PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1088 — 2005
Inwestycje finansowe i ubezpieczenia – tendencje światowe a polski rynek

Mirosław Szreder, Agnieszka Pobłocka
Uniwersytet Gdański

BAYESOWSKA AKTUALIZACJA ROZKŁADU LICZBY ODSZKODOWAŃ W UBEZPIECZENIACH KOMUNIKACYJNYCH

1. Wstęp

Skuteczne podejmowanie decyzji to m.in. umiejętność właściwego przetwarzania nowej informacji charakteryzującej sytuację decyzyjną. Twierdzenie Bayesa wskazuje na sposób aktualizacji bieżącej wiedzy na podstawie nowo pozyskanych informacji. W niniejszym opracowaniu przedstawimy możliwości wykorzystania podejścia bayesowskiego do aktualizacji rozkładu liczby odszkodowań na podstawie nowych, dostępnych informacji.

2. Twierdzenie Bayesa i wnioskowanie bayesowskie

Zwiększające się zasoby informacji z różnych dziedzin życia gospodarczego i społecznego, dostępne nie tylko naukowcom, działaczom gospodarczym, ale także zwykłym ludziom w ich codziennym życiu, każą zwrócić uwagę na problem przetwarzania informacji. Skuteczne poruszanie się w świecie zdarzeń stochastycznych uwarunkowane jest bowiem umiejętnością odpowiedniego przetwarzania dostępnych informacji, gotowością uczenia się na doświadczeniach z przeszłości i umiejętnością zmiany przekonań wynikających z nowych, dostępnych informacji. Jednym z najlepiej opisanych w literaturze i najczęściej stosowanych modeli „uczenia się" (warunkowania) na podstawie nowych informacji jest model oparty na twierdzeniu Bayesa (twierdzenie 1). Uniwersalność tego modelu polega na takim połączeniu dwóch źródeł informacji – wstępnej (*a priori*) oraz próbkowej (z doświadczenia) – w którym wagi obu tych rodzajów informacji zależą od stopnia prawdopodobieństwa (najczęściej subiektywnego) przypisanego im przez badacza. Podejście bayesowskie daje możliwość uwzględnienia w procesie przetwarzania informacji nie tylko różnej natury wiedzy apriorycznej (pochodzącej z doświadczenia, z poprzednich lub analogicznych badań, opartej na intuicji), ale także wyrażenia przez badacza jego stopnia niepewności co do poszczególnych elementów tej wiedzy. Wiedzę wstępną w modelu bayesowskim reprezentuje rozkład *a priori* $\pi(\cdot)$.

Twierdzenie 1. Twierdzenie Bayesa

Niech $\underline{X}$ i $\underline{Y}$ będą dowolnymi zmiennymi losowymi zadanymi na przestrzeni mierzalnej $(\Omega, \Im, \Pr(.))$ o wartościach w przestrzeni IR^{d_1}, $d_1 \geq 1$ i IR^{d_2}, $d_2 \geq 1$ odpowiednio. Niech $A \in B_{IR^{d_1}}$, $B \in B_{IR^{d_2}}$, $\Pr(\underline{X} \in B) > 0$. Zachodzi wtedy równość:

$$\Pr(\underline{Y} \in A \mid \underline{X} \in B) = \frac{\Pr(\underline{X} \in B \mid \underline{Y} \in A)\Pr(\underline{Y} \in A)}{\Pr(\underline{X} \in B)}. \tag{1}$$

Jeżeli istnieje gęstość łączna zmiennych $\underline{X}$ i $\underline{Y}$ względem miary Lesbesgue'a, to przy tradycyjnych oznaczeniach wzór Bayesa można zapisać następująco:

$$f_{(\underline{Y}|\underline{X})}(\underline{y}|\underline{x}) = \frac{f_{(\underline{X},\underline{Y})}(\underline{x},\underline{y})}{f_{\underline{X}}(\underline{x})} = \frac{f_{(\underline{X}|\underline{Y})}(\underline{x}|\underline{y}) \cdot f_{(\underline{Y})}(\underline{y})}{f_{\underline{X}}(\underline{x})}, \text{dla } \underline{x} \in IR^{d_1}, \underline{y} \in IR^{d_2} \text{ i } f_{\underline{X}}(\underline{x}) > 0. \tag{2}$$

Twierdzenie 1 stanowi model matematyczny charakteryzujący sposób uzyskiwania rozkładu *a posteriori* na podstawie rozkładu *a priori* $\pi(\cdot)$ oraz funkcji wiarygodności próby $L(\cdot, \underline{x})$[1]. Rozkład a *posteriori* $\pi^*(\cdot|\underline{x})$ uzyskany w wyniku zastosowania twierdzenia Bayesa przedstawia w sposób sformalizowany, wolny od aproksymacji, bieżący stan wiedzy badacza o nieznanych, szacowanych wielkościach, charakteryzujących określony problem decyzyjny. Każda nowa informacja próbkowa umożliwia aktualizację wiedzy badacza przez ponowne zastosowanie twierdzenia Bayesa, w którym uzyskany poprzednio rozkład *a posteriori* staje się rozkładem *a priori*, wyrażającym stan wiedzy przed uzyskaniem nowych informacji.

