Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 227 · 2015

Michał Trzęsiok

Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
Katedra Analiz Gospodarczych i Finansowych
michal.trzesiok@ue.katowice.pl

Alicja Wolny-Dominiak

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Metod Statystyczno-Matematycznych w Ekonomii
alicja.wolny-dominiak@ue.katowice.pl

ZŁOŻONY MIESZANY ROZKŁAD POISSONA – ZASTOSOWANIA UBEZPIECZENIOWE

Streszczenie: Celem niniejszej pracy jest zaproponowanie pewnego złożonego mieszanego rozkładu Poissona, korzystając z tego, że złożony rozkład Poisson–gamma jest szczególnym przypadkiem rozkładu Tweedie. Wykorzystujemy fakt, iż z kolei rozkład Tweedie należy do dyspersyjnej rodziny rozkładów wykładniczych. W tym celu w pierwszej części pracy rozważamy mieszany rozkład Poissona, w którym uwzględniamy zmienną nieobserwowalną. Następnie charakteryzujemy złożony mieszany rozkład Poissona. W drugiej części przedstawiamy możliwość wykorzystania tego rozkładu w szacowaniu oczekiwanej łącznej wartości szkód dla pojedynczego ryzyka w masowym portfelu ryzyk.

Słowa kluczowe: mieszany rozkład Poissona, złożony mieszany rozkład Poissona, ubezpieczenia, taryfikacja, estymacja.

Wprowadzenie

Rozkład Poissona odgrywa bardzo znaczącą rolę w modelowaniu zjawisk uwzględniających zmienne zliczające. Przyczyny tego faktu tkwią w szczególnej przydatności tego rozkładu w przypadkach, kiedy modelowana zmienna ma charakter *stricte* losowy, a rozpatrywana populacja jest jednorodna. Niestety w przypadku wielu rzeczywistych zbiorów danych przyjmowanie takich założeń jest nierealistyczne. Dla danych niejednorodnych alternatywą do modelowania z wykorzystaniem klasycznego rozkładu Poissona jest zastosowanie mieszanych rozkładów Poissona. Na szczególną uwagę zasługuje przypadek, w którym badana populacja w istocie składa się ze skończonej liczby jednorodnych podpopulacji. Rozkład prawdopodobieństwa zmiennej w całej badanej populacji można wówczas traktować jako mieszaninę rozkładu Poissona oraz rozkładu pewnej zmiennej nieobserwowalnej θ. Przykładem jest najpopularniejsza mieszanina Poisson–gamma.

Szczególnym przypadkiem jest analiza danych ubezpieczeniowych dotyczących krótkoterminowych ubezpieczeń majątkowych, takich jak ubezpieczenia komunikacyjne czy ubezpieczenia nieruchomości (mieszkań, domów itd.). Problem nieobserwowalnej (ukrytej) niejednorodności w danych ubezpieczeniowych w ubezpieczeniach komunikacyjnych wynika na przykład z faktu, że aktuariusz nie ma dostępu do danych dotyczących różnic w zachowaniach oraz stylu prowadzenia pojazdów poszczególnych kierowców. Jedną ze znanych konsekwencji owej niejednorodności w danych dla zmiennych zliczających jest tzw. efekt nadmiernej dyspersji, który oznacza, że wariancja tej zmiennej zliczającej jest większa niż jej wartość przeciętna. Oprócz wspomnianych konsekwencji wystąpienia niejednorodności dla wartości momentów niskich rzędów badanej zmiennej, ta ukryta niejednorodność wpływa również istotnie na postać estymowanego rozkładu. Dla wielu zbiorów danych ubezpieczeniowych charakterystyczne jest (wynikające właśnie z ukrytej niejednorodności) występowanie bardzo wielu wartości zerowych, jak również bardzo silna asymetria prawostronna (gruby prawy ogon rozkładu).

