Risikotheorie

2. Teil

Prof. Erika Hausenblas

Montanuniversität Leoben, Österreich

25. April 2018

Inhalt

Bonus-Malus System der Versicherung

Zwei Kategorien von Risiken:

- den *objektiven Risiken*: z.B. die PS zahl eines Autos, der Hubraum, das Gewicht, etc .. ;
- und den *subjektiven Risiken* (nicht objektiv messbare Risiken): Risikobereitschaft, das Können, das Fahrverhalten des Fahrers, Temperament, die genaue Kilometeranzahl, etc ... .

Bonus-Malus System der Versicherung

Bonus-Malus System der Versicherung

- Jeder Autofahrer a hat sein individuelles Risikoprofil, das durch den Parameter ϑ_a beschrieben wird. Dieser Parameter kann Werte aus D_Θ annehmen, wobei D_Θ die Menge aller möglichen Werte für ϑ_a darstellt.

Bonus-Malus System der Versicherung

- Jeder Autofahrer a hat sein individuelles Risikoprofil, das durch den Parameter ϑ_a beschrieben wird. Dieser Parameter kann Werte aus D_Θ annehmen, wobei D_Θ die Menge aller möglichen Werte für ϑ_a darstellt.
- Für einen bestimmten Autofahrer ist der genaue Wert von ϑ_a zumeist unbekannt. Aus Statistiken kann man aber Rückschlüsse auf die Verteilung von ϑ_a machen, so sind die meisten Autofahrer vorsichtig, allerdings gibt es einige Ausreißer, die immer wieder Unfälle verursachen.

Bonus-Malus System der Versicherung

- Jeder Autofahrer a hat sein individuelles Risikoprofil, das durch den Parameter ϑ_a beschrieben wird. Dieser Parameter kann Werte aus D_Θ annehmen, wobei D_Θ die Menge aller möglichen Werte für ϑ_a darstellt.
- Für einen bestimmten Autofahrer ist der genaue Wert von ϑ_a zumeist unbekannt. Aus Statistiken kann man aber Rückschlüsse auf die Verteilung von ϑ_a machen, so sind die meisten Autofahrer vorsichtig, allerdings gibt es einige Ausreißer, die immer wieder Unfälle verursachen.
- A'priori Verteilung des Parameters ϑ_a

Bonus-Malus System der Versicherung

- Jeder Autofahrer a hat sein individuelles Risikoprofil, das durch den Parameter ϑ_a beschrieben wird. Dieser Parameter kann Werte aus D_Θ annehmen, wobei D_Θ die Menge aller möglichen Werte für ϑ_a darstellt.
- Für einen bestimmten Autofahrer ist der genaue Wert von ϑ_a zumeist unbekannt. Aus Statistiken kann man aber Rückschlüsse auf die Verteilung von ϑ_a machen, so sind die meisten Autofahrer vorsichtig, allerdings gibt es einige Ausreißer, die immer wieder Unfälle verursachen.
- A'priori Verteilung des Parameters ϑ_a
- A'priori ist die Risikobereitschaft im Fahrverhalten bei Abschluss eines Vertrages hoch. Fahren Sie jetzt einige Zeit Unfallfrei, werden die Daten miteinbezogen und die Autohaftpflichtversicherung ordnet Ihre Risikobereitschaft geringer ein. Die Wahrscheinlichkeit im nächsten Jahr einen Unfall zu verursachen ist (statistisch gesehen) geringer und Ihre Prämie sinkt.

Statistischer Hintergrund

Variablen

- Jeder Autofahrer hat eine Risikobereitschaft Θ, diese Risikobereitschaft ist wird als Zufallsvariable interpretiert;
- Für $j = 1, \ldots, n$ ist X_j die Forderung eines Versicherten im Jahr j mit einer Verteilung F_ϑ die vom Parameter $\vartheta = \Theta$ abhängt.

Statistischer Hintergrund

Variablen

- Jeder Autofahrer hat eine Risikobereitschaft Θ, diese Risikobereitschaft ist wird als Zufallsvariable interpretiert;
- Für $j = 1, \ldots, n$ ist X_j die Forderung eines Versicherten im Jahr j mit einer Verteilung F_ϑ die vom Parameter $\vartheta = \Theta$ abhängt.