Niech wektor $\underline{x} \in IR^n$ oznacza możliwy przebieg (realizację) zjawiska empirycznego. Zakładamy, że pomiary $\underline{X} = (X_1, X_2 ... X_n)$, $\underline{X} \in IR^n$ stanowią próbę losową prostą z rozkładu o gęstości $f_{(\underline{X}|\underline{\Theta})}(\cdot|\underline{\theta})$ względem miary Lebesgue'a, która jest elementem rodziny rozkładów prawdopodobieństwa indeksowanych przez wektor $\underline{\theta} \in IR^n$:

$$P = \left\{ f_{(\underline{X}|\underline{\Theta})}(\cdot \mid \underline{\theta}) : \ \underline{\theta} \in IR_+ \cup \{0\} \right\}. \tag{3}$$

Na gruncie niebayesowskim rodzina P definiuje parametryczny model statystyczny $f_{(\underline{X}|\underline{\Theta})}(.|\underline{\theta})$, tzw. model próbkowy (ang. *sampling model*), w ramach którego

[1] Funkcja wiarygodności $L(\cdot, \underline{x}) = f_{(\underline{X}|\underline{\Theta})}(\underline{x} \mid \cdot) : IR^n \to IR_+$ jest funkcją dla $\underline{\Theta} \in E = IR^n$ i zaobserwowanego wektora obserwacji $\underline{X} \in X = IR^n$ definiowaną wzorem $L(\underline{\theta}, \underline{x}) = f_{(\underline{X}|\underline{\Theta})}(\underline{x}|\underline{\theta})$.

rozważa się np. zagadnienia estymacji wektora $\underline{\theta}$ i jego funkcji. Cechą charakterystyczną tego modelu jest wybór tylko jednej wartości parametru $\underline{\theta}$, a tym samym rozkładu prawdopodobieństwa rządzącego zjawiskiem badawczym. Często podejście niebayesowskie nazywane jest podejściem klasycznym lub częstościowym.

Na gruncie bayesowskim (subiektywnym) rodzina *P* nie jest jeszcze kompletnym modelem statystycznym. Oprócz modelu próbkowego $f_{(\underline{X}|\underline{\Theta})}(.|\underline{\theta})$, wprowadza się miarę probabilistyczną (lub przynajmniej miarę σ-skończoną) na zbiorze *E* o funkcji gęstości $f_{\underline{\Theta}}(.)$: $\underline{\Theta} \in E = IR^n$, czyli tzw. rozkład *a priori*, modelującą wstępne przekonania badacza o wektorze parametrów $\underline{\Theta}$. W dalszej części pracy rozkład *a priori* będziemy oznaczać przez $\pi(\cdot)$: $\pi(\theta) = f_{\underline{\Theta}}(\underline{\theta})$.

Bayesowski model statystyczny jest jednoznacznie określony przez funkcję łącznej gęstości dla wektora (potencjalnych) obserwacji $\underline{x}$ i wektora parametrów $\underline{\theta}$:

$$f_{(\underline{\Theta},\underline{X})}(\underline{x},\underline{\theta}) = f_{(\underline{X}|\underline{\Theta})}(\underline{x}|\underline{\theta}) \cdot \pi(\theta), \tag{4}$$

gdzie $f_{(\underline{X}|\underline{\Theta})}(\cdot|\underline{\theta})$ to gęstość rozkładu próbkowego, a $\pi(\cdot)$ to gęstość rozkładu *a priori*.

Wnioskowania o wektorze parametrów $\underline{\Theta}$ dokonuje się na podstawie rozkładu warunkowego tego wektora parametrów przy ustalonych obserwacjach, czyli na podstawie rozkładu *a posteriori* o gęstości wynikającej ze wzoru Bayesa:

$$f_{(\underline{\Theta}|\underline{X})}(\underline{\theta}|\underline{x}) = \frac{f_{(\underline{X},\underline{\Theta})}(\underline{x},\underline{\theta})}{f_{\underline{X}}(\underline{x})} = \frac{f_{(\underline{X}|\underline{\Theta})}(\underline{x}|\underline{\theta}) \cdot \pi(\underline{\theta})}{f_{\underline{X}}(\underline{x})}, \tag{5}$$

gdzie $f_{\underline{X}}(\underline{x}) = \int_E f_{(\underline{X}|\underline{\Theta})}(\underline{x}|\underline{\theta})\pi(\theta)d\underline{\theta}$ jest brzegową gęstością wektora obserwacji.

W dalszej części pracy, w celu uproszczenia oznaczeń, rozkład *a posteriori* oznaczać będziemy przez $\pi(\cdot|\underline{x})$: $\pi(\theta|\underline{x}) = f_{(\underline{\Theta}|\underline{X})}(\underline{\theta}|\underline{x})$.

Brak rozkładu *a posteriori* przy niewłaściwym rozkładzie *a priori* może być poważnym i trudnym problemem w zastosowaniu analizy bayesowskiej, dlatego w dalszych rozważaniach zakładamy specyfikację modelu bayesowskiego taką, że rozkład *a posteriori* istnieje.

Po wykorzystaniu kategorii funkcji wiarygodności bayesowski model statystyczny można zapisać jako $f_{(\underline{\Theta},\underline{X})}(\underline{x},\underline{\theta}) = L(\underline{\theta},\underline{x}) \cdot \pi(\theta)$, a gęstość rozkładu *a posteriori* jako $\pi^*(\theta|\underline{x}) \propto L(\underline{\theta},\underline{x}) \cdot \pi(\theta)$, gdyż $f_{\underline{X}}(\underline{x})$ nie zależy od parametru $\underline{\theta}$, a $\propto$ jest symbolem proporcjonalności.

