

Magdalena Sipa, Renata Gnitecka

APROKSYMACJA FUNKCJI INTENSYWNOŚCI ZGONÓW

1. Wstęp

Przez wiele lat ludzie próbowali znaleźć rozkład trwania życia. Jednak badania pokazały, że życia nie da się opisać za pomocą rozkładu deterministycznego, tzn. że nie można powiedzieć jednoznacznie, jak długie będzie życie poszczególnych osób. Dlatego trwanie życia zaczęto opisywać zmienną losową. Tak powstała matematyka ubezpieczeń na życie.

Na podstawie danych empirycznych buduje się tzw. tablice trwania życia, które służą do analizy umieralności. Dane zbierane są w przedziałach wieku, najczęściej długości jednego roku lub kilku lat, dlatego funkcje szacowane na ich podstawie są dyskretne. Następnie wyznacza się ich ciągłe odpowiedniki.

Tradycyjnie firmy ubezpieczeniowe wypłacały klientom świadczenia z tytułu zajścia zdarzenia ubezpieczeniowego na koniec okresu rozliczeniowego, np. miesiąca, kwartału. Obecnie coraz częściej proponuje się wypłatę świadczenia w dowolnym momencie trwania ubezpieczenia, po przedstawieniu odpowiednich dokumentów. Operacja taka wymaga określenia analitycznej postaci funkcji intensywności wystąpienia zdarzenia ubezpieczeniowego. W niniejszym artykule zajmujemy się najczęściej ubezpieczanym zdarzeniem – śmiercią ubezpieczonego.

Celem pracy jest aproksymacja funkcji intensywności zgonów na podstawie danych zawartych w polskich tablicach trwania życia i na podstawie znanych z literatury praw wymieralności kohorty. Wiek osoby oznacza się przez x , dla $x = 0, 1, 2, \dots, w$. Jeżeli na początku rozważanego okresu osoba jest w wieku x , pod koniec w wieku $x + 1$ (o jeden rok starsza), to ogólnie rozważa się osobę w przedziale wieku $(x, x + 1)$. Ponadto ostatni przedział pozostawia się nieskończony $(w, +\infty)$.

W tablicach trwania życia zawarte są przede wszystkim dwie funkcje:

- liczba osób dożywających,
- liczba zmarłych.

Liczba dożywających l_x określa liczbę osób, które z kohorty początkowej o liczebności l_0 , dożyły wieku x . Liczba zmarłych, oznaczana d_x , określa liczbę osób zmarłych w przedziale wieku $(x, x+1)$. Za początkową liczebność badanej kohorty, l_0 , przyjmuję się liczbę okrągłą, np. 100 000, 1 000 000; w ten sposób łatwiej porównać trwania życia różnych kohort.

Związek pomiędzy liczbą dożywających a liczbą zmarłych jest następujący:

$$l_{x+1} = l_x - d_x \text{ dla } x = 0, 1, 2, \dots, w-1,$$

przy czym $l_w = d_w$ oraz $d_0 + d_1 + \dots + d_w = l_0$.

W tablicy trwania życia rozpatruje się również funkcję prawdopodobieństwa zgonu, oznaczaną q_x . Określa ona prawdopodobieństwo warunkowe zgonu w wieku $(x, x+1)$ pod warunkiem dożycia wieku x . Ponadto wyznacza się funkcję prawdopodobieństwa przeżycia, oznaczaną symbolem p_x . Funkcja ta określa prawdopodobieństwo, że osoba w wieku x dożyje wieku $x+1$ (przeżyje 1 rok) i jest równa:

$$p_x = 1 - q_x.$$

W rzeczywistości w tablicy umieszcza się tylko oceny, obliczone na podstawie zaobserwowanej liczby zgonów i liczebności kohorty w kolejnych latach. Oceną prawdopodobieństwa zgonu wyznaczoną metodą największej wiarygodności (por. [Bolesławski 1973]) jest:

$$\hat{q}_x = \frac{d_x}{l_x}.$$

Tabela 1 przedstawia fragment tablicy trwania życia.

Funkcja, której aproksymację chcemy przeprowadzić, to funkcja intensywności zgonów, oznaczana $\mu(x)$, czyli taka, że wartość $\mu(x)h + o(h)$ jest warunkowym prawdopodobieństwem zgonu w nieskończenie małym przedziale $(x, x+h)$.

