PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 1 (1201) 2008
Statystyka aktuarialna – teoria i praktyka

Stanisław Heilpern
Akademia Ekonomiczna we Wrocławiu

ROZKŁAD SUM ZALEŻNYCH ZMIENNYCH LOSOWYCH, ZASTOSOWANIE W ZAGADNIENIACH AKTUARIALNYCH

1. Wstęp

Sumy zmiennych losowych stanowią podstawę najważniejszych modeli aktuarialnych [1; 17; 19]. Wystarczy wymienić klasyczne już modele ryzyka indywidualnego czy kolektywnego lub teorię ruiny. Jednakże w modelach tych przyjmuje się dość nierealistyczne, ale wygodne z matematycznego punktu widzenia, założenie o niezależności badanych zmiennych losowych reprezentujących poszczególne rodzaje ryzyka.

W praktyce rzadko możemy mówić o niezależności. W wielu sytuacjach rozpatrywane rodzaje ryzyka są zależne, i to w sposób istotny [7; 8; 20]. Dzieje się tak np. w przypadku ryzyka katastroficznego, gdzie występuje dość duża zależność między poszczególnymi rodzajami ryzyka [15]. Podobna sytuacja występuje w ubezpieczeniach komunikacyjnych [3] czy życiowych, gdzie można zauważyć występowanie zależności długości życia współmałżonków [2].

W prezentowanej pracy przedstawiona została analiza zależności rozpatrywanych zmiennych losowych w klasycznych już modelach aktuarialnych. Praca ma charakter przeglądowy i jest kontynuacją wcześniejszych prac autora [10; 12]. Omówione zostały sumy zależnych zmiennych losowych, w tym zależny rozkład dwumianowy, sumy losowe, stanowiące podstawę modelu kolektywnego ryzyka, model ryzyka indywidualnego i zagadnienie ruiny. W każdym zagadnieniu zwrócono uwagę na wpływ stopnia zależności na rozkład i własności otrzymanej sumy zmiennych losowych. Do wyznaczenia tych rozkładów stosowane były na ogół metody symulacyjne.

Przedstawione w artykule przykłady są proste. Można je oczywiście rozwinąć, uogólnić na bardziej skomplikowane przypadki, jednak zasadniczym celem prezentowanej pracy jest przedstawienie ogólnej idei aktuarialnych modeli opartych na sumach zależnych zmiennych losowych i uchwycenie występujących tam podstawowych prawidłowości.

2. Modelowanie zależności

W pracy wykorzystane zostały trzy wybrane metody modelowania zależności. Pierwsza metoda jest oparta na funkcjach łączących (ang. *copula*). Funkcja łącząca $C:[0,1]^n \to [0,1]$ jest łącznikiem między dystrybuantą F rozkładu łącznego zmiennych losowych X_1, X_2, ..., X_n a dystrybuantami rozkładów brzegowych F_i [7; 16]:

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)).$$

Jest to n-wymiarowa dystrybuanta zmiennych losowych U_1, U_2, ..., U_n o rozkładzie jednostajnym na [0, 1], gdzie $U_i = F_i(X_i)$. Niezależności odpowiada funkcja łącząca $\prod(u_1, u_2, \ldots, u_n) = u_1, u_2, \ldots, u_n$, a współmonotoniczności (ang. *comonotonicity*) funkcja $M(u_1, u_2, \ldots, u_n) = \min(u_1, u_2, \ldots, u_n)$. Zmienne losowe X_1, X_2, ..., X_n są współmonotoniczne, jeśli istnieją niemalejące funkcje f_1, f_2, ..., f_n i zmienna losowa Z taka, że wektory losowe $(X_1, X_2, \ldots, X_n)$ oraz $(f_1(Z), f_2(Z), \ldots, f_n(Z))$ mają ten sam rozkład [6].

