PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 60 — 2009
Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek

**Anna Szymańska**
Uniwersytet Łódzki

# ZASTOSOWANIE ESTYMATORÓW BAYESOWSKICH WIELKOŚCI SZKÓD DO TARYFIKACJI *A POSTERIORI* W UBEZPIECZENIACH KOMUNIKACYJNYCH OC

## 1. Wstęp

Podstawą działalności ubezpieczeniowej jest prawidłowe szacowanie składek ubezpieczeniowych. Składki powinny być tak oszacowane, aby towarzystwo nie ponosiło strat finansowych, natomiast ubezpieczony nie płacił za dużo lub za mało.

W ubezpieczeniach komunikacyjnych OC składki są wyznaczane w dwóch etapach. Pierwszy to obliczenie składki podstawowej na podstawie czynników *a priori* (obserwowalnych czynników ryzyka takich, jak na przykład rodzaj i rok produkcji samochodu, pojemność silnika, wiek i płeć kierowcy), drugi etap to taryfikacja *a posteriori* (historia szkodowości kierowcy). Zatem w pierwszym etapie ubezpieczyciel wyznacza składkę podstawową, a w drugim etapie szacuje, ile procent stawki podstawowej powinien płacić indywidualny kierowca [Lemaire 1995].

Teoretycznie istnieje wiele możliwych taryfikacji *a posteriori*. Towarzystwa ubezpieczeniowe najczęściej stosują systemy *bonus-malus*, w których podstawą taryfikacji jest liczba szkód spowodowana przez ubezpieczonego w przeszłości. Premią za rok bezszkodowej jazdy jest 10-procentowa zniżka w stosunku do składki podstawowej, karą za spowodowanie co najmniej jednej szkody w roku jest 10-procentowa zwyżka.

Celem pracy jest prezentacja metod taryfikacji *a posteriori*, w której podstawą nie jest liczba spowodowanych szkód, ale wielkość tych szkód. Stawki składki oszacowano za pomocą estymatorów bayesowskich. Tak oszacowane stawki składek są korzystniejsze zarówno dla ubezpieczyciela, jak i kierowców powodujących drobne szkody. Jednak zaufanie towarzystwa ubezpieczeniowego do tak oszacowanych składek powinno być duże tylko w odniesieniu do kierowców o kilkuletniej historii szkodowości.

Niech $X_{ij}$ oznacza całkowitą kwotę wypłaconych odszkodowań dla $i$-tego ubezpieczonego w $j$-tym roku trwania ubezpieczenia. Załóżmy, że ubezpieczyciel posiada obserwacje $x_{ij}$, $i = 1,..., N$, $j = 1,...,t$ będące realizacjami zmiennych losowych $X_{ij}$. Kwoty wypłat $x_{i,t+1}$ w roku $t + 1$ nie są znane.

Załóżmy, że dla każdego $i$ rozkład zmiennej losowej $X_{ij}$ zależy od parametru $\theta_i$ oraz że zmienne losowe $X_{ij}$ przy danym $\Theta_i = \theta_i$ są niezależne i mają jednakowy rozkład o dystrybuancie $F(x|\theta_i) = \Pr(X_{ij} \le x|\Theta_i = \theta_i), j = 1,...,t$. Wektor losowy $\mathbf{X}_i = (X_{i1},...,X_{it})$ oznacza indywidualną historię ubezpieczenia dla polisy $i$ w portfelu złożonym z $N$ polis. Ubezpieczyciel musi oszacować składkę netto w roku $t + 1$ dla kontraktu $i$, znając wektor $\mathbf{x}_i = (x_{i1},...,x_{it})$.

Jeżeli założy się równoważność roszczeń i składek, to składka netto $m(\theta_i)$ dla kontraktu $i$ wynosi

$$m(\theta_i) = E\left(X_{i,t+1}|\Theta_i = \theta_i\right). \tag{1}$$

Ponieważ nie znamy wartości parametru $\theta_i$ , wartość składki netto jest nieznana. Początkowo w metodach aktuarialnych w celu wyznaczenia składki netto stosowano dwa modele:

1) oszacowanie składki netto za pomocą składki kolektywnej $\mu = EX_j$;

2) oszacowanie składki netto na podstawie historii roszczeń w przeszłości, czyli jako średniej arytmetycznej $\overline{x}_i = \frac{1}{t}\sum_{j=1}^{t} x_{ij}$.