Tabela 1. Fragment tablicy trwania życia mężczyzn 2002

Wiek	Liczba dożywających	Liczba zmarłych	Prawdopodobieństwo zgonu
x	l_x	d_x	q_x
0	100 000	813	0,00813
1	99 187	53	0,000534
2	99 133	38	0,000383
3	99 096	27	0,000272
4	99 069	22	0,000222
5	99 047	20	0,000202
6	99 027	19	0,000192
7	99 009	20	0,000202

Wyprowadzenie wzoru na funkcję intensywności zgonów można znaleźć w pracach L. Bolesławskiego [1973] i M. Cieślak [1992].

W przedziale $(x, x+1)$ prawdopodobieństwo warunkowe dożycia wieku $x+1$, pod warunkiem dożycia wieku x , jest równe:

$$p_x = \exp \left\{ - \int_x^{x+1} \mu(s) ds \right\}.$$

Wobec tego prawdopodobieństwo śmierci w tym przedziale wynosi:

$$q_x = 1 - \exp \left\{ - \int_x^{x+1} \mu(s) ds \right\}. \quad (1)$$

Zależność (1) i prawa wymieralności, opisane w następnym punkcie, pozwalają aproksymować funkcję intensywności zgonów $\mu(x)$.

2. Rozkłady intensywności zgonów i prawdopodobieństwa zgonów

Na podstawie danych dostępnych w tablicach trwania życia nie można od razu wyznaczyć intensywności zgonów. W literaturze [Ostasiewicz 2000] podaje się pewne rozkłady zwane prawami wymierania, niestety żadne z nich nie opisują dokładnie rozkładu trwania życia. Są to m.in.:

1. Prawo Gompertza

$$\mu(t) = Bc^t, \quad B, t > 0, \quad c > 1.$$

2. Prawo Gompertza–Makehama

$$\mu(t) = A + Bc^t, \quad B, t > 0, \quad A \geq -B, \quad c > 1.$$

3. Prawo Weibulla

$$\mu(t) = Bt^n, \quad B, n, t > 0.$$

Zauważmy, że prawo Gompertza–Makehama jest uogólnieniem prawa Gompertza.

W niniejszej pracy aproksymacja funkcji intensywności $\hat{\mu}(t)$ sprowadza się do wyznaczenia parametrów A , B i c . W tym celu, na podstawie wymienionych praw i zależności (1), wyznaczono prawdopodobieństwa zgonu postaci:

$$\hat{q}_x = 1 - \exp \left\{ \frac{Bc^x(1-c)}{\ln c} \right\}, \quad (2)$$

$$\hat{q}_x = 1 - \exp\left\{\frac{Bc^x(1-c)}{\ln c} - A\right\}, \quad (3)$$

$$\hat{q}_x = 1 - \exp\left\{\frac{B}{n+1}\left(x^{n+1} - (x+1)^{n+1}\right)\right\}. \quad (4)$$

Parametry A , B i c można wyznaczyć na podstawie wzorów (2), (3), (4), minimalizując różnice między wartościami q_x , wziętymi z tablic trwania życia, a wartościami $\hat{q}_x$, występującymi we wzorach (2), (3) i (4).

3. Mierniki różnic między wartościami q_x a aproksymowanymi wartościami $\hat{q}_x$

Mierniki różnic M między wartościami q_x a $\hat{q}_x$ mogą być następujące:

- suma kwadratów odległości (tzw. błąd średniokwadratowy)

$$M_1 = \sum_{i=1}^n (q_i - \hat{q}_i)^2, \quad (5)$$

- suma odległości

$$M_2 = \sum_{i=1}^n |q_i - \hat{q}_i|, \quad (6)$$

- maksimum odległości

$$M_3 = \max_i |q_i - \hat{q}_i|, \quad (7)$$

- średni błąd procentowy

$$M_4 = \frac{1}{n} \sum_{i=1}^n \frac{|q_i - \hat{q}_i|}{q_i}. \quad (8)$$

Im mniejsze wartości wymienionych mierników, tym lepsze dopasowanie estymatora.