W praktyce wykorzystuje się zwykle indeksowane rodziny funkcji łączących. Do najbardziej popularnych należą rodziny archimedesowych funkcji łączących [7; 16]. Funkcja łącząca C jest archimedesowa, jeśli istnieje funkcja $\varphi: (0,1] \to R_+$, nazywana generatorem, taka, że $\varphi(1) = 0$,

$$(-1)^k \left(\varphi^{-1}\right)^{(k)}(x) \geq 0$$

dla $1 \leq k \leq n$, a $f^{(n)}$ jest n-tą pochodną funkcji f, oraz

$$C(u_1, u_2, \ldots, u_n) = \varphi^{-1}\left(\varphi(u_1) + \varphi(u_2) + \ldots + \varphi(u_n)\right).$$

Przykładem archimedesowej funkcji łączącej jest funkcja Claytona określona wzorem:

$$C_\alpha(u_1, u_2, \ldots, u_n) = \left(u_1^{-\alpha} + u_2^{-\alpha} + \ldots + u_n^{-\alpha} - n + 1\right),$$

gdzie $0 < \alpha$. Funkcje łączące Claytona C_α tworzą indeksowaną parametrem α archimedesową rodzinę funkcji łączących. W granicy, gdy parametr α dąży do zera, otrzymujemy funkcję łączącą niezależności Π, a dla α dążącej do nieskończoności funkcję łączącą współmonotoniczną M.

Dwuwymiarowa funkcja łącząca jednoznacznie wyznacza nam współczynnik korelacji rangowej Kendalla τ. Dla rodziny Claytona jest on związany z parametrem α wzorem [16]

$$\tau = \frac{\alpha}{\alpha + 2}.$$

W przypadku wielowymiarowych, archimedesowych funkcji łączących, brzegowe funkcje łączące wyznaczone dla każdej pary zmiennych losowych X_i, X_j są identyczne. Dlatego też każda para zmiennych ma ten sam współczynnik korelacji Kendalla.

Inny sposób modelowania zależności polega na wykorzystywaniu warunkowej niezależności, występującej dla konkretnych wartości θ pewnej zmiennej losowej Θ. Wtedy dla ustalonych wartości tej zmiennej stosujemy znane, klasyczne metody stosowane w przypadku niezależności. Dystrybuanta łączna przyjmuje wtedy postać mieszanki [8; 20]

$$F(x_1, x_2, \ldots, x_n) = \int_0^\infty F(x_1, x_2, \ldots, x_n | \theta) dF_\Theta(\theta) = \int_0^\infty F_1(x_1|\theta) F_2(x_2|\theta) \ldots F_n(x_n|\theta) dF_\Theta(\theta).$$

Indukowaną zmienną losową Θ możemy traktować jako wspólny czynnik wpływający na wszystkie zmienne losowe X_i. W przypadku gdy struktura zależności zmiennych X_1, X_2, ..., X_n opisana jest archimedesową funkcją łączącą, to istnieje zmienna losowa Θ, dla której zachodzi warunkowa niezależność [8; 11; 13].

W klasycznych zagadnieniach aktuarialnych, takich jak model kolektywnego ryzyka czy teoria ruiny, rozpatruje się losowe sumy zmiennych losowych, postaci

$$S = \sum_{i=1}^{N} X_i,$$

gdzie N jest pewną zmienną losową. W przypadku wieloklasowym, gdzie występują różne klasy szkód, zachodzi zwykle w sytuacjach praktycznych zależność między liczebnościami N_i poszczególnych klas. W najprostszym przypadku dwóch klas jednym ze sposobów modelowania zależności może być rozbicie zmiennej liczącej szkody na dwa czynniki [22]:

$$N_1 = M_1 + M_0,$$

$$N_2 = M_2 + M_0,$$

gdzie zmienne losowe M_0, M_1, M_2 są niezależne. Zmienne losowe M_1, M_2 przedstawiają czynniki wewnętrzne, charakterystyczne dla poszczególnych klas, a zmienna M_0 reprezentuje czynnik zewnętrzny oddziałujący na obydwie klasy. W przypadku większej liczby klas można też rozpatrywać czynniki działające na pary klas, na trójki, czwórki itp. [5].

Wektory losowe $\mathbf{X} = (X_1, X_2, \dots, X_n)$ możemy porządkować, wprowadzając relacje porządku uwzględniające strukturę zależności [13; 16]. Najprostszym porządkiem tego typu jest naturalne uporządkowanie funkcji łączących przedstawiających odpowiednie struktury zależności:

$$C \le C' \Leftrightarrow C(u_1, u_2, \dots, u_n) \le C'(u_1, u_2, \dots, u_n)$$

dla każdych $0 \le u_1, u_2, \dots, u_n \le 1$. Można pokazać, że dla wielowymiarowej ($n > 2$) archimedesowej funkcji łączącej zachodzi relacja [11; 16]

$$\Pi \le C \le M.$$

Jeśli indeksowana parametrem α rodzina funkcji łączących C_α jest uporządkowana względem relacji $\le$ zgodnie z naturalnym porządkiem parametrów, tzn. gdy zachodzi

$$\alpha \le \alpha' \Rightarrow C_\alpha \le C_{\alpha'},$$

to mówimy, że rodzina ta jest regularna. Przykładem takiej rodziny jest rodzina Claytona.