Ukrytą niejednorodność uwzględnia się w modelowaniu przez nałożenie dodatkowego założenia, że wartość przeciętna (parametr) rozkładu Poissona nie jest stały, lecz jest zmienną losową o pewnym rozkładzie (w modelowaniu mówimy wtedy o uwzględnieniu tzw. efektu losowego). Budowane w ten sposób modele nazywamy modelami mieszanek Poissona. Na przykład w modelowaniu szkód roczna oczekiwana frakcja polis związanych z wypłatą odszkodowania może być traktowana jako zmienna losowa.

Celem niniejszego artykułu jest zaproponowanie zastosowania złożonego mieszanego rozkładu Poissona w krótkoterminowych ubezpieczeniach majątkowych, w których dysponujemy masowym portfelem ryzyk. Pojedyncze ryzyko w portfelu rozumiemy tu jako zmienną losową o określonym rozkładzie prawdopodobieństwa. Może to być np. liczba szkód, wartość pojedynczej szkody czy łączna wartość szkód dla pojedynczego ryzyka. W pierwszej części definiujemy rozkłady: mieszany Poissona, złożony Poissona, złożony mieszany Poissona. Definiujemy wektor parametrów rozkładów, wyznaczamy pierwsze dwa momenty oraz omawiamy możliwe metody estymacji parametrów tych rozkładów. W drugiej części pracy przedstawiamy przykład empiryczny. Obliczenia wykonujemy korzystając z programu **R** [zob. R Core Team, 2014].

1. Mieszany rozkład Poissona (*MPOIS*)

Rozważmy dyskretną zmienną losową N o rozkładzie Poissona z parametrem λ (ozn. *POIS*). Dobrze znana wartość funkcji prawdopodobieństwa ma wtedy postać:

$$P[N = k] = \frac{\lambda^k \exp(-\lambda)}{k!}, \quad k = 0,1,2,\ldots \tag{1}$$

Łatwo można wykazać, że dla rozkładu Poissona zachodzi następująca równość:

$$E[N] = Var[N] = \lambda\,. \tag{2}$$

W wielu zagadnieniach, również ubezpieczeniowych, specyfika danych powoduje, iż często równanie (2) nie jest spełnione. Występuje wtedy, wspomniany we wprowadzeniu, efekt nadmiernej dyspersji. Efekt ten jest tłumaczony najczęściej faktem, iż nie uwzględnia się pewnej nieobserwowalnej zmiennej Θ wpływającej na zmienną N. Aby uwzględnić tę zmienną, randomizuje się parametr λ rozkładu Poissona poprzez wprowadzenie zmiennej losowej Θ, która może mieć zarówno rozkład dyskretny, jak również ciągły. Uzyskuje się wtedy mieszany rozkład Poissona (ozn. *MPOIS*), będący szczególnym przypadkiem mieszaniny rozkładów [zob. Winkelmann, 2003, s. 33, 134]. Brzegowa funkcja prawdopodobieństwa dana jest wtedy wzorem:

$$P[N = k] = \int_{\theta \in \Theta^c} P[N = k \mid \theta] f_\theta(\theta) d\theta, \tag{3}$$

gdzie $P[N = k \mid \theta]$ jest funkcją prawdopodobieństwa rozkładu Poissona zmiennej $N \mid \theta \sim POIS(\lambda\theta)$, $f_\theta(\cdot)$ jest gęstością rozkładu zmiennej nieobserwowalnej θ, natomiast Θ^c jest zbiorem wszystkich możliwych zmiennych Θ. Zachodzi więc równość:

$$E[N \mid \theta] = Var[N \mid \theta] = \lambda\,. \tag{4}$$

Wtedy wartość oczekiwana i wariancja rozkładu *MPOIS* wynoszą:

$$E[N] = E[E[N \mid \theta]] = E[\lambda\theta] = \lambda E[\theta] \tag{5}$$

$$\begin{aligned} Var[N] &= E[Var[N \mid \theta]] + Var[E[N \mid \theta]] = E[\lambda\theta] + Var[\lambda\theta] = \\ &= \lambda E[\theta] + \lambda^2 Var[\theta] \end{aligned}\,. \tag{6}$$

Wektor parametrów $\boldsymbol{\rho}$ rozkładu $MPOIS$ składa się zatem z parametru λ warunkowego rozkładu zmiennej $N \mid \theta$ oraz parametrów rozkładu zmiennej θ.