4. Przykład

Na podstawie danych dotyczących umieralności mężczyzn w roku 2002 podjęto próbę aproksymacji wartości q_x , a następnie funkcji intensywności zgonów. Rozważono przypadki, kiedy funkcje q_x przybierają postać (3), (4) lub postać wielomianu trzeciego stopnia. Przyjmowano jedną postać funkcji q_x dla całego rozwa-

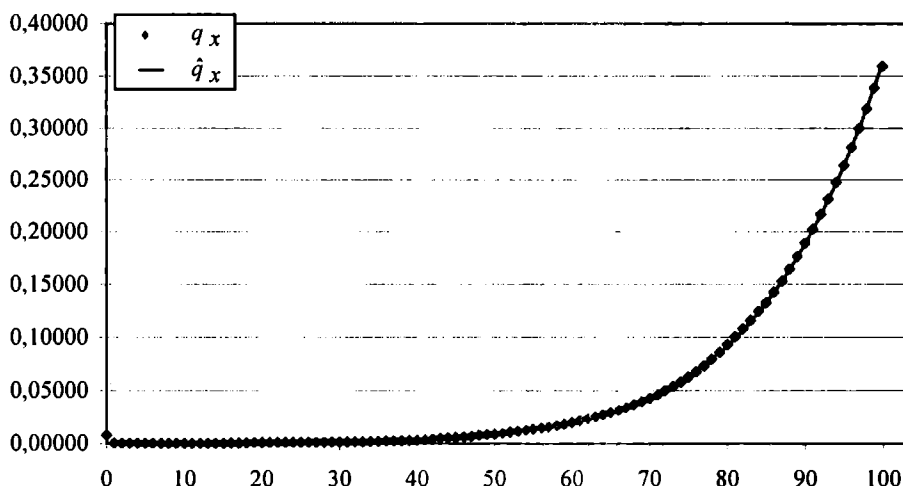
zanego okresu, albo oddzielne postaci funkcji q_x dla wyróżnionych okresów czasu (trzy segmenty). Uwzględniano różne mierniki różnic między wartościami q_x a wartościami $\hat{q}_x$.

W każdym przypadku (tj. dla ustalonego miernika różnic między q_x a wartościami $\hat{q}_x$ oraz dla ustalonej postaci funkcji q_x) wielokrotnie wyznaczano parametry funkcji $\hat{q}_x$ i ostatecznie wybierano ten „zestaw” parametrów, dla których miernik M przyjął wartość najmniejszą.

Jeśli funkcja $\hat{q}_x$ była wielomianem, wówczas do wyznaczenia jej parametrów zamiast metod numerycznych wykorzystano metodę najmniejszych kwadratów. Minimalizowano wtedy wartość M_1 .

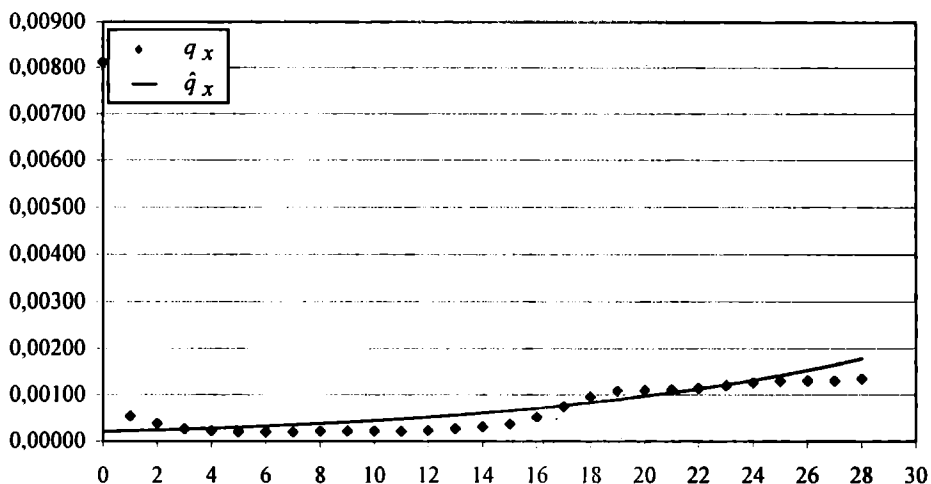
4.1. Aproksymacja parametrów funkcji Gompertza–Makehama

W pierwszej kolejności przyjęto, że $\hat{q}_x$ jest dane wzorem (3) i jako miernik dopasowania przyjęto M_1 . Otrzymano następujące parametry wymienionej funkcji: $A = 0$, $B = 0,0002$, $C = 1,0798$ oraz wartość $M_1 = 0,00011$. Wartości q_x oraz $\hat{q}_x$ podano na rys. 1 oraz na rys. 2. Z rysunku 2 wynika, że pojawia się szczególnie duża różnica między q_x oraz $\hat{q}_x$ dla roku zerowego. Dodatkowo widać spadek wartości q_x w latach: 0, 1, 2, 3 oraz wzrost wartości q_x w kolejnych latach, podczas gdy wartości $\hat{q}_x$ zawsze rosną.