Inną relacją porządkującą wektory losowe jest porządek supermodularny $\le_{\mathrm{sm}}$. Mówimy, że wektor losowy $\mathbf{X} = (X_1, X_2, \dots, X_n)$ jest mniejszy w porządku supermodularnym od wektora $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$, tzn. $\mathbf{X} \le_{\mathrm{sm}} \mathbf{Y}$, jeśli [15]

$$E(f(\mathbf{X})) \le E(f(\mathbf{Y}))$$

dla wszystkich funkcji supermodularnych f takich, że wartość oczekiwana istnieje. Funkcja f: $R^n \to R$ jest supermodularna, jeśli

$$f(\mathbf{x} \wedge \mathbf{y}) + f(\mathbf{x} \vee \mathbf{y}) \le f(\mathbf{x}) + f(\mathbf{y})$$

dla każdych $\mathbf{x}, \mathbf{y} \in R^n$, gdzie $\mathbf{x} \wedge \mathbf{y} = (\min\{x_1, y_1\}, \min\{x_2, y_2\}, \dots, \min\{x_n, y_n\})$ i $\mathbf{x} \vee \mathbf{y} = (\max\{x_1, y_1\}, \max\{x_2, y_2\}, \dots, \max\{x_n, y_n\})$.

Jeśli F oraz G są dystrybuantami wektorów losowych $\mathbf{X}$ oraz $\mathbf{Y}$ i zachodzi relacja $\mathbf{X} \le_{\mathrm{sm}} \mathbf{Y}$, to można też mówić o relacji supermodularnej między dystrybuantami, czyli $F \le_{\mathrm{sm}} G$. Ponadto, gdy zachodzi $\mathbf{X} \le_{\mathrm{sm}} \mathbf{Y}$, to dla dowolnych funkcji niemalejących f_1, f_2, ..., f_n otrzymujemy $(f_1(X_1), f_2(X_2), \dots, f_n(X_n)) \le_{\mathrm{sm}} (f_1(Y_1), f_2(Y_2), \dots, f_n(Y_n))$ [15]. Istnieje więc zgodność między uporządkowaniem supermodularnym między wektorami losowymi $\mathbf{X}$ oraz $\mathbf{Y}$ a uporządkowanymi według tej relacji, odpowiadającymi im funkcjami łączącymi, które są dystrybuantami łącznych rozkładów zmiennych losowych o rozkładzie jednostajnym na [0, 1]. Poniższe twierdzenie podane przez Wei i Hu w [21] umożliwia nam sprawdzenie, że dana rodzina funkcji łączących jest uporządkowana według porządku supermodularnego.

Twierdzenie 1 [21]. Jeśli φ_1 oraz φ_2 są generatorami archimedesowych funkcji łączących C_1 oraz C_2, spełniającymi warunek $\varphi_1 \circ \varphi_2^{-1} \in L_\infty^*$, gdzie $L_\infty^* = \left\{\psi: R_+ \to R_+ \middle| \psi(0)=0, \psi(\infty)=\infty, (-1)^{i-1}\psi^{(i)}(x) \geq 0, i \geq 1, x \geq 0\right\}$, to $C_1 \leq_{\mathrm{sm}} C_2$.

Przykład 1. Niech C_α oraz C_β, gdzie $\alpha < \beta$, będą archimedesowymi funkcjami łączącymi należącymi do rodziny Claytona. Wtedy otrzymujemy $\varphi_\alpha(u) = u^{-\alpha} - 1$, $\varphi_\beta^{-1}(t) = (1+t)^{-1/\beta}$ oraz

$$\varphi_\alpha \circ \varphi_\beta^{-1}(t) = (1+t)^{\alpha/\beta} - 1.$$

Łatwo sprawdzić, że $\varphi_\alpha \circ \varphi_\beta^{-1}(t) \in L_\infty^*$. W rezultacie otrzymujemy $C_\alpha \leq_{\mathrm{sm}} C_\beta$.