Rozkłady *a priori* i rozkłady warunkowe, dla których można analitycznie wyznaczyć rozkład *a posteriori*, to tzw. **sprzężone rodziny rozkładów *a priori*** [2] (por. [DeGroot 1981]). Ogólnie, gęstość rozkładu brzegowego $f_{\underline{X}}(.)$ wektora zmiennych losowych $\underline{X}$ zadanego na zbiorze X oraz gęstość rozkładu warunkowego *a posteriori* $\pi(\cdot \mid \underline{x})$ są wyznaczane numerycznie za pomocą odpowiednich procedur statystycznych (por. [Osiewalski 2001; Szreder 1994]). Obecnie szybko rozwijające się techniki komputerowe, pozwalające symulacyjnie określać rozkłady *a posteriori*, stwarzają nowe szanse dla aplikacji metod bayesowskich.

Załóżmy, że wektor losowy $\underline{Y} = (Y_1, Y_2 ... Y_n)$, $\underline{Y} \in \mathrm{X} = IR^n$ będzie wektorem dodatkowych obserwacji, którego składowe są niezależne o jednakowym rozkładzie (próba losowa prosta) i gęstości prawdopodobieństwa zadanej wzorem $g_{(\underline{Y}|\underline{\Theta})}(\cdot \mid \underline{\theta})$: $\underline{\theta} \in E = IR^n$. Zauważmy, że wektor parametrów $\underline{\theta}$ jest taki sam, zatem rozkład warunkowy dodatkowych obserwacji, pod warunkiem dotychczasowych obserwacji, będzie miał postać:

$$f_{(\underline{Y}|\underline{X})}(\underline{y}|\underline{x}) = \frac{f_{(\underline{Y},\underline{X})}(\underline{y},\underline{x})}{f_{\underline{X}}(\underline{x})} = \frac{\int_E f_{(\underline{Y}|(\underline{X},\underline{\Theta}))}(\underline{y}|(\underline{x},\underline{\theta})) \cdot L(\underline{\theta},\underline{x}) \cdot \pi(\underline{\theta}) \quad d\underline{\theta}}{\int_{(X,E)} g_{(\underline{Y},\underline{X},\underline{\Theta})}(\underline{y},\underline{x},\underline{\theta}) d\underline{y} \quad d\underline{\theta}}. \tag{6}$$

W związku z tym rozkład *a posteriori* dodatkowych informacji przbiera postać:

$$\pi^*(\underline{y}|\underline{x}) = \int_E g_{(\underline{Y}|\underline{\Theta})}(\underline{y}|\underline{\theta}) \cdot \pi^*(\underline{\theta}|\underline{y}) \quad d\underline{\theta}. \tag{7}$$

Predykcji rozkładu *a posteriori* możemy dokonać analogicznie do rozkładu *a posteriori* dla dodatkowej cechy. Załóżmy, że istnieje gęstość rozkładu próbkowego $f_{(\underline{X}|\underline{\Theta})}(\cdot \mid \underline{\theta}) : IR^n \to IR_+ \cup \{0\}$, $\underline{\theta} \in \mathrm{E} = IR^n$ $\underline{X} \in X = IR^n$ oraz wektor losowy $\underline{Y} = (Y_1, Y_2, ..., Y_n)$, $\underline{Y} \in \mathrm{X} = IR^n$ będzie wektorem przyszłych obserwacji, którego składowe są niezależne, o identycznym rozkładzie (próba losowa prosta), o gęstości prawdopodobieństwa zadanej wzorem $g_{(\underline{Y}|\underline{\Theta})}(\cdot \mid \underline{\theta})$: $\underline{\theta} \in E = IR^n$. Zauważmy, że parametr $\underline{\theta}$ nie zmienia się, zatem rozkład warunkowy przyszłych obserwacji, pod warunkiem dotychczasowych obserwacji, czyli predyktywny rozkład *a posteriori* będzie miał postać taką jak (7).

---

[2] Niech ***Q*** będzie rodziną rozkładów *a priori* zadaną poprzez gęstości $\pi(\cdot)$. Niech ***P*** będzie rodziną rozkładów warunkowych zadaną przez gęstości $f_{(X|\Theta)}(\cdot \mid \theta)$. **Rodzina *Q* jest sprzężona do rodziny *P***, jeżeli dla każdej gęstości $f_{(X|\Theta)}(\cdot \mid \theta)$ z rodziny ***P*** i dla każdej gęstości $\pi(.)$ z rodziny ***Q***, gęstość rozkładu *a posteriori*, wyznaczona zależnością $\pi^*(\theta \mid \underline{x}) \overset{\theta}{\propto} f_{(X|\Theta)}(\underline{x} \mid \theta) \cdot \pi(\theta)$, jest gęstością z rozkładu z rodziny ***Q***.

3. Probabilistyczne modele liczby odszkodowań

Do najczęściej stosowanych rozkładów liczby szkód w portfelach jednorodnych polis, w określonym przedziale czasu, należą:

1. Rozkład dwumianowy

W sytuacjach, gdy znana jest górna granica liczby szkód zaistniałych w okresie objętym ochroną ubezpieczeniową (najczęściej 1 rok), zakłada się, że liczba szkód zaistniałych ma rozkład dwumianowy. Funkcja prawdopodobieństwa rozkładu dwumianowego z parametrami n oraz p przybiera postać:

$$\Pr(N=k)=\binom{n}{k}p^k\, q^{n-k},\quad \text{gdzie } k\in IN,\ \ 0<p<1,\ \ q=1-p. \tag{8}$$

Rozkład dwumianowy jest także wykorzystywany w indywidualnych i kolektywnych modelach ryzyka do opisu łącznej wartości odszkodowań w danym okresie sprawozdawczym. Na przykład, jeżeli liczba szkód N w pewnym portfelu jednorodnego ryzyka (w ciągu roku) ma rozkład dwumianowy, a wartość pojedynczej szkody jest zmienną losową X określoną na zbiorze $\mathrm{IR}_+\cup\{0\}$, to łączna wartość szkód $\sum_{i=1}^{N}X_i$ (w ciągu roku) w badanym portfelu ryzyka ma rozkład dwumianowy.