Rys. 1. Wartości q_x oraz wartości $\hat{q}_x$ wyznaczone za pomocą funkcji Gompertza–Makehama

Źródło: opracowanie własne.



Rys. 2. Wartości q_x oraz $\hat{q}_x$ wyznaczone za pomocą funkcji Gompertza–Makehama, dane dla osób w wieku do 28 lat (0–28)

Źródło: opracowanie własne.

Z tych powodów wyznaczono funkcję Gompertza–Makehama ponownie, ale dla osób w wieku od 4 do 100 lat. Parametry wyznaczano trzema sposobami, za każdym razem minimalizując wartość jednego z mierników M_1 , M_2 , M_3 . Wartości parametrów, wymienionych mierników M oraz dodatkowo miernika M_4 podano w tab. 2. Widać, że trzema sposobami otrzymano dosyć podobne parametry. W każdym przypadku parametr A wynosi zero, co oznacza, że dla rozważanych danych funkcja Gompertza–Makehama jest równoważna funkcji Gompertza. We wszystkich przypadkach błąd procentowy M_4 był dość duży, rzędu 20–30%.

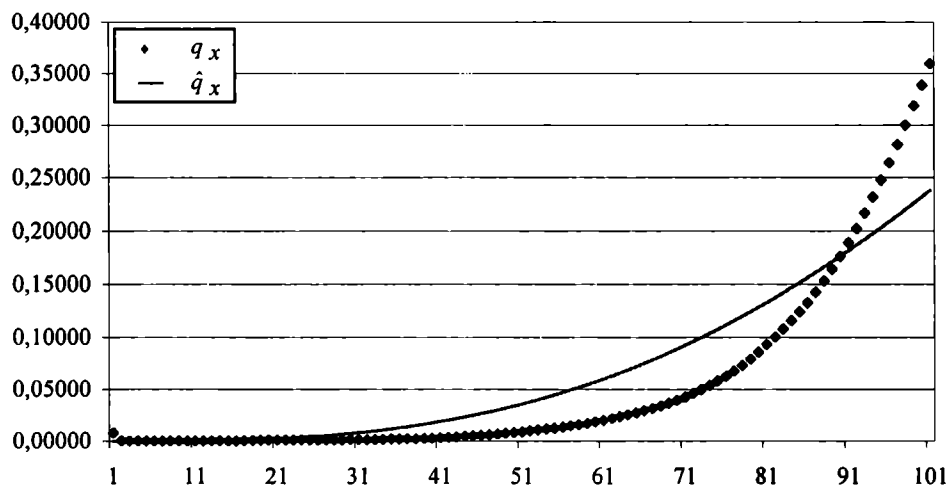
Tabela 2. Parametry funkcji Gompertza–Makehama dla różnych kryteriów oceny różnicy między q_x a $\hat{q}_x$

Kryterium różnicy między q_x a $\hat{q}_x$	A	B	C	M_1	M_2	M_3	M_4
min M_1	0	0,0002	1,07978	4,858E-05	0,056116	0,002476	19,4
min M_2	0	0,00019	1,080407	6,938E-05	0,050585	0,004537	17,1
min M_3	0	0,000234	1,077966	0,0001813	0,108104	0,00255	27,6

Źródło: opracowanie własne.

4.2. Aproksymacja parametrów funkcji Weibulla

W drugiej kolejności przyjęto, że funkcja q_x ma postać funkcji Weibulla, czyli dana jest wzorem (4). Funkcję tę wyznaczono dla osób w wieku do 100 lat (0-100), minimalizując wartość M_1 (rys. 3).



Rys. 3. Wartości q_x oraz wartości $\hat{q}_x$ wyznaczone za pomocą funkcji Weibulla

Źródło: opracowanie własne.