Najbardziej popularnym w zagadnieniach ubezpieczeniowych rozkładem $MPOIS$ jest mieszanina Poisson–gamma, w której uzyskuje się rozkład ujemny dwumianowy z wektorem parametrów $\boldsymbol{\rho} = (\lambda, \alpha)'$ (ozn. NB). W rozkładzie NB zakłada się, iż zmienna nieobserwowalna θ ma rozkład gamma, dla którego $E[\theta] = 1$, natomiast $Var[\theta] = \alpha$. Łatwo wykazać, iż tak skonstruowany rozkład NB charakteryzuje się zawsze efektem nadmiernej dyspersji. Korzystając z założenia o rozkładzie warunkowym $N \mid \theta \sim POIS(\lambda\theta)$, brzegowa wartość oczekiwana i brzegowa wariancja zmiennej N wynoszą:

$$E[N] = \lambda, \tag{7}$$

$$Var[N] = \lambda + \lambda^2 Var[\theta] = \lambda(1 + \lambda\alpha). \tag{8}$$

Podobną konstrukcją jak wyżej charakteryzują się również inne popularne mieszaniny, w których zakłada się różnorodne rozkłady dla zmiennej nieobserwowalnej. Tab. 1 zawiera stosowne zestawienie.

Tab. 1. Zestawienie wybranych mieszanek rozkładów Poissona

Nazwa rozkładu *MPOIS* ze wskazaniem rozkładu zmiennej nieobserwowalnej
Hermite'a (Poisson–normalny)
Poisson–Odwrotny–normalny
Lindleya (Poisson–Lindley)
Poisson–Ujemny dwumianowy
Neymana (Poisson–Poisson)
Poisson–Logarytmiczno–normalny

Źródło: Na podstawie: [Karlis, 2001].

Estymacja parametrów $\boldsymbol{\rho}$ mieszanego rozkładu Poissona $MPOIS$ jest mniej lub bardziej skomplikowana w zależności od założonego rozkładu zmiennej θ. W przypadku gdy dla konkretnej mieszaniny całkę (3) można obliczyć analitycznie, jak np. w przypadku rozkładu ujemnego dwumianowego, bezpośrednie zastosowanie znajduje metoda największej wiarygodności. Jeśli natomiast całka ta nie ma swojej analitycznej postaci, można wykorzystać takie algorytmy, jak algorytm *EM* (*Expectation–Maximization algorithm*) [Dempster, Laird, Rubin, 1977] czy algorytm *H–IWSL* (*Hierarchical Iterative Weighted Least Square*) [Lee, Nelder, 1996].

Przedstawiony powyżej rozkład *MPOIS* znajduje bezpośrednie zastosowanie w zagadnieniu taryfikacji (grupowania ryzyk) w krótkoterminowych ubezpieczeniach majątkowych, takich jak ubezpieczenia komunikacyjne czy ubezpieczenia domów/mieszkań. Istotne jest wtedy szacowanie wartości $E[N_i]$, gdzie N_i oznacza liczbę szkód wygenerowaną przez *i*-te ryzyko w portfelu ryzyk [Denuit i in., 2007; Antonio, Valdez, 2012; Wolny-Dominiak, 2013]. W punkcie 3. rozważamy liczbę szkód dla portfela ryzyk, gdzie parametry rozkładu *MPOIS* szacujemy metodą *EM*, gdzie postać *E*-kroku oraz *M*-kroku została zaczerpnięta z pracy Karlisa [2001].

