstellung in der Form $\exp(\ln(P(Y = y)))$ und führe explizit die Umformungen durch, die Einsetzen der in f) gefundenen Funktion $b(\vartheta)$ ermöglichen und leite auf die Darstellung (1) über. Man bezeichne die Funktion c!
h) Gegeben seien 10 Kollektive von zu Jahresanfang lebenden Männern: Diese gruppieren sich in die Altersklassen $x \in \{40; 41; 42; 43; 44\}$ und in die Tarifwerke $i \in \{A; B\}$. Bekannt seien die Anzahlen n_{xi} zu Jahresbeginn Lebender und die Anzahlen k_{xi} im Beobachtungsjahr Gestorbener.

n_{xi}	Tarifwerk i	k_{xi}	Tarifwerk i		
Alter x	A	B	Alter x	A	B
40	10.000	5.000	40	15	5
41	10.000	5.000	41	17	7
42	10.000	5.000	42	16	4
43	10.000	5.000	43	18	6
44	10.000	5.000	44	17	7

Man zeichne die Mittelwert-Plots der relativen Häufigkeiten!

i) Es handle sich um Bestände aus Versicherungen mit Todesfallcharakter, die Auswertung basiere auf Zahlen des Jahres 2008. Die Bestände aus Tarifwerk A wurden zwischen 1995 und 2001 aufgebaut, die Bestände aus Tarifwerk B ab 2002 (Tarifwerk B ist derzeit noch offen). Es ist bekannt, dass eine strenge Risikoselektion angewandt wurde. Es handelt sich um sehr langfristige Verträge (durchschnittliche Restlaufzeit 20 Jahre).
   - Welche qualitative Aussage zu den Sterblichkeiten lässt sich aus den Mittelwert-Plots ableiten?
   - Welcher Unterschied im Einfluss der Selektion liegt zur Erklärung der Beobachtung nahe? Warum?
   - Welchen Warnhinweis wird der Aktuar aufgrund dieser Beobachtung und Interpretation hinsichtlich der Angemessenheit der biometrischen Rechnungsgrundlagen äußern?
   - Was kann augrund der vorliegenden Information nicht abschließend beurteilt werden?
   - Welche Auswertung liegt nahe, diese Befürchtung zu erhärten oder zu widerlegen?

Lösung

Zu a)

Es eignet sich die Binomialverteilung.
Begründung: Für $n = 1$ sind die Elementarereignisse $k = 0$ mit $P(K = 0) = 1 - q_x$ und $k = 1$ mit $P(K = 1) = q_x$. Diese Verteilung heißt Bernoulli-Verteilung. Die Faltung von n unabhängigen identischen Bernoulli-Verteilungen ist eine Binomialverteilung $B(n; q_x)$ Dies kann mit Hilfe der erzeugenden Funktion gezeigt werden.

$$P(K = k) = \binom{n}{k} q_x^k \cdot (1 - q_x)^{n-k}$$

Zu b)

$$Y = \frac{K}{n} \quad E(Y) = q_x$$

Zu c)

Kanonische Link Funktion:

$$\text{logit:} \quad \vartheta = \ln\left(\frac{\mu}{1-\mu}\right)$$

Zu d)

Eingesetzt werden kann $q_x = \mu$, weil dies der Erwartungswert der rel. Häufigkeit ist, und der Parameter ϑ über die Linkfunktion mit μ zusammenhängt.

Zu e)

Invertieren: $q_x = \mu(\vartheta) = \frac{e^{\vartheta}}{1+e^{\vartheta}}$ und einsetzen:

$$P(Y=y) = \binom{n}{ny} \cdot \left(\frac{e^{\vartheta}}{1+e^{\vartheta}}\right)^{ny} \cdot \left(\frac{1}{1+e^{\vartheta}}\right)^{n-ny}$$

Zu f)

$$E(Y) = q_x = \mu(\vartheta) = b'(\vartheta) = \frac{e^{\vartheta}}{1+e^{\vartheta}} \rightarrow \quad b(\vartheta) = \int \frac{e^{\vartheta}}{1+e^{\vartheta}} d\vartheta = \ln(1+e^{\vartheta})$$