Z tych samych powodów co poprzednio wyznaczono dodatkowo funkcję Weibulla dla osób w wieku od 4 do 100 lat. Parametrów tej funkcji szukano trzema sposobami, tak by zminimalizować wartość ustalonego błędu: M_1 , M_2 lub M_3 . Wyniki podano w tab. 3. Wynika z nich, że w każdym przypadku otrzymano bardzo duże procentowe różnice między q_x a $\hat{q}_x$ – wartości M_4 są rzędu 100-1000%.

Tabela 3. Parametry funkcji Weibulla dla różnych kryteriów oceny różnicy między q_x a $\hat{q}_x$

Kryterium różnicy między q_x a $\hat{q}_x$	A	B	C	M_1	M_2	M_3
min M_1	4,05E-07	2,91028	0,118889	2,486776	0,121029	190,9
min M_2	4,05E-07	2,91028	0,145823	2,268907	0,1623	142,5
min M_3	0,000203	1,573974	0,479272	5,888284	0,109094	1261,0

Źródło: opracowanie własne.

4.3. Aproksymacja parametrów funkcji q_x w segmentach

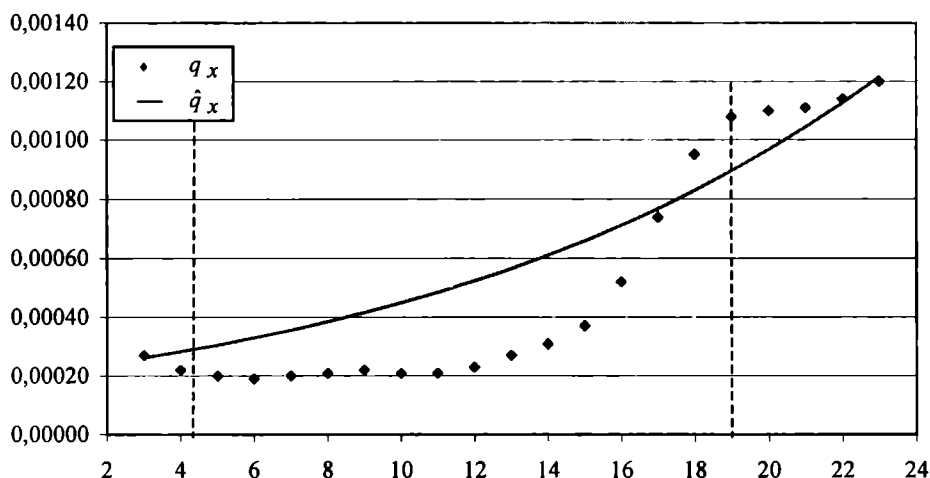
W poprzednich przypadkach wyznaczanie funkcji q_x dla wszystkich lub prawie wszystkich danych powodowało duże błędy procentowe M_4 . Dlatego postanowiono wyznaczyć oddzielne funkcje dla wyróżnionych okresów. Zbadano zmiany wartości q_x i przykładowej funkcji Gompertza–Makehama w poszczególnych latach i wyróżniono trzy okresy (segmenty) danych, (rys. 4, 5 i 6):

- okres 1 (segment 1), wiek 4-19 lat,
- okres 2 (segment 2), wiek 20-50 lat,
- okres 3 (segment 3), wiek 51-100 lat.

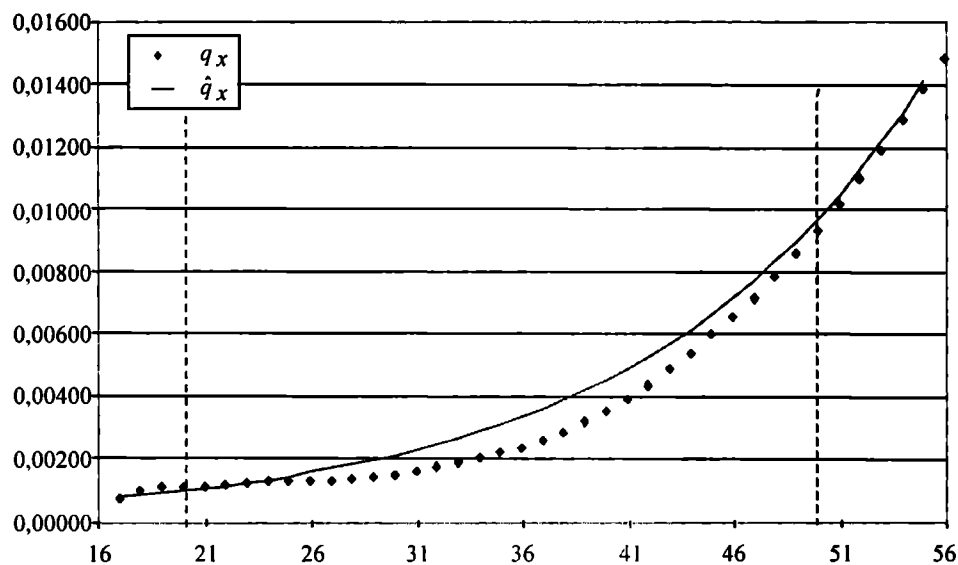
Dla poszczególnych segmentów osobno wyznaczono funkcję Gompertza–Makehama oraz wielomian trzeciego stopnia. W każdym przypadku minimalizowano wartość M_1 . Parametry funkcji q_x oraz wartości M_1 i dodatkowo pozostałych mierników M podano w tab. 4 i 5.