Gegeben im Jahre n:

Forderung von den letzten Jahren, d.h. $\mathbf{X} = (X_1, \ldots, X_n)'$

Statistischer Hintergrund

Variablen

- Jeder Autofahrer hat eine Risikobereitschaft Θ, diese Risikobereitschaft ist wird als Zufallsvariable interpretiert;
- Für $j = 1, \ldots, n$ ist X_j die Forderung eines Versicherten im Jahr j mit einer Verteilung F_ϑ die vom Parameter $\vartheta = \Theta$ abhängt.

Gegeben im Jahre n:

Forderung von den letzten Jahren, d.h. $\mathbf{X} = (X_1, \ldots, X_n)'$

Aufgabe:

Die individuelle Prämien für das nächste Jahr aller Autofahrer berechnen.

Statistischer Hintergrund

Variablen

- Jeder Autofahrer hat eine Risikobereitschaft Θ, diese Risikobereitschaft ist wird als Zufallsvariable interpretiert;
- Für $j = 1, \ldots, n$ ist X_j die Forderung eines Versicherten im Jahr j mit einer Verteilung F_ϑ die vom Parameter $\vartheta = \Theta$ abhängt.

Gegeben im Jahre n:

Forderung von den letzten Jahren, d.h. $\mathbf{X} = (X_1, \ldots, X_n)'$

Aufgabe:

Die individuelle Prämien für das nächste Jahr aller Autofahrer berechnen.

Dazu ist die Verteilung von X_{n+1} zu schätzen. Da unter der Bedingung dass ϑ bekannt ist, der Typ der Verteilung F_ϑ bekannt ist, heißt dies dass der Parameter ϑ geschätzt werden soll. Um ϑ zu schätzen, können wir nur auf die Daten $\mathbf{X}$ zurückgreifen.

Statistischer Hintergrund

Variablen

- Jeder Autofahrer hat eine Risikobereitschaft Θ, diese Risikobereitschaft ist wird als Zufallsvariable interpretiert;
- Für $j = 1, \ldots, n$ ist X_j die Forderung eines Versicherten im Jahr j mit einer Verteilung F_ϑ die vom Parameter $\vartheta = \Theta$ abhängt.

Gegeben im Jahre n:

Forderung von den letzten Jahren, d.h. $\mathbf{X} = (X_1, \ldots, X_n)'$

Aufgabe:

Die individuelle Prämien für das nächste Jahr aller Autofahrer berechnen.

Dazu ist die Verteilung von X_{n+1} zu schätzen. Da unter der Bedingung dass ϑ bekannt ist, der Typ der Verteilung F_ϑ bekannt ist, heißt dies dass der Parameter ϑ geschätzt werden soll. Um ϑ zu schätzen, können wir nur auf die Daten $\mathbf{X}$ zurückgreifen.

Modellannahmen:

Um das Modell mathematisch formulieren zu können werden folgende Annahmen gemacht:

- Ist $\Theta = \vartheta$ gegeben, dann sind $X_1, X_2, \ldots$ unabhängig, identisch verteilt und $X_1 \sim F_\vartheta$.
  Identische verteilt ist gleich zu setzten mit stationärität - damit kann man auf historische Daten zurückgreifen - Inflation und andere Einflussfaktoren müssen dabei herausgerechnet werden.

Modellannahmen:

Um das Modell mathematisch formulieren zu können werden folgende Annahmen gemacht:

- Ist $\Theta = \vartheta$ gegeben, dann sind $X_1, X_2, \ldots$ unabhängig, identisch verteilt und $X_1 \sim F_\vartheta$.
  Identische verteilt ist gleich zu setzten mit stationärität - damit kann man auf historische Daten zurückgreifen - Inflation und andere Einflussfaktoren müssen dabei herausgerechnet werden.
- Θ ist eine Zufallsvariable mit Verteilungsfunktion U und Dichtefunktion u und Verteilungsfunktion U. Diese Verteilungsfunktion heißt *Struktur Funktion des Kollektives* (structural function of the collective).

