Risikotheorie

3. Teil

Prof. Erika Hausenblas

Montanuniversität Leoben, Österreich

25. April 2018

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 Prämie für das nächste Jahr jedes Autofahrers 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 Prämie für das nächste Jahr jedes Autofahrers 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 Prämie für das nächste Jahr jedes Autofahrers 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.

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

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

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

Das Bonus Malus System

Vorgangsweise

- Schätzen der Strukturalen Parameter α und β anhand der Gesamtdaten;
- Berechnen des Individuellen Prämie mit Hilfe der beobachteten Daten eines Autofahrers;

Das Bonus Malus System

Vorgangsweise

- Schätzen der Strukturalen Parameter α und β anhand der Gesamtdaten;
- Berechnen des Individuellen Prämie mit Hilfe der beobachteten Daten eines Autofahrers;

Literatur

- Bühlmann, Gisler: A Course in Credibility Theory and its Applications
- Hoff: A First Course in Bayesian Statistical Methods;

Kunjugierte Klassen von Verteilungen

Das Problem:

- Sei D der Parameterbereich und $p : D \to \mathbb{R}_0^+$ eine *struktur Funktion* abhängig von Parametern α und β;
- Sei $\{F_\theta : \theta \in D\}$ eine gegebene Familie von Verteilungen, $\{f(x \mid \theta) : \theta \in D\}$ die Familie der zugehörigen Verteilungsfunktionen;
- Sei X verteilt mit Dichtefunktion

$$f(x) = \int_D f(x \mid \theta) p(\theta) d\theta, \quad x \in \mathbb{R}.$$

- Gegeben ist $\mathbf{X} = (X_1, \ldots, X_n)$, schätzen Sie α und β;
- Gegeben ist $\mathbf{X}_{ind} = (X_1, \ldots, X_n)$, schätzen Sie Θ;
- Berechnen Sie die Prämie anhand von Θ.

Kunjugierte Klassen von Verteilungen

- Binominal Verteilung $\Leftrightarrow$ Beta Verteilung
- Normal Verteilung $\Leftrightarrow$ Normal Verteilung
- Gamma Verteilung $\Leftrightarrow$ Gamma Verteilung
- Geometrische Verteilung $\Leftrightarrow$ Beta Verteilung
- Pareto Verteilung $\Leftrightarrow$ Gamma Verteilung

Kunjugierte Klassen von Verteilungen

- Binominal Verteilung $\Leftrightarrow$ Beta Verteilung
- Normal Verteilung $\Leftrightarrow$ Normal Verteilung
- Gamma Verteilung $\Leftrightarrow$ Gamma Verteilung
- Geometrische Verteilung $\Leftrightarrow$ Beta Verteilung
- Pareto Verteilung $\Leftrightarrow$ Gamma Verteilung

Maschinelles Lernen

Der Binomial – Beta Fall

Die Story:

Ein Betrieb produziert Werkstücke . Diesen Werkstücke werden von verschiedenen Maschinen produziert jede Maschine hat eine bestimmte Ausfall Rate. Diese Ausfallrate kann bei verschiedenen Maschinen unterschiedlich sein. Allerdings hat jede Maschine eine konstante Ausfall Rate die nicht in der Zeit variiert. Die Aufgabe ist es diese Ausfallrate in Abhängigkeit von der Maschine zu schätzen.
Mathematisch gesehen, hat man verschiedene Gruppen von Werkstücken. Man ist an der Wahrscheinlichkeit des Auftretens eines Schadenfalles innerhalb einer bestimmten Gruppe interessiert. Hier nehmen wir an dass jedes Werkstück (Element) der Gruppe (oder Klasse) mit gleicher Wahrscheinlichkeit defekt wird und das dies unabhängig von den anderen Gruppenelementen passiert. Auch nehmen wir an dass ein Element in Falle eines Defektes oder Schadenfalls die Gruppe verlässt.

Der Binomial – Beta Fall

Modellannahmen:

- Unter der Bedingung dass der Parameter ϑ gegeben ist, sind die N_j, $j = 1, 2, \ldots$ unabhängig und binominal verteilt, i.e.

$$\mathbb{P}\left(N_j = k \mid P = \vartheta\right) = \binom{V_{n+1}}{k} \vartheta^k (1-\vartheta)^{V_{n+1}}.$$

- Der Parameter Θ ist Beta(a, b) verteilt mit $a, b > 0$, bzw.

$$u(\vartheta) = \frac{1}{B(a,b)} \vartheta^{a-1}(1-\vartheta)^{b-1}, \quad 0 \le \vartheta \le 1,$$

wobei gilt $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{(a+b)}$.

Der Binomial – Beta Fall

Bemerkung

- Die ersten beide Momente der Beta Verteilung lauten

$$\mathbb{E}\Theta = \frac{a}{a+b}, \quad \mathbf{Var}(\Theta) = \frac{a\,b}{(1+a+b)(a+b)^2}$$

- Wir wollen den *Ausfallrate* schätzen;
- Die Familie der Beta Verteilungen die mit a und b beschrieben werden ist ziemlich gross.