Analiza ostatniego wiersza tab. 4, w którym podano sumaryczne wartości M_1 , M_2 , M_3 oraz uśrednioną wartość M_4 dla wszystkich segmentów łącznie, pozwala stwierdzić, że wyznaczanie funkcji Gompertza–Makehama w segmentach powoduje wyraźne zmniejszenie wartości mierników M w porównaniu z przypadkiem, kiedy funkcja ta jest wyznaczana dla danych 4-100. Na przykład błąd M_4 zmniejsza się z wartości rzędu 20-30% do wartości 2,48%.

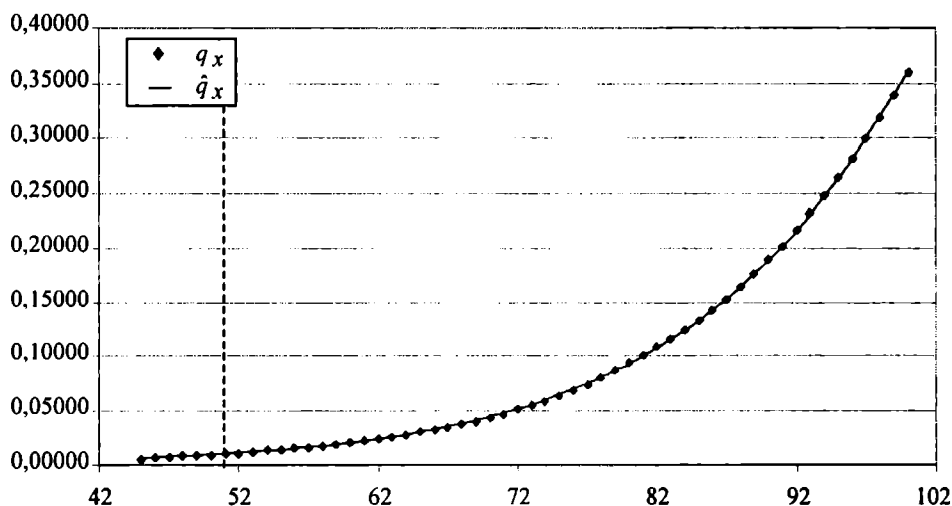
Jak pokazuje dodatkowo tab. 5, jeszcze mniejsze wartości M można uzyskać, stosując wielomian. Ostatecznie najmniejsze różnice między q_x a $\hat{q}_x$ występują



Rys. 4. Wartości q_x dla osób w wieku od 2 do 24 lat

Rys. 5. Wartości q_x dla osób w wieku od 17 do 56 lat

Źródło: opracowanie własne.

Rys. 6. Wartości q_x dla osób w wieku od 44 do 100 lat

Źródło: opracowanie własne.

w segmentach pierwszym i drugim dla wielomianu oraz w segmencie trzecim dla funkcji Gomperta–Makehama. W tym przypadku średni błąd M_4 wyniesie 2,09%.

Tabela 4. Funkcja Gompertza–Makehama w poszczególnych segmentach

Segmenty	A	B	C	M_1	M_2	M_3	$M_4[\%]$
4-19	0,0002	1E-06	1,4335	5E-05	0,0561	0,00248	19,4
20-50	0,0007	3E-05	1,1202	4E-07	0,0034	0,00025	4,9
51-100	0	0,0002	1,0797	3E-05	0,036	0,0023	1,4
Wszystkie segmenty				3E-05	0,0353	0,00192	2,48

Źródło: opracowanie własne.