Zu g)

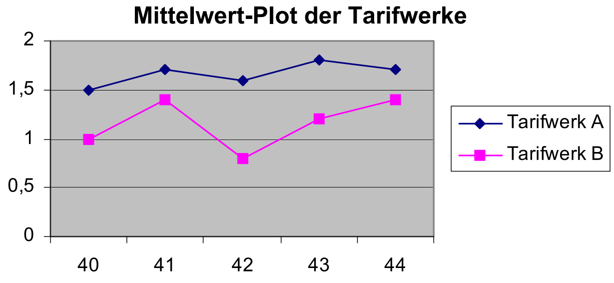
$$\begin{aligned}
&\exp\left(\ln\left(P(Y=y)\right)\right)\\
&= \exp\left[\ln\binom{n}{ny} + ny\ln\left(e^{\vartheta}\right) - ny\ln\left(1+e^{\vartheta}\right) - (n-ny)\ln\left(1+e^{\vartheta}\right)\right]\\
&= \exp\left[ny\vartheta - n\ln\left(1+e^{\vartheta}\right) - ny\ln\left(1+e^{\vartheta}\right) + ny\ln\left(1+e^{\vartheta}\right) + \ln\binom{n}{ny}\right]\\
&= \exp\left[n\cdot\left(y\vartheta - \ln\left(1+e^{\vartheta}\right)\right) - \left(-\ln\binom{n}{ny}\right)\right]
\end{aligned}$$

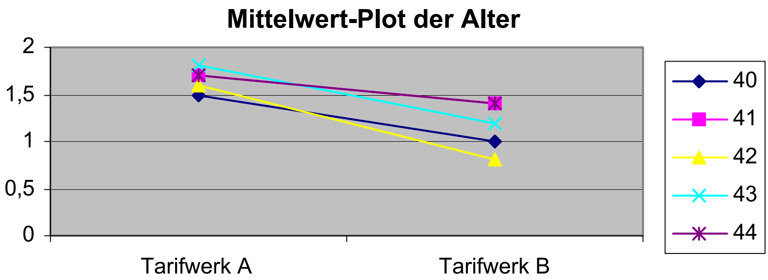
Mit $\Psi = 1$, $w = n$, $b(\vartheta)$ wie in f) und $c\left(y; \frac{w}{\Psi}\right) = \ln\binom{n}{ny}$ folgt (1).

Zu h)

Rel. Häufigkeiten (‰)

	Tarifwerk Tarifwerk A	Tarifwerk B
40	1,5	1
41	1,7	1,4
42	1,6	0,8
43	1,8	1,2
44	1,7	1,4





Zu i)

- Die Sterblichkeiten des jungen Tarifwerkes liegen im allen beobachtbaren Altersgruppen deutlich unter denjenigen des alten Tarifwerkes.
- Es handelt sich um Versicherungen mit Todesfall-Charakter. Die Risikoselektion bewirkt demzufolge eine Senkung der Sterbewahrscheinlichkeiten. Die Selektion greift im Wesentlichen bei Vertragsabschluß. Damit liegt im jungen Tarifwerk der Zeitpunkt der Selektion kürzer zurück als beim alten Tarifwerk. Die Vermutung liegt nahe, dass der Unterschied in der beobachteten Sterblichkeit darauf zurückzuführen ist, dass der Einfluss der Selektion zu Vertragsbeginn im Laufe der Zeit abnimmt.
  - Mit Alterung des Tarifwerkes wird die aktuell beobachtbare Untersterblichkeit geringer. Dies gilt insbes. für geschlossene Tarifwerke.
  - Es kann nicht beurteilt werden, ob dieser Verschlechterungsprozess aktuell abgeschlossen ist, oder ob die Verschlechterung weiter fortschreiten wird.
  - Es liegt nahe, eine 2-dimensionale Auswertung nach Alter und Bestandszugehörigkeit durchzuführen und die Faktoren der Bestandszugehörigkeit auf einen Trend hin zu analysieren.