2. Rozkład ujemny dwumianowy

W sytuacjach, gdy występuje bardzo zróżnicowane ryzyko, można przyjąć, że liczba szkód N w portfelu polis ma rozkład ujemny dwumianowy z parametrami r oraz p. Funkcja prawdopodobieństwa rozkładu ujemnego dwumianowego przybiera postać:

$$\begin{gathered}\Pr(N=n)=\binom{r+n-1}{n}p^r\, q^n,\quad \text{gdzie}\ \ n\in IN,\ \ r>0,\ \ 0<p<1,\ \ q=1-p,\\ \binom{r+n-1}{n}=\frac{(r+n-1)(r+n-2)\dots r}{n!},\end{gathered} \tag{9}$$

gdzie p jest prawdopodobieństwem sukcesu (prawdopodobieństwem wystąpienia szkody), a n to liczba zdarzeń niezbędna do uzyskania r sukcesów (szkód).

Rozkład ujemny dwumianowy jest stosowany, gdy występuje bardzo zróżnicowane ryzyko. Na przykład, jeżeli liczba szkód ma rozkład Poissona ze średnią Λ i heterogeniczność ryzyka powodują, że rozkład zmiennej losowej Λ jest rozkładem gamma, to zmienna losowa N ma rozkład ujemny dwumianowy.

3. Rozkład Poissona

Często zakłada się (ze względu ma własności), że rozkład całkowitej liczby szkód zaistniałych w badanym okresie objętym ochroną ubezpieczeniową jest rozkładem Poissona z parametrem λ. Funkcja prawdopodobieństwa rozkładu Poissona ma postać:

$$\Pr(N=n)=\frac{\lambda^n}{n!}e^{-\lambda},\quad \text{gdzie}\ \ n\in IN,\ \lambda>0. \tag{10}$$

Rozkład Poissona jest często stosowany, gdyż ma bardzo ważną własność: zmienna losowa, będąca sumą niezależnych statystycznie zmiennych losowych, każdej o rozkładzie Poissona ze średnią odpowiednio λ_i, $i = 1, 2, ..., n$, ma rozkład Poissona ze średnią będącą sumą wszystkich średnich, tj. $\lambda = \sum_{i=1}^{n} \lambda_i$. Własność ta pozwala szybko wyznaczać wartość całkowitej szkody w badanym portfelu jednorodnych typów ryzyka.

4. Mieszany rozkład Poissona

Całą rodzinę rozkładów liczby szkód można otrzymać, przyjmując, że parametr λ w rozkładzie Poissona jest zmienną losową o gęstości $u(.)$ dla $\lambda > 0$ oraz że rozkład warunkowy zmiennej losowej N, gdy $\Lambda = \lambda$ jest rozkładem Poissona z parametrem λ. Po wykorzystaniu wzoru na prawdopodobieństwo całkowite, mamy:

$$\Pr(N = n) = \int_0^\infty \Pr(N \mid \Lambda = \lambda) u(\lambda) \quad d\lambda = \int_0^\infty \frac{\lambda^n}{n!} e^{-\lambda} \; u(\lambda) \quad d\lambda. \qquad (11)$$

Mieszany rozkład Poissona znajduje zastosowanie w kolektywnych modelach ryzyka, gdy uwzględniamy franszyzę redukcyjną[3].

4. Empiryczny przykład bayesowskiej analizy średniej liczby wypłaconych odszkodowań

Rozpatrzmy portfel jednorodnych typów ryzyka komunikacyjnego autocasco (AC). W ubezpieczeniu komunikacyjnym AC, zgodnie z ogólnymi warunkami ubezpieczenia[4], przedmiotem ubezpieczenia są pojazdy wraz z wyposażeniem określonym we wniosku. Ochroną ubezpieczeniową są objęte szkody powstałe w pojeździe bądź w jego wyposażeniu w związku z ruchem i postojem pojazdu. Do zakresu ubezpieczenia mogą być także włączone szkody powstałe wskutek zdarzeń losowych (powodzi, pożaru, uderzenia pioruna) i działania osób trzecich (włamania, kradzieży).

W ubezpieczeniu autocasco górną granicę odpowiedzialności ubezpieczyciela wyznacza suma ubezpieczenia odpowiadająca rzeczywistej wartości pojazdu. Za odpowiednio zmniejszoną składkę umowa ubezpieczenia AC może zostać zawarta z zastosowaniem udziału własnego w każdej szkodzie. Zakres ubezpieczenia może być rozszerzony o umowy dodatkowe za odpowiednio zwiększoną składkę. Wypłata odszkodowania za określoną szkodę pomniejsza sumę ubezpieczenia w wysokości odpowiadającej tej wypłacie. Zbadamy liczbę wypłaconych odszkodowań przypadającą na 1000 jednorodnych polis AC w jednym okresie sprawozdawczym (równym rok)[5]. Przyjmujemy, że z jednej

---

[3] Szerzej na temat rodzajów franszyzy redukcyjnej por. m.in. [Monkiewicz, Gąsiorkiewicz, Hadyniak 1999, s. 52-54]; [Ostasiewicz 2000, rozdz. 3].

[4] Ogólne warunki ubezpieczenia autocasco z dnia 25 kwietnia 2003 r. ze zmianami zatwierdzonymi Uchwałą nr UZ/432/2003 z dnia 7 października 2003 r.