Model pierwszy jest skuteczny dla portfeli jednorodnych pod względem generowanych roszczeń. W przypadku portfeli niejednorodnych ubezpieczeni o niskiej szkodowości płacą za dużo, a o wysokiej szkodowości − za mało. Po kilku latach w portfelu zostają tylko ubezpieczeni o dużej szkodowości.

Model drugi funkcjonuje prawidłowo w portfelach niejednorodnych tylko w przypadkach, gdy historia roszczeń jest dostatecznie długa.

Na początku XX wieku składkę netto zaczęto liczyć jako średnią ważoną składki kolektywnej $\mu$ i indywidualnej składki $\overline{x}_i$ oszacowanej na podstawie historii roszczeń w przeszłości, czyli jako

$$m(\theta_i) = Z_i\overline{x}_i + (1 - Z_i)\mu, \tag{2}$$

gdzie $Z_i \in [0,1]$. Tak zdefiniowaną składkę $m(\theta_i)$ nazywa się *składką zaufania* dla $i$-tego kontraktu, natomiast $Z_i$ − *współczynnikiem zaufania*. Estymator zmiennej $X_{i,t+1}$ nazywany jest *predyktorem* tej zmiennej, natomiast wartość predyktora − prognozą dla $X_{i,t+1}$ na podstawie obserwacji $x_{i1},...,x_{it}$ [Jasiulewicz 2005].

## 2. Bayesowska teoria zaufania

Podstawą teorii zaufania jest bayesowska analiza statystyczna z kwadratową funkcją straty.

Załóżmy, że zmienna losowa $X_j$ oznacza wielkość roszczeń pojedynczego ubezpieczonego w $j$-tym roku ubezpieczenia i ma funkcję gęstości $f(x_j)$. Rozkład zmiennej $X_j$ zależy od parametru $\theta$.

Zakłada się, że parametr ryzyka $\theta$ ubezpieczonego jest taki sam przez cały okres trwania ubezpieczenia i jest realizacją zmiennej losowej $\Theta$ o funkcji gęstości rozkładu prawdopodobieństwa $\pi(\theta)$.

Niech $f\left(x_j | \theta\right)$ będzie funkcją gęstości warunkowego rozkładu zmiennej losowej $X_j$ przy warunku $\Theta = \theta$. Indywidualna składka netto dla ubezpieczonego o parametrze ryzyka $\Theta = \theta$ jest wartością oczekiwaną rozkładu warunkowego $f\left(x_j | \theta\right)$ i wynosi $m(\theta) = E\left(X_j | \theta\right)$.

W celu oszacowania przyszłego roszczenia $(X_{t+1} | \theta)$ ubezpieczonego z nieznanym parametrem $\theta$ minimalizuje się wartość oczekiwaną kwadratowej funkcji straty $\left(X_{t+1} - d(X_1,...,X_t)\right)^2$, gdzie $d(x_1,...,x_t)$ oznacza dowolną funkcję rzeczywistą określoną na przestrzeni obserwacji. Najmniejszy przeciętny błąd kwadratowy

$$\min_d E\left(X_{t+1} - d(X_1,...,X_t)\right)^2 = E\left(X_{t+1} - d^*(X_1,...,X_t)\right)^2, \tag{3}$$

jest osiągnięty dla predyktora bayesowskiego o postaci:

$$d^*(X_1,...,X_t) = E\left(X_{t+1} | X_1,...,X_t\right). \tag{4}$$

Oznacza to, że prognozą przyszłego roszczenia $X_{t+1}$ o minimalnym błędzie średniokwadratowym dla danych $X_1 = x_1,...,X_t = x_t$ jest wartość warunkowej wartości oczekiwanej $E\left(X_{t+1} | X_1,...,X_t\right)$.