3. Sumy zmiennych losowych

Rozpatrzmy na początek prosty przypadek zagregowanej wartości wypłat, traktowanej jako suma zmiennych losowych X_1, X_2, ..., X_n o tym samym rozkładzie z dystrybuantą F

$$S = X_1 + X_2 + \ldots + X_n.$$

W odróżnieniu od klasycznego podejścia [1; 19], nie zakładamy tutaj niezależności występujących zmiennych losowych X_i. Założymy, że struktura zależności opisana jest funkcją łączącą C.

W przypadku niezależności, czyli funkcji łączącej Π, dystrybuanta sumy F_S jest n-tym splotem dystrybuanty F:

$$F_S(x) = F^{*n}(x).$$

Gdy wypłaty X_i są współmonotoniczne otrzymujemy $S \overset{d}{=} nX_1$. W pozostałych pośrednich przypadkach, bardziej interesujących z punktu widzenia praktycznych zastosowań, niestety nie można podać jawnego wzoru, lecz jedynie można wykorzystać metody symulacyjne.

Do porządkowania ryzyka związanego z sumą wypłat S dla różnych struktur zależności może być wykorzystany porządek zatrzymania straty $\leq_{\mathrm{sl}}$ oraz poniżej przedstawione twierdzenie 2 podane przez Müllera w [15]. Między dwiema zmiennymi losowymi S oraz S' zachodzi relacja zatrzymania straty, tzn. $S \leq_{\mathrm{sl}} S'$ [19], jeśli $E(S-d)^+ \leq E(S'-d)^+$ dla każdej retencji $d \geq 0$. Jest to równoważne ze stwierdzeniem, że dla każdej wypukłej i rosnącej funkcji f zachodzi $E(f(S)) \leq E(f(S'))$.

Twierdzenie 2 [15]. Jeśli dwa wektory losowe $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ oraz $\mathbf{X}' = (X_1', X_2', \ldots, X_n')$ są uporządkowane według porządku supermodularnego, tzn. $\mathbf{X} \leq_{\mathrm{sm}} \mathbf{X}'$, to dla ich sum zachodzi porządek zatrzymania straty, czyli $S \leq_{\mathrm{sl}} S'$.

W naszym przypadku wszystkie zmienne losowe X_i, X_i', przedstawiające wielkości wypłat, mają ten sam rozkład, a struktura zależności wektorów $\mathbf{X}$ oraz $\mathbf{X}'$ jest wyznaczona przez różne funkcje łączące C oraz C'. Jeśli stwierdzimy, że zachodzi $\mathbf{X} \leq_{\mathrm{sm}} \mathbf{X}'$, co jest równoznaczne z $C \leq_{\mathrm{sm}} C'$, to na podstawie twierdzenia 2 łączna wypłata $S' = \sum_{i=1}^{n} X_i'$ jest „większa", bardziej ryzykowna dla ubezpieczyciela, od S.

Stosując do modelowania zależności archimedesową funkcję łączącą C, otrzymujemy zależności $\Pi \leq_{\mathrm{sm}} C \leq_{\mathrm{sm}} M$, które pociągają za sobą relację

$$S_{\Pi} \leq_{\mathrm{sl}} S \leq_{\mathrm{sl}} S_M.$$

Z przykładu 1 i z powyższych rozważań wynika, że rodzina Claytona jest uporządkowana według porządku supermodularnego, a odpowiadające sumy wypłat będą wraz ze wzrostem parametru α uporządkowane według porządku zachowania straty. Innymi słowy, wzrost zależności między poszczególnymi wypłatami spowoduje wzrost ryzyka.

Gdy struktura zależności opisana jest funkcją łączącą C należącą do regularnej rodziny archimedesowych funkcji łączących, wzrost zależności zwiększa wariancję zagregowanej wypłaty S. Wariancja sumy zmiennych losowych jest wtedy równa

$$V(S) = \sum_{j=1}^{n} V\left(X_j\right) + 2\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \mathrm{Cov}\left(X_j, X_k\right). \qquad (1)$$

Wartości współczynnika korelacji Pearsona r_C między zmiennymi losowymi X_j oraz X_k są uporządkowane zgodnie z naturalnym porządkiem funkcji łączących [18]. Zachodzi wtedy relacja

$$C_1 \leq C_2 \Rightarrow r_{C_1} \leq r_{C_2},$$

gdzie C_1 oraz C_2 są funkcjami łączącymi należącymi do tej rodziny. Widzimy, że dla funkcji łączącej C_2, dla której zachodzi większa zależność, kowariancja $\mathrm{Cov}(X_i, X_j)$ jest dla każdej pary zmiennych losowych również większa, co powoduje wzrost wariancji sumy S.