Problem:

Gegeben $\mathbf{X} = (X_1, \ldots, X_n)'$, zu schätzen ist die Prämie für das nächste Jahr. Die Prämie hängt vom Parameter ϑ ab, aber dieser Parameter hängt von $(X_1, \ldots, X_n)'$. Damit ist T also ist eine Funktion

$$T : \mathbb{R}^n \longrightarrow D,$$

wobei D die Menge der möglichen Schätzewerte bezeichnet. Mittels $\hat{\vartheta} = T(X_1, \ldots, X_n)$ wird jetzt die Prämie errechnet.

Methoden die Prämie zu Schätzen

Allgemein gibt es bestimmte Möglichkeiten Prämien (oder Rücklagen) zu berechnen. Dabei hat man eine Zufallsvariable X, die den möglichen Verlust (Risiko, Forderungen) bezeichnet. In unseren Fall, gilt $X = X_{n+1}$ ist die Forderung im Jahre $n+1$:

Methoden die Prämie zu Schätzen

Allgemein gibt es bestimmte Möglichkeiten Prämien (oder Rücklagen) zu berechnen. Dabei hat man eine Zufallsvariable X, die den möglichen Verlust (Risiko, Forderungen) bezeichnet. In unseren Fall, gilt $X = X_{n+1}$ ist die Forderung im Jahre $n+1$:

Möglichkeiten eine Prämien festzusetzten

- $X \mapsto (1+\alpha)\mathbb{E}X, \quad \alpha > 0$ (expectation principle)
- $X \mapsto \mathbb{E}X + \beta \mathbf{Std}(X), \quad \beta > 0$ (Standard Deviation Principle)
- $X \mapsto \mathbb{E}X + \gamma \mathbf{Var}(X), \quad \gamma > 0$ (Variance Principle)
- $X \mapsto \frac{1}{\delta} \ln(\mathbb{E}e^{\delta X}), \quad \delta > 0$ (Exponential Principle)

Methoden die Prämie zu Schaetzen

Aufgrund der Daten $\mathbf{X} = (X_1, \ldots, X_n)'$, ist der Wert $\mathbb{E}[X_{n+1} \mid \vartheta]$, bzw. $\mathbb{E}\left[\mu(\Theta) \mid \Theta\right] = \mu(\vartheta)$ zu Schätzen.

Definition

Die Korrekte individuelle Prämie ist gegeben durch

$$\mathcal{P}^{ind} = \mathbb{E}[X_{n+1} \mid \vartheta] = \int_0^\infty x\, dF_\vartheta(x) =: \mu(\vartheta).$$

Definition

Die Kollektive Prämie ist gegeben durch

$$\mathcal{P}^{coll} = \int_{D_\Theta} \mu(\vartheta)\, dU(\vartheta) =: \mu_0.$$

Hier bezeichnet D_Θ die Menge aller zulässigen (und möglichen) Parameter Θ, i.e. $D_\Theta := \{\vartheta \mid \vartheta \text{ ist als Parameter möglich }\}$.

Schätzmethode

Minimierung der Abweichung:

Sei $L(\vartheta, T(\mathbf{x}))$ der *Verlust*, die sich ergibt, falls ϑ der richtige Parameter ist und $T(\mathbf{x})$ der aus der Beobachtung $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ geschätzte Wert von $\mu(\vartheta)$ ist. Eine mögliche Verlustfunktion wäre z.B. die quadratische Funktion

$$L(\vartheta, T(\mathbf{x})) = (\mu(\vartheta) - T(\mathbf{x}))^2 .$$

Schätzmethode

Minimierung der Abweichung:

Sei $L(\vartheta, T(\mathbf{x}))$ der *Verlust*, die sich ergibt, falls ϑ der richtige Parameter ist und $T(\mathbf{x})$ der aus der Beobachtung $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ geschätzte Wert von $\mu(\vartheta)$ ist. Eine mögliche Verlustfunktion wäre z.B. die quadratische Funktion

$$L(\vartheta, T(\mathbf{x})) = (\mu(\vartheta) - T(\mathbf{x}))^2 .$$

Risikofunktion

$$R_T(\vartheta, \mathbf{x}) := \mathbb{E}_\vartheta \left[L(\vartheta, T(\mathbf{x})) \right] = L(\vartheta, T(\mathbf{x}))\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n).$$