[5] Dane źródłowe są danymi rzeczywistymi, pochodzącymi z jednego z oddziałów PZU SA.

szkody wypłacone zostaje co najwyżej jedno odszkodowanie oraz pojedynczy klient jest tylko osobą fizyczną (tzn. nie jest osobą prawną).

Niech zmienna losowa N oznacza liczbę wypłaconych odszkodowań. Z badań wynika, że N jest zmienną losową o rozkładzie Poissona z parametrem Θ, ($\Theta \in E = IR_+$), który interpretujemy jako **średnią liczbę wypłaconych odszkodowań**. Przyjmujemy, że rozkład liczby wypadków jest zależny od średniej liczby wypłaconych odszkodowań, tj. zmiennej losowej Θ. Z wcześniejszych analiz wynika, że rozkład średniej liczby wypłaconych odszkodowań (rozkład *a priori*) w ostatnich latach jest zgodny z rozkładem gamma z parametrami $a = 2350$ i $\beta = 6$. Zakładamy, że liczby wypłaconych odszkodowań w kolejnych latach występują niezależnie i podlegają rozkładowi Poissona z tym samym parametrem $\Theta = \theta$. Dokonujemy obserwacji w zakresie produktu AC przez 10 lat. Przyjmujemy, że w każdym roku wypłacono odpowiednio $n_1, n_2, \ldots, n_{10}$ odszkodowań i w sumie wypłacono 4321 odszkodowań. Wyznaczymy rozkład *a posteriori* średniej liczby wypłaconych odszkodowań oraz dokonamy predykcji otrzymanego rozkładu *a posteriori* na kolejny, przyszły, jedenasty rok badań.

Budując model probabilistyczny, przyjmujemy, że $X = IN$, $E = IR_+$, $B_X = 2^{IN}$, $B_E = B_{IR+}$, a przestrzeń produktowa zmiennych losowych N i Θ będzie postaci $(IN \times IR_+, 2^{IN} \otimes B_{IR+})$. Gdy skorzysta się z (3), częstość próbkowa liczby szkód N pod warunkiem parametru $\Theta = \theta$, $\theta \in IR_+$ będzie miała postać:

$$f_{(N|\Theta)}(n|\Theta = \theta) = \Pr(N = n|\Theta = \theta) = \frac{\theta^n}{n!}e^{-\theta}, \quad n \in IN,\ \theta \in IR_+.$$

Jeżeli składowe wektora $\underline{n} = (n_1, n_2, \ldots, n_{10}) \in IN^{10}$ oznaczają liczby wypłaconych odszkodowań z $t = 10$ lat (próba losowa prosta w sensie modelu Bayesa), to funkcja wiarygodności (dla $\theta \in IR_+$, $\underline{n} \in IN^{10}$) będzie miała postać:

$$L(\theta, \underline{n}) = f_{(\underline{N}|\Theta)}(\underline{n}|\Theta = \theta) = \prod_{i=1}^{10}\frac{\theta^{n_i}}{n_i!}e^{-\theta} = \frac{e^{-10\theta}\theta^{\sum_{i=1}^{10}n_i}}{\prod_{i=1}^{10}n_i!}.$$

Z założeń wynika, że rozkład *a priori* parametru Θ jest rozkładem gamma z parametrami $a = 2350$ i $\beta = 6$ o gęstości postaci[6]:

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \overset{\theta}{\propto} \theta^{2350-1}e^{-6\theta}, \qquad \theta \in IR_+,$$

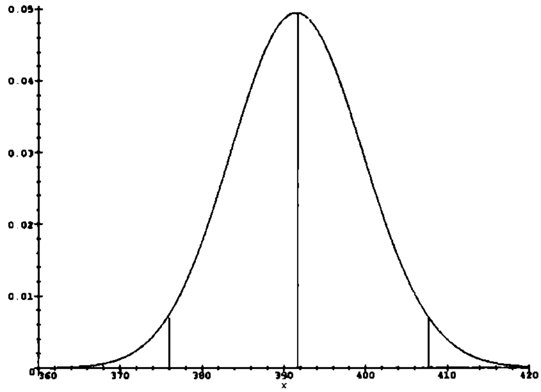
stąd, podstawowe charakterystyki rozkładu *a priori* parametru Θ są równe:

$$IE(\Theta) = \frac{\alpha}{\beta} = \frac{2350}{6} \cong 391{,}67, \qquad Var(\Theta) = \frac{\alpha}{\beta^2} = \frac{2350}{6^2} \cong 65{,}28.$$

Wartość oczekiwana zmiennej losowej Θ, opisującej średnią liczbę wypłaconych odszkodowań przypadających na 1000 jednorodnych polis AC w ciągu roku, jest równa 391,67, a odchylenie standardowe $\sqrt{65{,}28} \cong 8{,}08$. Rozkład *a priori*

[6] Wyrażenie $\overset{\theta}{\propto}$ oznacza proporcjonalność względem parametru θ.

średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC przedstawiono na rys. 1.



Rys. 1. Rozkład *a priori* średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC
Źródło: opracowanie własne.