Przykład 2. Rozpatrzmy sumę $n = 50$ zmiennych losowych X_i o jednakowym rozkładzie wykładniczym z wartością oczekiwaną $\mu = 10$, tzn. $X_i \sim W(10)$, gdzie struktura zależności opisana jest za pomocą funkcji łączącej Claytona C_α. W celu zbadania rozkładu sumy S_α dla różnych wartości parametru α przeprowadzono symulacje oparte na 20 000 powtórzeń. Symulacje dotyczyły przypadków gdy parametr α przyjmował wartości: 0; 0,86; 2; 8 oraz ∞. Odpowiada to wartościom współczynnika korelacji Kendalla: 0; 0,3; 0,5; 0,8 oraz 1. Odpowiednie histogramy przedstawiające uzyskane wyniki zamieszczone zostały na rys. 1. Widzimy, że wraz ze wzrostem wartości parametru α, czyli wzrostem zależności, zmienia się istotnie kształt otrzymanych wykresów, zwiększają się ogony tych rozkładów oraz rośnie wariancja. Dla rozpatrywanych wartości parametru α odchylenie standardowe sumy S_α jest odpowiednio równe: 70,71; 261,70; 339,05; 427,54 oraz 500.

Jednym z przykładów sumy zależnych zmiennych losowych jest warunkowy rozkład dwumianowy [12; 14], gdzie zmienne losowe X_i mają rozkład dwupunktowy z prawdopodobieństwem sukcesu $p_i = \Pr(X_i = 1)$. Rozkłady te stosuje się m.in. w reasekuracji nadwyżki szkody [16]. Zmienna losowa X_i przybiera wtedy wartość 1 gdy szkoda przekroczy ustalony próg retencji.

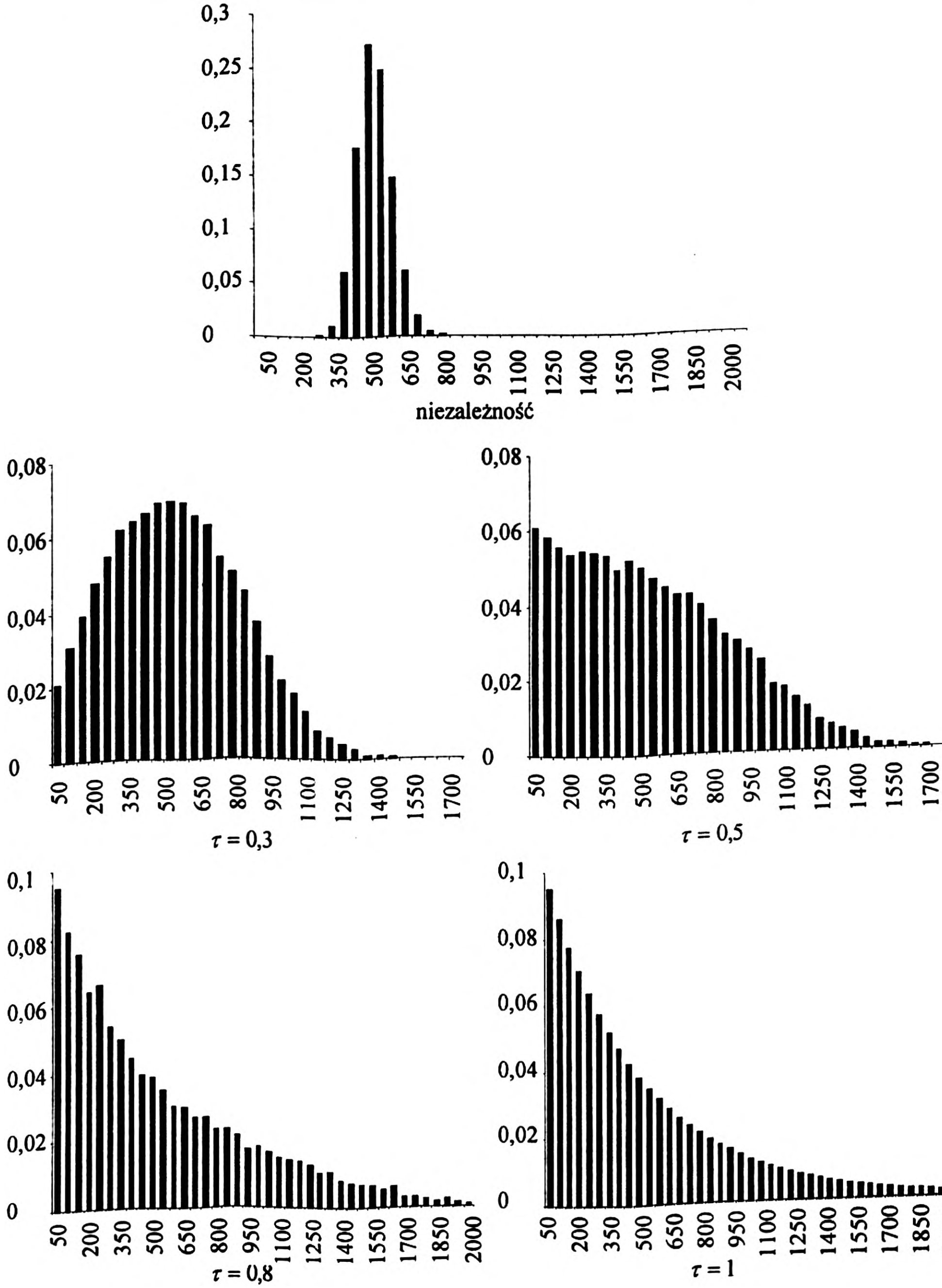
Dla zmiennej losowej $S = X_1 + X_2 + \ldots + X_n$ mającej warunkowy rozkład dwumianowy, gdy struktura zależności zmiennych X_i opisana jest funkcją łączącą C, można w sposób jawny określić jej rozkład [12]. W przypadku gdy zmienne losowe X_i mają ten sam rozkład, rozkład sumy S określony jest wzorem