2. Złożony mieszany rozkład Poissona (*CMPOIS*)

Rozważmy zmienną losową Y skonstruowaną następująco:

$$Y = \sum_{k=1}^{N} V_k , \tag{9}$$

gdzie N jest zmienną dyskretną o rozkładzie Poissona, zmienne V_k, $k = 1,\dots,N$ są niezależnymi zmiennymi losowymi o tych samych rozkładach oraz niezależnymi od N. Ponadto $Y = 0$, jeśli $N = 0$. Wtedy zmienna *Y* posiada złożony rozkład Poissona (ozn. *CPOIS*). Wartość oczekiwana $E[Y]$ oraz wariancja $Var[Y]$ są dane wzorem:

$$E[Y] = E[V_1]E[N], \tag{10}$$

$$Var[Y] = E^2[V_1]Var[N] + E[N]Var[V_1]. \tag{11}$$

Oczywiście również i w tym przypadku może wystąpić efekt nadmiernej dyspersji dla zmiennej *N*, tłumaczony występowaniem nieobserwowalnej zmiennej θ. W celu uwzględnienia tego efektu w rozkładzie zmiennej *Y*, przyjmuje się więc założenie, iż N ma mieszany rozkład Poissona $N \sim MPOIS$. Uzyskuje się wtedy złożony mieszany rozkład Poissona (ozn. $CMPOIS$). Korzystając z tego, iż $N \mid \theta \sim POIS(\lambda\theta)$, można wyprowadzić wartość oczekiwaną i wariancję zmiennej $Y \sim CMPOIS$:

$$E[Y] = E[V_1]E[N] = E[V_1]E[E[N \mid \theta]] = \lambda E[V_1]E[\theta], \tag{12}$$

$$\begin{aligned} Var[Y] &= E^2[V_1]Var[N] + E[N]Var[V_1] = \\ &= \lambda\{E^2[V_1](E[\theta] + \lambda Var[\theta]) + Var[V_1]E[\theta]\} \end{aligned}. \tag{13}$$

Wektor parametrów $\boldsymbol{\rho}$ rozkładu *CMPOIS*, analogicznie jak dla *MPOIS*, zawiera parametr λ warunkowego rozkładu zmiennej $N \mid \theta$ oraz parametru rozkładu zmiennej θ.

Funkcja gęstości rozkładu złożonego nie ma swojej analitycznej postaci, co prowadzi do komplikacji estymacji parametrów tego rozkładu. Bezpośrednie zastosowanie znajduje tutaj metoda *PQL* [Breslow, Clayton, 1993]. Wynika to z faktu, iż w metodzie tej korzysta się jedynie z dwóch pierwszych momentów rozkładu.

W zastosowaniach ubezpieczeniowych często przyjmuje się rozkład gamma dla zmiennych V_k. Wtedy alternatywnym rozwiązaniem może być wykorzystanie pewnej własności złożonego rozkładu Poisson–gamma. W pracy [Jorgensen, Paes De Souza, 1994] autorzy przedstawili ten rozkład jako szczególny przypadek rozkładu Tweedie z parametrem $1 < p < 2$. Rozkład Tweedie należy do dyspersyjnej rodziny rozkładów wykładniczych i charakteryzuje się tym, iż funkcja wariancji $V(\mu) = \mu^p$. Rozkład ten posiada zatem trzy parametry: $\boldsymbol{\rho} = (\mu, \phi, p)'$, gdzie ϕ jest parametrem dyspersji (o dyspersyjnej rodzinie rozkładów wykładniczych zob. [Jorgensen, 1987]). Przyjmując, że:

1) N jest zmienną licznikową o rozkładzie Poissona z parametrem λ,
2) zmienne $Y_k, k = 1, \ldots, N$ są niezależne o rozkładzie gamma z parametrem skali α' i parametrem kształtu β' oraz niezależne od N,

parametry złożonego rozkładu Poisson–gamma jako przypadku rozkładu Tweedie mają postać:

$$\mu = \lambda\alpha'\beta', \quad p = \frac{\beta'+2}{\beta'+1}, \quad \phi = \frac{\lambda^{1-p}(\alpha'\beta')^{2-p}}{2-p}. \tag{14}$$

Wtedy $E[Y] = \mu$ oraz $Var[Y] = \phi\mu^p$. Łatwo dalej uzyskać złożony rozkład mieszany *CMPOIS* poprzez randomizację parametru μ, przyjmując, iż zmienna N ma rozkład *MPOIS* oraz $N \mid \theta \sim POIS(\lambda\theta)$. Wektor parametrów rozkładu *CMPOIS* składa się z: wektora $(p, \phi)'$, parametru λ warunkowego rozkładu zmiennej $N \mid \theta$ oraz parametrów rozkładu zmiennej θ.