Jeżeli predyktor bayesowski $E\left(X_{t+1} | X_1,...,X_t\right)$ zmiennej losowej $X_{t+1}$ ma postać

$$E\left(X_{t+1} | X_1,...,X_t\right) = Z\overline{X} + (1-Z)\mu, \tag{5}$$

dla $Z \in [0,1]$, $\overline{X} = \frac{1}{t}\sum_{j=1}^{t} X_j$, $\mu = EX_j$, to nazywa się go *bayesowskim predyktorem zaufania* [Jasiulewicz 2005].

Jeżeli współczynnik zaufania $Z$ jest bliski zera, to oznacza to, że składkę kolektywną powinno się obdarzać większym zaufaniem niż indywidualną średnią z obserwacji. Współczynnik zaufania $Z$ bliski jedności oznacza, że ubezpieczyciel większe zaufanie powinien mieć do indywidualnej średniej z obserwacji. Współczynnik zaufania $Z$ jest w przybliżeniu równy jeden, gdy historia ubezpieczenia dla danej polisy jest długa i wykazuje małą zmienność w czasie lub gdy kontrakty są bardzo zróżnicowane.

Predyktor beyesowski $E\left(X_{t+1}\middle|X_1,...,X_t\right)$ jest bayesowską prognozą zaufania, jeżeli ma postać (5).

## 3. Bayesowskie predyktory wielkości roszczeń

Niech $X_j$ będzie zmienną losową oznaczającą wielkość roszczeń w roku $j$ dla danej polisy oraz $(X_1, X_2,\ldots,X_t)$ − wektorem obserwacji wielkości roszczeń przez $t$ lat dla danej polisy. Niech rozkład zmiennej losowej $X_j$ zależy od parametru $\theta$. W ubezpieczeniach parametr $\theta$ jest nazywany parametrem ryzyka i reprezentuje wszystkie nieuwzględnione w taryfikacji indywidualne cechy ryzyka ubezpieczonego. Zakłada się, że parametr ryzyka $\theta$ ubezpieczonego jest taki sam przez cały okres trwania ubezpieczenia i jest realizacją zmiennej losowej $\Theta$.

Załóżmy, że zmienne losowe $X_j$ dla ustalonego $\Theta=\theta$ są niezależne, mają jednakowe rozkłady oraz że ubezpieczeni generują straty niezależnie od siebie.

Zakładamy, że ubezpieczyciel zna wielkość roszczeń ubezpieczonego z okresu $t$ lat. Niech $X_{t+1}$ będzie nieznaną wielkością szkód w roku $t + 1$ dla polisy opisanej wektorem obserwacji $(X_1, X_2,\ldots, X_t)$, który można oszacować za pomocą estymatora bayesowskiego.

Niech zmienna losowa $X$ ma rozkład wykładniczy z parametrem $\theta$ o funkcji gęstości o postaci:

$$f(x|\theta)=\theta e^{-\theta x}\ ,x>0,\theta>0. \tag{6}$$

Jeżeli parametr $\Theta$ jest zmienną losową o rozkładzie gamma z parametrami $\alpha$ i $\beta$, o funkcji gęstości o postaci

$$\pi(\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\theta}\theta^{\alpha-1},\quad \alpha>0,\beta>0,\theta>0, \tag{7}$$

to zmienna losowa $X$ ma rozkład o funkcji gęstości o postaci

$$f(x)=\int_0^\infty f(x|\theta)\pi(\theta)d\theta=\int_0^\infty \theta e^{-\theta x}\frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\theta}\theta^{\alpha-1}d\theta=\frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty\theta^\alpha e^{-(\beta+x)\theta}d\theta=\frac{\alpha\beta^\alpha}{\left(\beta+x\right)^{\alpha+1}}$$

$$f(x)=\frac{\alpha\beta^\alpha}{\left(\beta+x\right)^{\alpha+1}},\quad x>0. \tag{8}$$