$$P(S = k) = \sum_{j=0}^{k} (-1)^j \frac{n!}{(n-k)!\,j!\,(k-j)!} F_{k-j,n},$$

gdzie $F_{k,n} = \Pr(Y_{k+1} = 0, \ldots, Y_n = 0) = C(\underbrace{1, \ldots, 1}_{k}, \underbrace{q, \ldots, q}_{n-k})$ i $q = 1 - p$. Gdy zmienne losowe X_i są niezależne otrzymujemy klasyczny rozkład dwumianowy, dla archimedesowej funkcji łączącej z generatorem φ mamy

$$F_{k,n} = \varphi^{-1}((n-k)\varphi(q)),$$

a dla współmonotoniczności zachodzi zależność



Rys. 1. Rozkłady sum zależnych zmiennych losowych

Źródło: opracowanie własne.

$$P(S=k)=\begin{cases} q, & k=0, \\ p, & k=n. \end{cases}$$

Więcej informacji na temat tego warunkowego rozkładu dwumianowego czytelnik znajdzie w [12; 14].

Do wyznaczania rozkładu sumy zależnych zmiennych losowych może być wykorzystana warunkowa niezależność archimedesowych funkcji łączących [4; 5]. Gdy zależność jest opisana archimedesową funkcją łączącą C, to dystrybuanta sumy S jest równa mieszance

$$F_S(x)=\int_0^{\infty} F_{S_\theta}(x)dF_\Theta(\theta),$$

gdzie zmienna losowa Θ jest indukowana przez funkcję łączącą C. Warunkową dystrybuantę F_{S_θ} można wyznaczyć, stosując dla każdej wartości θ zmiennej losowej Θ klasyczne metody dotyczące niezależnych zmiennych losowych.