Schätzmethode

Risikofunktion

$$R_T(\vartheta, \mathbf{x}) := \mathbb{E}_\vartheta \left[L(\vartheta, T(\mathbf{x})) \right] = L(\vartheta, T(\mathbf{x}))\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n).$$

Schätzmethode

Risikofunktion

$$R_T(\vartheta, \mathbf{x}) := \mathbb{E}_\vartheta \left[L(\vartheta, T(\mathbf{x})) \right] = L(\vartheta, T(\mathbf{x}))\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n).$$

Definition

Bayes'scher Riskoschätzer von T mit Anfangsverteilung $U(\vartheta)$:

$$R_T(\mathbf{x}) := \int_D R_T(\vartheta, \mathbf{x})\, dU(\vartheta).$$

Schätzmethode

Risikofunktion

$$R_T(\vartheta, \mathbf{x}) := \mathbb{E}_\vartheta \left[L(\vartheta, T(\mathbf{x})) \right] = L(\vartheta, T(\mathbf{x}))\, dF_\vartheta(x_1) \cdots dF_\vartheta(x_n).$$

Definition

Bayes'scher Riskoschätzer von T mit Anfangsverteilung $U(\vartheta)$:

$$R_T(\mathbf{x}) := \int_D R_T(\vartheta, \mathbf{x})\, dU(\vartheta).$$

Definition

Der *Bayes'scher Schätzer* von T ist gegeben durch

$$T(\mathbf{x}) := \arg\min_{t:\mathbb{R}^n \to D} R_t(\mathbf{x}),$$

wobei das Minimum über alle möglichen Abbildungen $T : \mathbb{R}^n \to \mathbb{R}$ genommen wird.

Poisson Verteilung

Verteilungsfunktion:

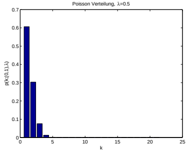
$$\begin{aligned} & \mathbb{P}\left(K = k \text{ im Zeitintervall } (0,t)\ \right) \\ = \quad & p(k;(0,t),\lambda) = \frac{(\lambda t)^k}{k!} e^{-k\lambda t}, \quad k = 1,2,\ldots . \end{aligned}$$

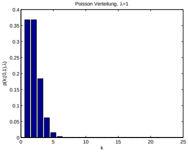
Der Erwartungswert lautet

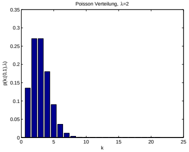
$$\mathbb{E}X = \lambda t.$$

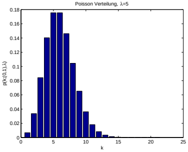
Die Varianz lautet

$$\mathrm{Var}[X] = \lambda t.$$









Gamma Verteilung

Die Dichtefunktion einer Gamma-verteilten Zustandsvariable $X \sim \Gamma(k;\lambda)$ mit reellwertigen Parametern $k > 0$ und $\lambda > 0$ hat für $x > 0$ die Gestalt

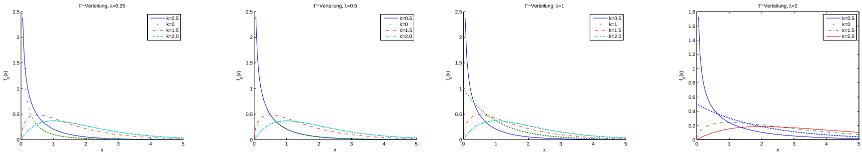
$$f(x) = \frac{\lambda^k}{\Gamma(k)} \cdot x^{k-1} \cdot e^{-\lambda x}.$$

Dabei ist

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

die sogenannte Gamma Funktion.
Erwartungswert: $\mathbb{E}X = \frac{k}{\lambda}$, die Varianz: $\mathrm{Var}[X] = \frac{k}{\lambda^2}$.

Die Summe zweier unabhängig Γ-verteilter Zufallsvariablen $X_1 \sim \Gamma(k_1,\lambda)$, $X_2 \sim \Gamma(k_2,\lambda)$ $(k_1, k_2 > 0)$ mit demselben Verteilungsparameter $\lambda > 0$ ist selbst auch Γ verteilt mit Parameter λ und es gilt $X_1 + X_2 \sim \Gamma(k_1 + k_2, \lambda)$. Insbesondere lässt sich für k ganzzahlig eine Γ-verteilte Zufallsvariable als k–fache Summe von k unabhängigen exponentialverteilter Zufallsvariablen interpretieren.