Rozkład *a posteriori*

Po skorzystaniu z twierdzenia 1 oraz wzoru (6) gęstość rozkładu warunkowego *a posteriori* dla przyjętych parametrów ($a = 2350, \beta = 6, t = 10, \sum_{i=1}^{10} n_i = 4321$) ma postać:

$$\pi(\theta \mid \underline{n}) \overset{\theta}{\propto} \theta^{\left(\alpha+\sum_{i=1}^{10} n_i - 1\right)} e^{-(10+\beta)\theta} = \theta^{(6671-1)} e^{-16\theta}.$$

Otrzymany rozkład jest rozkładem gamma z parametrami odpowiednio 6671 i 16, którego podstawowe charakterystyki są równe:

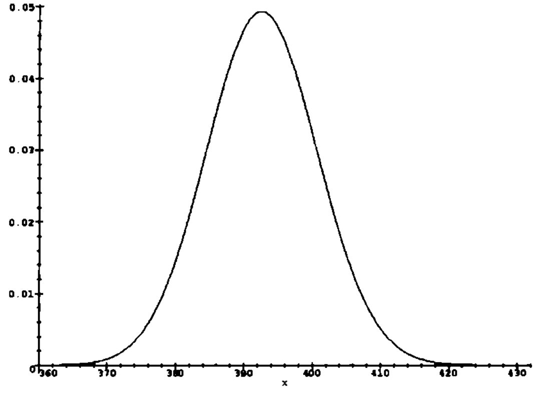
$$IE(\Theta \mid \underline{\mathrm{N}} = \underline{\mathrm{n}}) = \frac{\alpha + \sum_{i=1}^{10} n_i - 1}{10+\beta} = \frac{6671}{16} \cong 416{,}94,$$

$$Var(\Theta \mid \underline{\mathrm{N}} = \underline{\mathrm{n}}) = \frac{\alpha + \sum_{i=1}^{10} n_i - 1}{(10+\beta)^2} = \frac{6671}{16^2} \cong 26{,}06.$$

Na podstawie tych wyników stwierdzamy, że w badanym okresie 10 lat wartość oczekiwana średniej liczby wypłaconych odszkodowań jest równa 416,94, a odchylenie standardowe $\sqrt{26{,}06} = 5{,}11$.

Wartość oczekiwana średniej liczby wypłaconych odszkodowań w jednym okresie sprawozdawczym jest wielkością niezwykle ważną dla ubezpieczyciela, gdyż na jej podstawie wyznacza się różne wskaźniki ekonomiczne, mówiące o kondycji finansowej towarzystw ubezpieczeniowych, np. wskaźnik retencji czy też wielkość rezerwy (IBNR).

Zauważmy, że wartość oczekiwana rozkładu *a priori* wynosi 391,67 i odchylenie standardowe – 8,25. Wartość oczekiwana rozkładu *a posteriori* wynosi 416,94 i odchylenie standardowe 5,11. Wnioskujemy zatem, że dodatkowa informacja ze zgłoszonych roszczeń i wypłaconych odszkodowań w ciągu 10 lat spowodowała zwiększenie precyzji oszacowań nieznanej średniej liczby wypłaconych odszkodowań z $1/8{,}08 \cong 0{,}13$ do $1/5{,}11 \cong 0{,}2$. Na rysunku 2 graficznie przedstawiono rozkład *a posteriori* średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC.



Rys. 2. Rozkład *a posteriori* średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC
Źródło: opracowanie własne.

Rozkład brzegowy średniej liczby wypłaconych odszkodowań

Wyznaczymy *rozkład brzegowy liczby wypłaconych odszkodowań.* Gęstość zmiennej losowej N będzie wyrażona wzorem:

$$f_N(n) = P(N=n) = \int_0^\infty f_{(N|\Theta)}(n\,|\,\Theta=\theta)\pi(\theta)d\theta = \int_0^\infty \frac{\theta^n}{n!}e^{-\theta}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta$$

$$= \frac{\beta^\alpha}{n!\Gamma(\alpha)}\int_0^\infty \theta^{n+\alpha-1}e^{-(\beta+1)\theta}d\theta = \frac{\beta^\alpha}{n!\Gamma(\alpha)}\int_0^\infty \frac{(\beta+1)^{n+\alpha}}{\Gamma(n+\alpha)}\theta^{n+\alpha-1}e^{-(\beta+1)\theta}d\theta\frac{\Gamma(n+\alpha)}{(\beta+1)^{n+\alpha}}$$

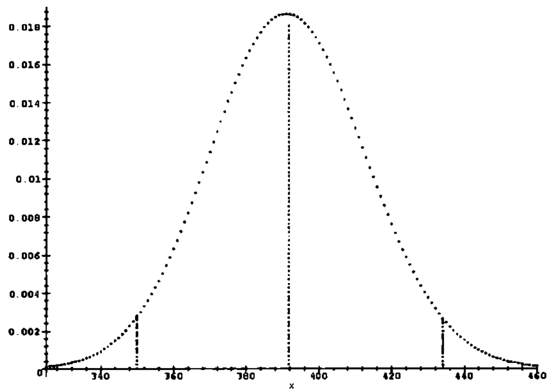
$$= \frac{\beta^\alpha}{n!\Gamma(\alpha)}.1.\frac{\Gamma(n+\alpha)}{(\beta+1)^{n+\alpha}} = \frac{(n+\alpha-1)(n+\alpha-2)...(\alpha+1)\alpha}{n!}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^n$$

$$= \binom{n+\alpha-1}{n}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^n, \text{ gdzie: } n \in IN, \theta \in IR_+.$$

Otrzymany rozkład jest rozkładem ujemnym dwumianowym z parametrami $r = \alpha$ oraz $p = 1/(\beta + 1)$, który w praktyce jest wykorzystywany do globalnej oceny portfela jednorodnych typów ryzyka. Podstawowe charakterystyki w rozkładzie ujemnym dwumianowym z parametrami $r = 2350$ i $p = 1/7$, są równe:

$$IE(N) = \frac{rp}{1-p} = \frac{2350 \cdot \frac{1}{7}}{\frac{6}{7}} \cong 391{,}66, \quad Var(N) = \frac{rp}{(1-p)^2} = \frac{2350 \cdot \frac{1}{7}}{\left(\frac{6}{7}\right)^2} \cong 456{,}00.$$

Wartość oczekiwana liczby odszkodowań jest równa 391.66, a odchylenie standardowe wynosi $\sqrt{456} \cong 21{,}35$. Na rysunku 3 przedstawiono brzegowy rozkład liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC.