Istotą powyższego podejścia jest fakt, iż wektor parametrów $\boldsymbol{\rho}$ rozkładu *CPOIS* można szacować, wykorzystując klasyczny algorytm *IWSL* stosowany w modelach regresyjnych klasy *GLM* [Mccullagh, Nelder, 1989]. W przypadku rozkładu *CMPOIS* można z kolei wykorzystać wspomniany już wcześniej algorytm *PQL* lub hierarchiczny algorytm *H–IWSL* [Lee, Nelder, 1996; Wolny-Dominiak, 2013]. W każdej z tych metod nie jest wymagana znajomość postaci funkcji gęstości, a jedynie dwóch pierwszych momentów. Alternatywnie można bezpośrednio zastosować funkcję brzegowej wiarygodności, co wymaga całkowania numerycznego [Zhang, 2013].

Omówiony rozkład *CMPOIS* (w szczególności jako przypadek rozkładu Tweedie), podobnie jak rozkład *MPOIS*, znajduje bezpośrednie zastosowanie w taryfikacji. W tym przypadku poszukuje się wartości $E[Y_i]$, gdzie Y_i oznacza łączną wartość szkód wygenerowaną przez *i*-te ryzyko w portfelu ryzyk [Smyth, 2002; Wolny-Dominiak, 2014]. Rozkład ten jest również stosowany w zagadnieniu szacowania rezerwy szkodowej [Peters i in., 2009]. W punkcie 3. rozważamy problem szacowania $E[Y_i]$. Do estymacji parametrów rozkładu *CMPOIS* stosujemy funkcję brzegowej wiarygodności i całkowanie metodą Laplace'a.

3. Zastosowania *MPOIS* oraz *CMPOIS* w krótkoterminowych ubezpieczeniach majątkowych

Rozważmy portfel ryzyk w krótkoterminowych ubezpieczeniach majątkowych. Każdemu ryzyku w portfelu są przypisane zmienne losowe N_i oraz Y_i, oznaczające odpowiednio: liczbę szkód oraz łączną wartość szkód dla *i*-tego ryzyka. Zakładamy jednakową ekspozycję na ryzyko dla całego portfela. W celu uwzględnienia zmiennej nieobserwowalnej θ przyjmujemy odpowiednio: mieszany rozkład Poissona $N_i \sim MPOIS$ oraz złożony mieszany rozkład Poissona $Y_i \sim CMPOIS$. Zmienna θ jest interpretowana jako zmienna nieobserwowalna, opisująca osobiste cechy osoby ubezpieczającej się, wpływające na szkodowość danego ryzyka (*risk profile*). Jest ona zatem charakterystyczna dla indywidualnego ryzyka. W obliczeniach wykorzystujemy portfel zaczerpnięty z pakietu insuranceData [Wolny-Dominiak, Trzęsiok, 2014] programu **R** o nazwie ClaimsLong. Poniżej przedstawiamy przykłady proponowanych zastosowań rozkładów *MPOIS* oraz *CMPOIS* w ubezpieczeniach.

Przykład 1. Szacowanie oczekiwanej liczby szkód dla pojedynczego ryzyka w portfelu

Zakładamy rozkład *MPOIS* liczby szkód dla pojedynczego ryzyka w portfelu. Liczba szkód ma alternatywnie rozkład *MPOIS* – mieszanina Poisson–gamma lub *MPOIS* – mieszanina Poisson–Lindley. Oczywiście można to rozważać jako rozkład ujemny dwumianowy i szacować parametry rozkładu, korzystając z metody *MLE*. My przedstawiamy inne rozwiązanie, w którym korzystamy z algorytmu *EM*. Algorytm ten jest elastyczny i daje możliwość zmiany rozkładu zmiennej nieobserwowalnej.