Tabela 5. Wielomian trzeciego stopnia w poszczególnych segmentach

Segmenty	D	C	B	A	M_1	M_2	M_3	$M_4[\%]$
4-19	-5E-05	0,0001	-2E-05	7E-07	1E-08	0,0004	6,1E-05	6,4
20-50	-0,0031	0,0005	-2E-05	4E-07	5E-08	0,001	7,2E-05	1,4
51-100	-0,7038	0,0354	-0,0006	4E-06	3E-05	0,0339	0,00192	1,9

Źródło: opracowanie własne.

4.4. Funkcja intensywności

W przypadku kiedy funkcja q_x była aproksymowana w segmentach otrzymano niższe wartości mierników M niż w przypadku kiedy funkcja q_x była aproksymowana dla danych z zakresu wieku 4-100 lat. Dlatego podaną niżej funkcję intensywności wyznaczono na podstawie funkcji q_x , aproksymowanych w określonych wyżej segmentach.

Na podstawie oszacowanych parametrów (tab. 4) funkcja intensywności, spełniająca prawo Gompertza–Makehama, ma następującą postać

$$\mu(t) = \begin{cases} 0,0002 + 0,000001 \cdot 1,4335^t & \text{dla } t \in [4, 20), \\ 0,0007 + 0,00003 \cdot 1,1202^t & \text{dla } t \in [20, 51), \\ 0 + 0,0002 \cdot 1,079^t & \text{dla } t \in [51, 100). \end{cases}$$

Poniżej podano także funkcję intensywności wyznaczoną na podstawie q_x , dla której otrzymano najmniejszą wartość miernika M_1 (w dwóch pierwszych segmentach q_x jest wielomianem trzeciego stopnia, w ostatnim segmencie q_x to funkcja Gompertza–Makehama). Jeśli q_x jest wielomianem trzeciego stopnia, to nie można podać jawnej postaci funkcji intensywności. Dlatego funkcję intensywności zapisano w postaci uwikłanej, jako:

$$1 - \exp \left\{ - \int_x^{x+1} \mu(t) dt \right\} = q_x = \begin{cases} -0,00005 + 0,0001t - 0,00002t^2 & \text{dla } t \in [4, 20), \\ -0,0031 + 0,0005t - 0,00002t^2 + 4 \cdot 10^{-7} t^3 & \text{dla } t \in [20, 51), \\ 1 - \exp \left\{ \frac{0,0002 \cdot 1,0797^t (1 - 1,0797)}{\ln 1,0797} \right\} & \text{dla } t \in [51, 100). \end{cases}$$

5. Wnioski

1. W badanym przypadku otrzymuje się duże różnice między wartościami q_x a wartościami aproksymowanymi $\hat{q}_x$, kształtujące się zgodnie z prawem Gompertza–Makehama lub Weibulla.
2. Wyznaczenie funkcji $\hat{q}_x$ Gompertza–Makehama w segmentach wyraźnie zmniejsza te różnice.
3. Funkcja $\hat{q}_x$ postaci wielomianu trzeciego stopnia jest lepiej dopasowana do danych dla pierwszych dwóch segmentów niż funkcja $\hat{q}_x$ postaci Gompertza–Makehama. Jednak przy założeniu, że q_x jest wielomianem, nie można funkcji intensywności przedstawić w postaci jawnej.

Literatura

Bolesławski L. (1973), *Budowa tablic trwania życia*, GUS, Warszawa.
Demografia. Metody analizy i prognozowania (1992), red. M. Cieślak, PWN, Warszawa.
 Ostasiewicz S. (2000), *Składki w wybranych typach ubezpieczeń życiowych*, AE, Wrocław.

THE APPROXIMATION OF THE FORCE OF MORTALITY

Summary

The mortality analysis of the population is carried out on the basis of the life tables. They include the discrete functions (e.g. death probability function) but their continuous equivalents can be approximated (e.g. the force of mortality).

In the paper we have carried out the approximation of the force of mortality on the basis of the data taken from the Polish life tables. In order to do that we have used the analytical distribution of the future lifetime (the simple formula expressing the force of mortality) known from the insurance literature. 3rd order polynomial was also utilized.

Mgr inż. Magdalena Sipa i dr Renata Gnitecka są pracownikami Katedry Statystyki Akademii Ekonomicznej we Wrocławiu.