Der Binomial – Beta Fall

Proposition

Unter den obigen Modellannahmen gilt

$$
\begin{aligned}
F^{ind} &= \mathbb{E}[X_{n+1}|\Theta] = \Theta, \\
F^{coll} &= \mathbb{E}E[\Theta] = \frac{a}{a+b} \\
F^{Bayes} &= \frac{a + \sum_{j=1}^n N_j}{a+b+\sum_{j=1}^n V_j} = \alpha \bar{N} + (1-\alpha)\frac{a}{a+b} \\
&\text{wobei gilt} \quad \bar{N} = \frac{\sum_{j=1}^n N_j}{\sum_{j=1}^n V_j}, \ \alpha = \frac{\sum_{j=1}^n V_j}{a+b+\sum_{j=1}^n V_j}.
\end{aligned}
$$

Der quadratische Verlust des Schätzers F^{Bayes} beträgt

$$
\mathbb{E}(F^{Bayes} - \Theta)^2 = (1-\alpha)\mathbb{E}(F^{coll} - \Theta)^2 = \alpha \mathbb{E}(\bar{N} - \Theta)^2.
$$

Der Normal $\Leftrightarrow$ Normal Fall

Die Story

Angenommen eine Maschine produziert Bauteile einer bestimmten Länge L. Die Maschinen müssen adjustiert werden, dabei passiert es öfters das die Länge der Bauteile die erzeugt werden variieren, d.h. um einen Mittelwert schwanken. Wird so ein Bauteil gemessen, gibt es aber Messfehler, bzw. passieren bei der Produktion Fehler.

Der Normal Normal Fall

- L_j die Länge des j ten Bauteils
- die Maschine produziert Bauteile der Länge Θ

Modellannahmen:

- Unter der Bedingung dass der Parameter ϑ gegeben ist, sind die L_j, $j = 1, 2, \ldots$ unabhängig und normal verteilt mit Mittelwert ϑ und Varianz σ^2, i.e. $L \sim \mathcal{N}(\vartheta, 0.01)$.
- Der Parameter Θ ist Normalverteilt mit Mittelwert μ und Varianz τ^2.

Der Normal Normal Fall

Proposition

Unter den obigen Modellannahmen gilt

$$\begin{aligned}
P^{ind} &= \mathbb{E}[X_{n+1}|\Theta] = \Theta, \\
P^{coll} &= \mathbb{E}[\Theta] = \mu, \\
P^{Bayes} &= \frac{\tau^2\mu + \sigma^2\left(\sum_{j=1}^n X_j\right)}{\tau^2 + n\sigma^2} = \alpha\bar{X} + (1-\alpha)\mu \\
& \text{wobei gilt} \quad \bar{X} = \frac{1}{n}\sum_{j=1}^n X_j,\ \alpha = \frac{n}{n^2 + \frac{\sigma^2}{\tau^2}}.
\end{aligned}$$

Der quadratische Verlust des Schätzers F^{Bayes} beträgt

$$\mathbb{E}(F^{Bayes} - \Theta)^2 = (1-\alpha)\mathbb{E}(F^{coll} - \Theta)^2 = \alpha\mathbb{E}(\bar{X} - \Theta)^2.$$

Pareto $\Leftrightarrow$ Gamma Fall

Die Story

Man betrachtet nur Beträge von verschiedenen Forderungen die größer als eine bestimmten Schwellenwert x_0 sind. Diese Fragestellung ist für Rückversicherungen (Müncher Rück, Züricher Rück) wichtig. Aus Grenzwertsätzen weiss man, dass diese Verteilungen annähernd Pareto verteilt sind.

Pareto ⇔ Gamma Fall

Die Pareto-Verteilung liefert nur oberhalb eines Schwellenwertes x_0 positive Wahrscheinlichkeiten. Die Dichtefunktion f_X und die Verteilungsfunktion F_X einer Pareto verteilten Zufallsvariablen $X \sim Pareto(x_0; \alpha)$ mit den reelwertigen Parametern $x_0 > 0$ und $\alpha > 0$ lauten

$$f_{\alpha,x_0}(x) = \begin{cases} \alpha \cdot x_0^{\alpha} \cdot (x-\theta)^{-\alpha-1} & \text{für } x > x_0 \\ 0 \text{ sonst.} \end{cases}$$

und

$$F_X(x) = \begin{cases} 1 - \left(\frac{x}{x_0}\right)^{-\alpha} & \text{für } x > x_0 \\ 0 \text{ sonst.} \end{cases}$$

Der Erwartungswert lautet

$$\mathbb{E}X = x_0 \cdot \frac{\alpha}{\alpha - 1} \quad (\text{falls } \alpha > 1) \ ,$$

die Varianz lautet

$$\mathrm{Var}[X] = x_0^2 \cdot \frac{\alpha}{(\alpha-1)^2(\alpha-2)} \quad (\text{falls } \alpha > 2) \ .$$