Przyjmijmy dalej, że liczba szkód ma rozkład warunkowy $N \mid \theta \sim POIS(\theta)$. Zmienna nieobserwowalna ma z kolei rozkład $\theta \sim Gamma(\alpha, \beta)$ o funkcji gęstości $f_\theta(x) = \dfrac{x^{\alpha-1}\exp(-\beta x)\beta^\alpha}{\Gamma(\alpha)}$, $x > 0, \alpha, \beta > 0$ oraz $E[\theta] = 1$, $Var[\theta] = \dfrac{\alpha}{\beta^2}$.

Do szacowania parametrów α, β stosujemy algorytm *EM*, w którym dla danych wartości α_{old} oraz β_{old} iteracyjnie są wykonywane następujące kroki [por. Karlis, 2001, s. 11]:

1. **Krok *E***: wyznaczenie wartości:

$$t_i = E(\theta_i \mid x_i) = \frac{x_i + \alpha_{old}}{1 + \beta_{old}} \text{ oraz } s_i = \Psi(\alpha_{old} + x_i) - \ln(\beta_{old} + 1),$$

   dla $i = 1, \ldots, n$, gdzie $\Psi(\cdot)$ oznacza funkcję digamma.

2. **Krok *M***: etap maksymalizacji, który w przypadku tego rozkładu sprowadza się do obliczenia nowych wartości parametrów α i β, zgodnie z wzorami:

$$\beta_{new} = \frac{\alpha_{old}}{\bar{t}} \text{ oraz } \alpha_{new} = \alpha_{old} - \frac{\Psi(\alpha_{old}) + \ln(\beta_{new}) - \bar{s}}{\Psi_3(\alpha_{old})},$$

   gdzie $\Psi_3(x) = \dfrac{d\Psi}{dx}$ oznacza funkcję trigamma.

Dalej zmieniono założenie odnośnie do rozkładu zmiennej nieobserwowalnej i przyjęto rozkład Lindleya $\theta \sim LD(p)$ o funkcji gęstości $f_\theta(x) = \dfrac{p^2}{p+1}(x+1)\exp(-px)$, $x, p > 0$. Wtedy $E[\theta] = \dfrac{p+2}{p(p+1)}$ oraz $Var[\theta] = \dfrac{2(p+3)}{p^2(p+1)}$. Podobnie jak w przypadku mieszaniny Poisson–gamma, do szacowania parametru p stosujemy algorytm *EM*, w którym dla danej wartości p_{old} iteracyjnie są wykonywane następujące kroki [por. Karlis, 2001, s. 8]:

1. **Krok *E***: wyznaczenie wartości:

$$t_i = E(\theta_i \mid x_i) = \frac{(p_{old} + x_i + 3)(x_i + 1)}{(p_{old} + x_i + 2)(p_{old} + 1)}, \text{ dla } i = 1, \ldots, n.$$

2. **Krok *M***: etap maksymalizacji, który w przypadku tego rozkładu sprowadza się do obliczenia nowej wartości parametru p zgodnie ze wzorem
$$p_{new} = \frac{-(\bar{t}-1)+\sqrt{\bar{t}^2+6\bar{t}+1}}{2\bar{t}}.$$