Jest to funkcja gęstości zmiennej losowej o rozkładzie Pareto z parametrami $\alpha$ i $\beta$, odpowiednio o wartości oczekiwanej i wariancji [Hossack, Pollard, Zehnwirth 1999]:

$$EX = \frac{\beta}{\alpha - 1}, \quad \alpha > 1, \tag{9}$$

$$D^2 X = \frac{\alpha\beta^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad \alpha > 2\cdot \tag{10}$$

Estymatory parametrów $\alpha$ i $\beta$ wyznaczone metodą momentów odpowiednio wynoszą:

$$\tilde{\alpha} = \frac{2S_x^2}{S_x^2 - \overline{x}^2}, \tag{11}$$

$$\tilde{\beta} = \overline{x}\,\frac{S_x^2 + \overline{x}^2}{S_x^2 - \overline{x}^2}, \tag{12}$$

gdzie: $\overline{x}$ – jest średnią wielkością roszczeń w portfelu,
$S_x^2$ – wariancją wielkości roszczeń w portfelu [Domański, Pruska 2000].

Gęstość *a posteriori* $\pi_x^*(\theta)$ obliczana ze wzoru Bayesa ma postać:

$$\pi_x^*(\theta) = \frac{\pi(\theta)\prod_j f_\theta(x_j)}{f(x_1,\ldots x_t)} = \frac{\left(\tilde{\beta} + \sum_j x_j\right)}{\Gamma\left(\tilde{\alpha} + t\right)} e^{-\theta\left(\tilde{\beta} + \sum_j x_j\right)} \theta^{\tilde{\alpha} + t - 1}. \tag{13}$$

Rozkład *a posteriori* parametru $\Theta$ jest więc rozkładem gamma o parametrach $\hat{\alpha}$ i $\hat{\beta}$ o postaci:

$$\hat{\alpha} = \tilde{\alpha} + t \quad \text{i} \quad \hat{\beta} = \tilde{\beta} + \sum x_j\ . \tag{14}$$

Rozkład predyktywny wielkości roszczeń $X_{t+1}$ przy warunku $X_1 = x_1, \ldots, X_t = x_t$ ma postać:

$$f\left(x \middle| x_1, \ldots, x_t\right) = \int f\left(x \middle| \theta\right) \pi\left(\theta \middle| x_1, \ldots, x_t\right) d\theta = \frac{\hat{\alpha}\hat{\beta}^{\hat{\alpha}}}{\left(\hat{\beta} + x\right)^{\hat{\alpha}+1}}\cdot \tag{15}$$

Jest to rozkład Pareto o parametrach $\hat{\alpha}$ *i* $\hat{\beta}$. Wartość oczekiwana rozkładu wyraża się wzorem

$$E\left(X_{t+1} \middle| X_1, \ldots, X_t\right) = \frac{\hat{\beta}}{\hat{\alpha} - 1} = \frac{\tilde{\beta} + \sum x_j}{\tilde{\alpha} + t - 1}\cdot \tag{16}$$

Dla takich rozkładów $f\left(x\ \middle|\theta\right)$ i $\pi(\theta)$ mamy:

$$\mu = \frac{\tilde{\beta}}{\tilde{\alpha}-1}, Z = \frac{t}{t+\tilde{\alpha}-1},$$

$$m(\theta) = Z\overline{x} + (1-Z)\mu = \frac{t}{t+\tilde{\alpha}-1}\frac{1}{t}\sum x_j + \left(1 - \frac{t}{t+\tilde{\alpha}-1}\right)\frac{\tilde{\beta}}{\tilde{\alpha}-1} = \frac{\tilde{\beta}+\sum x_j}{\tilde{\alpha}+t-1},$$

stąd indywidualna składka w roku $t + 1$ o postaci:

$$m(\theta) = \frac{\tilde{\beta}+\sum x_j}{\tilde{\alpha}+t-1}, \tag{17}$$

jest bayesowskim predyktorem zaufania.