Weitere Modellannahmen

1. Gegeben sei das Risikoprofil ϑ des Autofahrers. Dann gilt

$$\mathbb{E}\left[X_j \mid \Theta = \vartheta\right] = C \cdot \mathbb{E}\left[N_j \mid \Theta = \vartheta\right]$$

   wobei C nur von der PS-Zahl des Autos abhängt und $\mathbb{E}\left[N_j \mid \Theta = \vartheta\right]$ vom Fahrer (bzw. dessen Risikobereitschaft) anhängt.

2. Gegeben ist $\Theta = \vartheta$, dann sind die $\{N_j : j = 1, \ldots, n, n+1\}$ unabhängig und Poisson verteilt mit Parameter ϑ, i.e.

$$f_\vartheta(N_j) = \mathbb{P}\left(N_j = k \mid \Theta = \vartheta\right) = e^{-\vartheta} \frac{\vartheta^k}{k!}.$$

3. Die Zufallsvariable Θ ist Gamma verteilt mit Parametern γ und β. Genauer, die Dichtefunktion lautet

$$u(\vartheta) = \begin{cases} \frac{\beta^\gamma}{\Gamma(\gamma)} \vartheta^{\gamma-1} e^{-\beta\vartheta}, & \vartheta > 0, \\ 0 & \text{sonst .} \end{cases}$$

Weitere Modellannahmen

Bemerkung

Es gilt

$$\mathbb{E}[\Theta] = \frac{\gamma}{\beta} \quad \text{und} \quad \mathbf{Var}[\Theta] = \frac{\gamma}{\beta^2}.$$

Durch eine kurze Rechnung lässt sich folgendes zeigen

$$\begin{aligned}\mathbb{E}[N_{n+1} \mid \Theta = \vartheta] &= \sum_{k=0}^{\infty} \mathbb{P}(N_{n+1} = k \mid \Theta = \vartheta) \cdot k = \sum_{k=0}^{\infty} e^{-\vartheta} \frac{\vartheta^k}{k!} \cdot k \\ &= \vartheta e^{-\vartheta} \sum_{k=1}^{\infty} \frac{\vartheta^{k-1}}{(k-1)!} = \vartheta e^{-\vartheta} e^{\vartheta} = \vartheta.\end{aligned}$$

Wichtige Eigenschaften

Proposition

Unter den oben genannten Annahmen gilt

$$\begin{aligned}
\mathcal{P}^{ind} &= \mathbb{E}\left[N_{n+1} \mid \Theta\right] = \Theta, \\
\mathcal{P}^{coll} &= \mathbb{E}[\Theta] = \frac{\gamma}{\beta}, \\
\mathcal{P}^{bayes} &= \frac{\gamma + \sum_{j=1}^{n} N_j}{\beta + n} = \alpha \bar{\mathbf{N}} + (1-\alpha)\frac{\gamma}{\beta},
\end{aligned}$$

wobei $\alpha = n/(n+\beta)$ und $\bar{\mathbf{N}} = \frac{1}{n}\sum_{j=1}^{n} N_j$. Weiters gilt für die Abweichung der geschätzten Prämie zur eigentlichen Prämie

$$\mathbb{E}\left[\left(\mathcal{P}^{Bayes} - \Theta\right)^2\right] = (1-\alpha)\,\mathbb{E}\left[\left(\mathcal{P}^{coll} - \Theta\right)^2\right] = \alpha\mathbb{E}\left[(\bar{N} - \Theta)^2\right].$$

Einsetzten der Daten

k	0	1	2	3	4	5	6	Total
$N(k) = \#$ Polizen mit k Forderungen	103 704	14 075	1766	255	45	6	2	119 853

Modellvergleich

k	beobachtete	Poisson ($\lambda = 0.1555$)	Negative Binomial ($\gamma = 1.001, \beta = 6.458$)
0	103 704	102 629	103 757
1	14 075	15 922	13 934
2	1 766	1 234	1 871
3	255	64	251
4	45	3	34
4	6	0	5
6	2	0	1