Rys. 3. Brzegowy rozkład liczby wypłaconych odszkodowań przypadających na rok na 1000 jednorodnych polis AC
Źródło: opracowanie własne.

Predykcja rozkładu *a posteriori*

Na podstawie historycznych i bieżących informacji firmy ubezpieczeniowe regularnie sporządzają prognozy średniej liczby zgłoszonych roszczeń i wypłaconych odszkodowań na kolejne okresy sprawozdawcze. Prognozy przygotowywane są na potrzeby sprawozdawczości wewnętrznej (np. zarządu firmy) i zewnętrznej (np. KNUiFE). W tym celu wyznacza się predyktywny warunkowy rozkład *a posteriori* średniej liczby wypłaconych odszkodowań, przy dotychczasowej, bieżącej informacji.

Niech wektor $\underline{N}$ będzie próbą losową prostą opisującą liczbę wypłaconych odszkodowań w kolejnych 10. latach ($t = 10$) w danym portfelu jednorodnych typów ryzyka komunikacyjnego autocasco. Z badań wynika, że składowe wektora $\underline{N}$ są zmiennymi losowymi o częstościach z rozkładu Poissona z parametrem Θ, ($\Theta \in E = IR_+$), który jest interpretowany jako średnia liczba wypłaconych odszkodowań. Niech zmienna losowa Y zadana na przestrzeni mierzalnej (IN, 2^{IN}) będzie liczbą wypłaconych odszkodowań w kolejnym, tj. 11. roku badania. Zakładamy, że Y i $\underline{N}$ są niezależne w sensie podejścia bayesowskiego. Przyjmujemy, że częstość zmiennej losowej Y, pod warunkiem Θ, jest częstością z rozkładu Poissona z parametrem $\Theta = \theta = 391,67$:

$$g_{(Y|\Theta)}(y \mid \Theta = \theta) = \frac{\theta^y}{y!}e^{-\theta}, n \in IN, \theta = 391,67.$$

Po skorzystaniu ze wzoru dla dodatkowej cechy w modelu bayesowskim, **predyktywny warunkowy rozkład a posteriori** zmiennej Y, pod warunkiem dotychczasowej liczby odszkodowań, będzie rozkładem o funkcji gęstości ($\underline{n} \in IN^{10}$, $y \in IN$) wyrażonej wzorem:

$$f_{(Y|\underline{N})}(y|\underline{n}) \overset{\theta}{\propto} \int_0^{\infty} g_{(Y|\Theta)}(y|\Theta=\theta)\pi^*(\theta|\underline{n})d\theta = \int_0^{\infty} \frac{\theta^y}{y!}e^{-\theta} \frac{\theta^{\left(\alpha+\sum_{i=1}^t n_i-1\right)} e^{-(t+\beta)\theta}(t+\beta)^{\left(\alpha+\sum_{i=1}^t n_i\right)}}{\Gamma\left(\alpha+\sum_{i=1}^t n_i\right)}d\theta =$$

$$= \int_0^{\infty} \frac{\theta^{\left(\alpha+y+\sum_{i=1}^t n_i-1\right)} e^{-(t+1+\beta)\theta}(t+\beta)^{\left(\alpha+\sum_{i=1}^t n_i\right)}}{y!\Gamma\left(\alpha+\sum_{i=1}^t n_i\right)}d\theta =$$

$$= \frac{(t+\beta)^{\left(\alpha+\sum_{i=1}^t n_i\right)} \cdot \Gamma\left(\alpha+y+\sum_{i=1}^t n_i\right)}{(t+1+\beta)^{\left(\alpha+y+\sum_{i=1}^t n_i\right)} y!\Gamma\left(\alpha+\sum_{i=1}^t n_i\right)} =$$

$$= \frac{\Gamma\left(\alpha+y+\sum_{i=1}^t n_i\right)}{y!\Gamma\left(\alpha+\sum_{i=1}^t n_i\right)} \left(\frac{t+\beta}{t+1+\beta}\right)^{\left(\alpha+\sum_{i=1}^t n_i\right)} \left(\frac{1}{t+\beta+1}\right)^y,$$

gdzie $\dfrac{\Gamma\left(\alpha+y+\sum_{i=1}^t n_i\right)}{y!\Gamma\left(\alpha+\sum_{i=1}^t n_i\right)} = \begin{pmatrix} \alpha+y+\sum_{i=1}^t n_i - 1 \\ \alpha+\sum_{i=1}^t n_i \end{pmatrix}$.