## 4. Szacowanie stawek składki netto

W ubezpieczeniach komunikacyjnych OC indywidualna składka netto w okresie $t + 1$ wynosi

$$m(\theta) = (EX)\cdot(E\Lambda)\cdot b_{t+1}(x_1,...,x_t), \tag{18}$$

gdzie: $m(\theta)$ – indywidualna składka netto w okresie $t$+1,
$(EX)$ – wartość oczekiwana wielkości pojedynczej szkody,
$(E\Lambda)$ – wartość oczekiwana liczby szkód w okresie rozrachunkowym (w ciągu jednego roku),
$b_{t+1}(x_1,...,x_t)$ – stawka szacowanej składki [Lemaire 1995].

Jeżeli założy się , że $EX = \frac{\tilde{\beta}}{\tilde{\alpha}-1}$ oraz $E\Lambda = 1$, to równanie (18) ma postać:

$$m(\theta) = \frac{\tilde{\beta}}{\tilde{\alpha}-1}\cdot b_{t+1}(x_1,...,x_t), \tag{19}$$

stąd kierowca, który po $t$ latach zgłosił roszczenia wielkości $\sum_{j=1}^{t} x_j$, powinien płacić stawkę szacowanej składki równą

$$b_{t+1}(x_1,...,x_t) = \frac{\tilde{\alpha}-1}{\tilde{\beta}} m(\theta)\cdot 100\%\cdot \tag{20}$$

Najprostszą zasadą kalkulacji składki w ubezpieczeniach komunikacyjnych jest zasada wartości oczekiwanej [Lemaire 1995]. Według tej zasady szacowana indywidualna składka netto powiększona o dodatek bezpieczeństwa Q wynosi:

$$m(\theta) = (1+Q)E\left(X_{t+1} \middle| X_1, ..., X_t\right) = (1+Q)\frac{\tilde{\beta} + \sum x_j}{\tilde{\alpha} + t - 1}. \tag{21}$$

Na podstawie wzorów (20) i (21) kierowca, który po $t$ latach zgłosił roszczenia w wysokości $\sum_{j=1}^{t} x_j$ w roku $t + 1$, powinien płacić stawkę składki równą

$$b_{t+1}(x_1, ..., x_t) = (1+Q)\frac{(\tilde{\alpha}-1)(\tilde{\beta} + \sum x_j)}{\tilde{\beta}(\tilde{\alpha} + t - 1)} \cdot 100\%. \tag{22}$$

Jeśli się przyjmie, że Q = 0, to otrzyma się

$$b_{t+1}(x_1, ..., x_t) = \frac{(\tilde{\alpha}-1)(\tilde{\beta} + \sum x_j)}{\tilde{\beta}(\tilde{\alpha} + t - 1)} \cdot 100\%. \tag{23}$$

## 5. Zastosowania

Badanie przeprowadzono w łódzkim towarzystwie ubezpieczeniowym, które zastrzegło sobie anonimowość. Z jednego z portfeli ubezpieczeń komunikacyjnych OC złożonego z 10230 polis wylosowano 100 polis. Średnia roczna wielkość roszczeń wynosiła 5,05 tys. zł z odchyleniem standardowym 7,1 wynoszącym tys. zł (parametry rozkładu wielkości roszczeń $\tilde{\alpha} = 4,048$ $,\tilde{\beta} = 15,391$). Założono, że rozkład wielkości szkód w tym portfelu jest zbliżony do rozkładu Pareto. Dla wylosowanej próby oszacowano stawki składek za pomocą estymatorów bayesowskich oraz obliczono współczynniki zaufania. Wyniki przedstawiają tab. 1 i 2.