W obliczeniach wykorzystujemy implementację własną algorytmu *EM* w programie **R**. Zmienną *N* jest liczba szkód dla pojedynczego ryzyka (claim.num), dla którego zarejestrowano przynajmniej jedną szkodę. Jako wartości początkowe oszacowywanych metodą *EM* parametrów przyjęto: dla mieszanki Poisson–gamma $\alpha_{old} = 1,\ \beta_{old} = 0{,}3$, zaś dla mieszanki Poisson–Lindley $p_{old} = 0{,}1$. Prostymi metodami symulacyjnymi sprawdzono również wrażliwość na zadane wartości początkowe przyjęte do oszacowania metodą *EM* i stwierdzono brak istotnych różnic w wyznaczonych ocenach parametrów. Zakończenie iteracyjnej procedury estymacji metodą *EM*, opisanej powyżej, następowało wtedy, gdy oszacowania nowych wartości parametrów w kroku *M* różniły się od wartości wyznaczonych w poprzedniej iteracji o nie więcej niż $\varepsilon > 0$. Przyjęto $\varepsilon = 10^{-8}$. Oszacowane parametry dla Poisson–gamma wynoszą $\hat{\alpha} = 1{,}8$ oraz $\hat{\beta} = 1{,}03$, natomiast parametr *p* w rozkładzie Lindleya $\hat{p} = 0{,}86$. Wtedy wartość oczekiwana liczby szkód dla pojedynczego ryzyka w portfelu wynosi: $E[N] = 1{,}7442$ (Poisson–gamma) oraz $E[N] = 1{,}7736$ (Poisson–Lindley). W tym bardzo uproszonym podejściu jest ona taka sama dla wszystkich ryzyk w portfelu, jednak sytuację zmieni wprowadzenie regresorów (czynników wpływających na liczbę szkód, opisujących osobę ubezpieczającą się, przedmiot ubezpieczenia oraz obszar geograficzny). Regresory spowodują zróżnicowanie portfela i w wyniku otrzymamy nie jedną wartość $E[N]$, ale kilka wartości w zależności od liczby regresorów oraz liczby ich kategorii.

Przykład 2. Szacowanie łącznej wartości szkód dla pojedynczego ryzyka w portfelu

Zakładamy rozkład *CMPOIS* łącznej wartości szkód dla pojedynczego ryzyka w portfelu, w którym liczba szkód dla pojedynczego ryzyka ma warunkowy rozkład $N \mid \theta \sim POIS(\lambda\theta)$ oraz zmienna nieobserwowalna wynosi $\theta = \exp(\theta')$, gdzie θ' ma rozkład $N(0, \sigma^2_{\theta'})$. Szacujemy zatem wektor parametrów $(\lambda, p, \phi, \sigma^2_{\theta'})'$. Z zależności (14) można dalej łatwo wyznaczyć parametry α', β' rozkładu zmiennej nieobserwowalnej.

W obliczeniach wykorzystujemy funkcję cpglmm{cplm} oraz całkowanie metodą Laplace'a. Dane dotyczą polis zawartych w kolejnych trzech okresach.

Zmienną Y jest łączna wartość szkód dla pojedynczego ryzyka (claim.cost), dla którego zarejestrowano przynajmniej jedną szkodę w przeciągu trzech okresów. Oszacowane parametry rozkładu *CMPOIS* są równe: $\lambda = 0{,}7$, $p = 1{,}51$ $\phi = 87{,}77$, $\alpha' = 0{,}97$, $\beta' = 1498{,}48$. Wyznaczona oczekiwana łączna wartość szkód dla pojedynczego ryzyka jest równa $E[Y] = 1013{,}11$. Podobnie jak w przykładzie 1., jest to przypadek prosty, w którym nie różnicujemy w żaden sposób ryzyk w portfelu poprzez np. wprowadzenie regresorów.

Podsumowanie

W modelowaniu wielu zjawisk o charakterze społeczno-ekonomicznym wymagane jest uwzględnienie nieobserwowalnej niejednorodności (heterogeniczności) występującej w danych. W pracy przedstawiono podejście, w którym ta niejednorodność jest reprezentowana przez pewną nieobserwowalną zmienną losową o zadanym przez nas rozkładzie prawdopodobieństwa. Wykorzystanie mieszanego rozkładu Poissona, przedstawianego bezpośrednio jako pewna nieskończona mieszanina, jest zdecydowanie bardziej elastycznym podejściem i daje znacznie większe możliwości uwzględniania nieobserwowalnej niejednorodności w danych poprzez różne założenia odnośnie do rozkładu zmiennej θ. O ile parametry rozkładu *MPOIS* można w miarę łatwo oszacować za pomocą algorytmu *EM*, o tyle dla rozkładu *CMPOIS* jest to ciągle problem złożony. Zaproponowane przez nas rozwiązanie, wykorzystujące rozkład Tweedie, daje praktyczne możliwości modelowania łącznej wartości szkód dla pojedynczego ryzyka w portfelu z uwzględnieniem nieobserwowalnej niejednorodności.