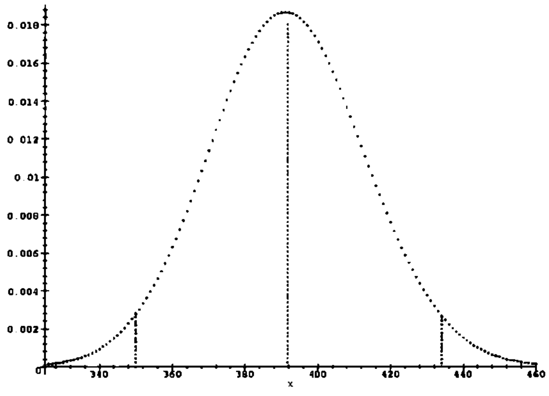
Otrzymany rozkład jest rozkładem ujemnym dwumianowym z parametrami $r = \alpha+y+\sum_{i=1}^t n_i - 1$ i $p = \dfrac{1}{t+\beta+1}$, po podstawieniu wartości liczbowych, zgodnych z założeniami, otrzymujemy:

$$f_{(Y|\underline{N})}(y \mid \underline{n}) \overset{\theta}{\propto} \frac{\Gamma(2350+y+4321)}{y!\Gamma(2350+4321)}\left(\frac{16}{17}\right)^{6671}\left(\frac{1}{17}\right)^y, \text{dla } y \in IN.$$

Podstawowe charakterystyki badanego predyktywnego rozkładu *a posteriori*, tj. rozkładu ujemnego dwumianowego z parametrami $r = 6671$ i $p = \dfrac{1}{t+\beta+1} = \dfrac{1}{17}$, są równe:

$$IE(Y \mid \underline{N}) = \frac{rp}{1-p} = \frac{6671 \cdot \frac{1}{17}}{\frac{16}{17}} \cong 416,94, \quad Var(Y \mid \underline{N}) = \frac{rp}{(1-p)^2} = \frac{6671 \cdot \frac{1}{17}}{\left(\frac{16}{17}\right)^2} \cong 442,99.$$

Na rysunku 4 zaprezentowano predyktywny rozkład *a posteriori* średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 polis AC.



Rys. 4. Predyktywny rozkład *a posteriori* średniej liczby wypłaconych odszkodowań przypadających na rok na 1000 polis AC
Źródło: opracowanie własne.

Jeżeli zakładamy, że wartość oczekiwana średniej liczby wypłaconych odszkodowań przypadająca na rok na 1000 jednorodnych polis ubezpieczeń komunikacyjnych autocasco w okresie od 1. do 10. roku badań wyniesie $IE(\Theta|\underline{N}) = 416{,}94$, a odchylenie standardowe 5,11, przy założeniu *ceteris paribus*, to oczekujemy, że w 11. roku badań wartość oczekiwana średniej liczby wypłaconych odszkodowań będzie równa $IE(Y|\underline{N}) = 416{,}94$, a odchylenie standardowe wyniesie $\sqrt{442{,}99} = 21{,}05$. Zauważmy, że precyzja rozkładu *a posteriori* wynosi $1/_{5.11} \cong 0{,}2$ i jest czterokrotnie większa od precyzji predyktywnego rozkładu równej $1/_{21.05} \cong 0{,}05$.

5. Zakończenie

Podejście bayesowskie jest jedną z niewielu teorii, pozwalającą wykorzystać wcześniejsze informacje w modelowaniu zjawisk bieżących. Zastosowanie teorii bayesowskiej w ubezpieczeniach jest niezwykle cenne ze względu na możliwość badania informacji zebranych na podstawie dłuższego czasu i modyfikacji danych na bazie aktualnych informacji.

Przedstawiony w artykule przykład obrazuje, jak w towarzystwach ubezpieczeniowych, na podstawie historycznych oraz bieżących informacji o średniej liczbie wypłaconych odszkodowań i informacji *a priori*, po skorzystaniu z podejścia bayesowskiego, można modelować charakterystyki liczbowe przyszłej średniej liczby wypłaconych odszkodowań. Oczywiście, przykład nie wyczerpuje wszystkich możliwości zastosowania podejścia bayesowskiego w aktuarialnej analizie ubezpieczeń, lecz przedstawia kilka możliwości. Więcej informacji na temat rozkładów prawdopodobieństwa wykorzystywanych w ubezpieczeniach majątkowych można znaleźć w literaturze przedmiotu (por. [Bawers 1986; Buhimann 1970; Ostasiewicz 2000; Szreder 1999]).

Literatura

Bawers N.L. (red.), *Actuarial Mathematics*, The Society of Actuaries 1986.

Buhimann H., *Mathematical Methods in Risk Theory*, Springer-Verlag, Berlin, Heidelberg, New York 1970.

DeGroot M., *Optymalne decyzje statystyczne*, PWN, Warszawa 1981.

Monkiewicz J., Gąsiorkiewicz L., Hadyniak B., *Zarządzanie finansami ubezpieczeń*, Poltext, Warszawa 1999.

Osiewalski J., *Ekonometria bayesowska w zastosowaniach*, AE, Kraków 2001.

Ostasiewicz W. (red.), *Modele aktuarialne*, AE, Wrocław 2000.

Szreder M., *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*, UG, Sopot 1994.

Szreder M., *Use of Prior Probabilities in Bayesian Inference*, „Statistics in Transition" 1999, vol. 4, nr 2, s. 245-258.

BAYESIAN UPDATING THE DISTRIBUTION OF CLAIMS IN AUTOMOBILE INSURANCE

Summary

An increasing amount of information collected in various areas of economic and social activities makes the problem of processing information more and more important. Efficient decision making requires the use of proper rules for processing and updating information used by decision makers. The Bayesian approach is considered to be one of the most appealing ways for updating the present knowledge on the basis of newly collected data. The paper presents opportunities of the Bayesian approach in updating the distribution of the number of claims when new information becomes available. An empirical example discussed in the paper, which uses actual data obtained from a Polish insurance company, exhibits some opportunities of Bayesian tools in actuarial problems.