Tabela 1. Stawka składki netto $b_{t+1}(x_1, ..., x_t)$ w roku $t$ + 1 w zależności od czasu trwania ubezpieczenia $t$ i sumy $\sum_{j=1}^{t} x_j$ roszczeń zgłoszonych w latach 1,..., $t$

| $\sum_{j=1}^{t} x_j$ (w tys. zł) | *t* | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Poniżej 0,2 | 100% | 80% | 67% | 58% | 51% |
| 0,2-1 | 105% | 84% | 71% | 61% | 53% |
| 1-2 | 112% | 90% | 75% | 64% | 56% |
| 2-3 | 118% | 95% | 79% | 68% | 60% |
| 3-4 | 124% | 100% | 83% | 72% | 63% |
| 4-5 | 131% | 105% | 88% | 75% | 66% |
| 5-6 | 137% | 110% | 92% | 79% | 69% |
| 6-7 | 144% | 115% | 96% | 83% | 73% |

Źródło: obliczenia własne.

Tabela 2. Współczynnik zaufania $Z$ do obserwacji $x_1,...,x_t$ w zależności od czasu trwania ubezpieczenia $t$ i sumy $\sum_{j=1}^{t} x_j$ roszczeń zgłoszonych w latach 1,..., $t$

| $\sum_{j=1}^{t} x_j$ (w tys. zł.) | $t$ | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Poniżej 0,2 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 0,2-1 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 1-2 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 2-3 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 3-4 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 4-5 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 5-6 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |
| 6-7 | 0,24 | 0,39 | 0,49 | 0,56 | 0,62 |

Źródło: obliczenia własne.

## 6. Podsumowanie

Oszacowane stawki składek wahają się od 51 do 144% składki podstawowej. Oznacza to, że w tak funkcjonującym systemie *bonus-malus* największa zniżka wynosiłaby 51% składki podstawowej, a największa zwyżka − 144% składki podstawowej. W tym systemie przyjęto, że kierowca, który w pierwszym roku ubezpieczenia spowodował szkodę nie większą niż 200 zł, płaci 100% składki podstawowej, ale ubezpieczyciel ma tutaj dość duże pole manewru, przyjmując w opisywanej klasie na przykład 105-procentową zwyżkę. Wówczas wszystkie stawki wzrosną.

Taryfikacja na podstawie liczby szkód może być niekorzystna zarówno dla ubezpieczyciela, jak i ubezpieczonego. Ubezpieczyciel nie rozróżnia kierowców powodujących małe i wielkie szkody, stosując takie same zwyżki i zniżki w stosunku do kierowcy powodującego szkodę w wysokości 500 zł i 8000 zł. Nie można w tym przypadku powiedzieć, że idea ubezpieczenia jest „sprawiedliwa" dla kierowcy powodującego niewielkie szkody.

Stawki składek szacowane na podstawie estymatorów wielkości szkód są korzystniejsze dla ubezpieczonych powodujących nawet kilka niewielkich szkód. W przypadku taryfikacji na podstawie liczby szkód kierowcy powodujący wiele szkód o małej wartości będą płacić duże zwyżki. Jednak i taka taryfikacja ma wady. Spowodowanie jednej dużej szkody powoduje wysokie zwyżki, a współczynniki zaufania są zadowalające dla kierowców z długą historią szkodowości. Na podstawie tab. 2 można stwierdzić, że po 5 latach ubezpieczyciel może mieć 62% zaufania do indywidualnej średniej wielkości szkód, a 38% zaufania do kolektywnej średniej wielkości szkód.

# Literatura


Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.

Hossack I.B., Pollard J.H., Zehnwirth B., *Introductory statistics with applications in general insurance*, Cambridge 1999.

Jasiulewicz H., *Teoria zaufania. Modele aktuarialne*, Wydawnictwo AE we Wrocławiu, Wrocław 2005.

Lemaire J., *Bonus-malus systems in automobile insurance*, Kluwer Nijhoff, Boston 1995.


## APPLICATION OF THE BAYESIAN ESTIMATORS OF THE SIZE OF DAMAGES TO THE POSTERIOR TARIFFICATION IN CR AUTOMOBILE LIABILITY INSURANCE


### Summary

In the paper an application of the bayesian estimators of the size of damages to the posterior tariffication in CR automobile liability insurance is presented. Net premiums are calculated by means of the expected value method with assumption that distribution of the size of damages is the Pareto distribution. The rates of net premium are estimated by use of bayesian estimators. The research was performed on real data from an insurance company from Łódź.