Literatura

Antonio K., Valdez E.A. (2012), *Statistical Concepts of a priori and a posteriori Risk Classification in Insurance*, „Advances in Statistical Analysis", 96(2), s. 187-224.

Breslow N.E., Clayton D.G. (1993), *Approximate Inference in Generalized Linear Mixed Models*, „Journal of the American Statistical Association", 88(421), s. 9-25.

Dempster A.P., Laird N.M., Rubin D.B. (1977), *Maximum Likelihood from Incomplete Data Via the EM Algorithm*, „Journal of the Royal Statistical Society. Series B (Methodological)", 39(1), s. 1-38.

Denuit M., Maréchal X., Pitrebois S., Walhin J.F. (2007), *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus–Malus Systems*, John Wiley&Sons, Chichester.

Jorgensen B. (1987), *Exponential Dispersion Models*, „Journal of the Royal Statistical Society. Series B (Methodological)", 49(2), s. 127-162.

Jørgensen B., Paes De Souza M.C. (1994), *Fitting Tweedie's Compound Poisson Model to Insurance Claims Data*, „Scandinavian Actuarial Journal", 1994(1), s. 69-93.

Karlis D. (2001), *A General EM Approach for Maximum Likelihood Estimation in Mixed Poisson Regression Models*, „Statistical Modelling", 1(4), s. 305-318.

Lee Y., Nelder J.A. (1996), *Hierarchical Generalized Linear Models*, „Journal of the Royal Statistical Society. Series B (Methodological)", 58(4), s. 619-678.

McCullagh P., Nelder J.A. (1989), *Generalized Linear Models*, Chapman and Hall, London.

Peters G.W., Shevchenko P.V., Wüthrich M.V. (2009), *Model Uncertainty in Claims Reserving within Tweedie's Compound Poisson Models*, „Astin Bulletin", 39(1), s. 1-33.

R Core Team (2014), **R**, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R–project.org/

Smyth G.K. (2002), *Fitting Tweedie's Compound Poisson Model To Insurance Claims Data: Dispersion Modeling*, „Astin Bulletin", 32(1), s. 143-157.

Winkelmann R. (2003), *Econometric Analysis of Count Data*, Springer, New York.

Wolny-Dominiak A. (2013), *Modeling Claim Severity via h–likelihood* [w:] H. Vojackova, red., *Proceedings of 31st International Conference Mathematical Methods in Economics 2013*, College of Polytechnics Jihlava, Jihlava.

Wolny-Dominiak A. (2014), *Taryfikacja w ubezpieczeniach majątkowych z wykorzystaniem modeli mieszanych*, Wydawnictwo UE, Katowice.

Wolny-Dominiak A., Trzęsiok M. (2014), *insuranceData: A Collection of Insurance Datasets Useful in Risk Classification in Non–life Insurance*, R package version 1.0.

Zhang Y. (2013), *Likelihood-based and Bayesian Methods for Tweedie Compound Poisson Linear Mixed Models*, „Statistics and Computing", 23(6), s. 743-757.

COMPOUND MIXED POISSON DISTRIBUTION AND ITS APPLICATION IN INSURANCE

Summary: The paper presents compound mixed Poisson distributions in the context of Tweedie distribution, since e.g. the compound Poisson-gamma distribution is its special case. We use the fact, that Tweedie distribution belongs to overdispersed exponential family of distributions. In the first part of the paper the problem of overdispersion is addressed by considering mixed Poisson distribution, where the unobserved variable is introduced. Then we specify compound mixed Poisson distribution. In the second part we present the possibilities of making use of this distribution in estimating the total claim costs for the individual risk in the automobile insurance portfolio.

Keywords: mixed Poisson, compound mixed Poisson, non–life insurance, ratemaking, estimation.