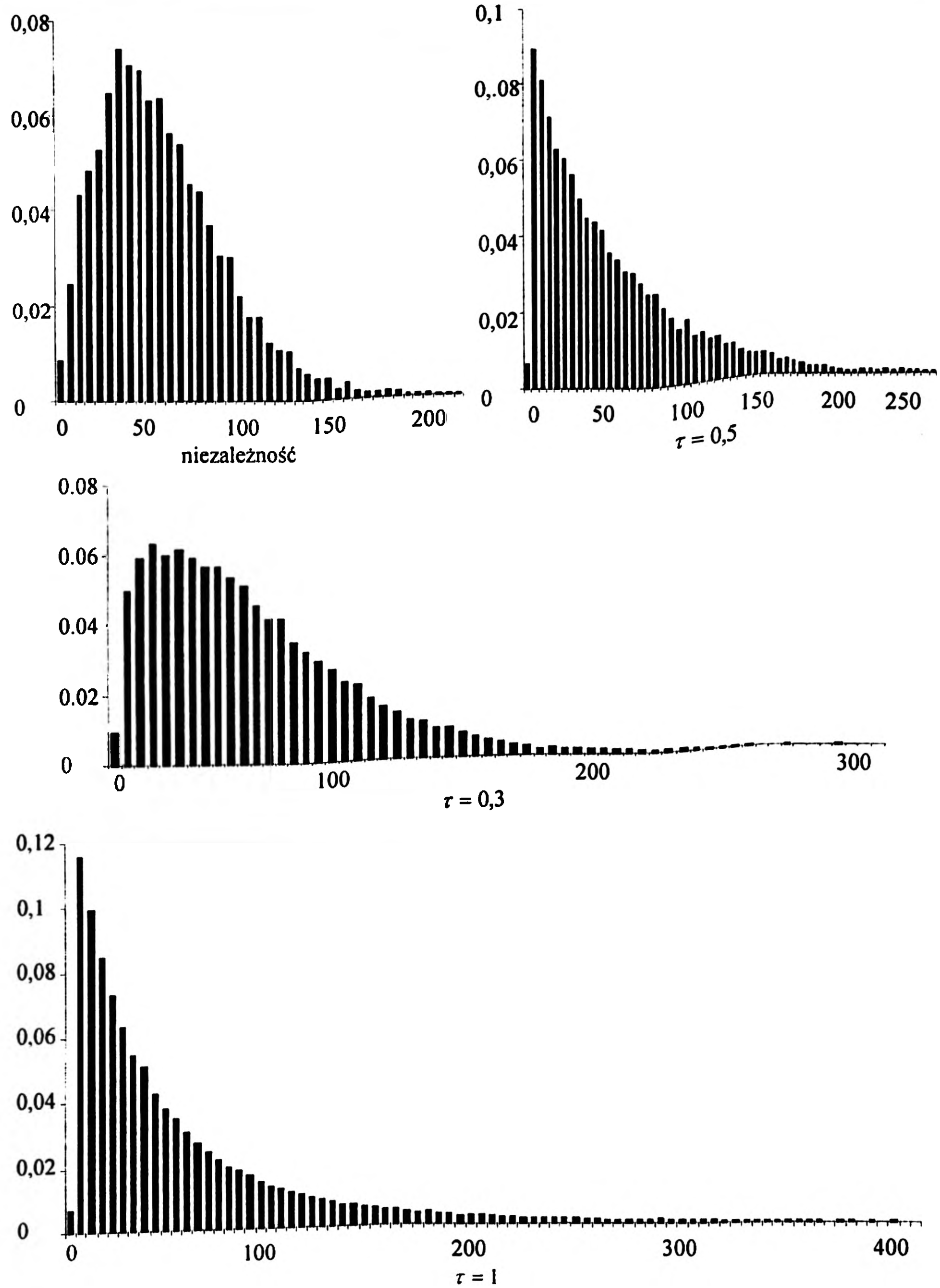
4. Sumy losowe

W zagadnieniach związanych z ryzykiem kolektywnym rozpatruje się sumy zmiennych losowych X_i, będących wartościami poszczególnych wypłat, których liczebność jest zmienną losową N [1, 17, 19]. Są to zmienne losowe postaci

$$S=X_1+X_2+\ldots+X_N.$$

W naszej pracy interesować nas będzie przypadek, gdy zmienne X_i mogą być zależne i ich struktura zależności opisana jest funkcją łączącą C. Natomiast zmienna losowa N przedstawiająca liczbę wypłat będzie niezależna od ich wielkości X_i. Jednakże, w tej sytuacji, nawet w najprostszych przypadkach nie można podać jawnych wzorów wyznaczających rozkład sumy. Możemy jedynie wykorzystać w tym celu metody symulacyjne.

Przykład 3. Niech poszczególne wypłaty X_i mają rozkład wykładniczy z wartością oczekiwaną $\mu = 10$, a liczba wypłat jest zmienną losową o rozkładzie Poissona z parametrem $\lambda = 5$. Załóżmy, że struktura zależności między wypłatami opisana jest funkcją łączącą Claytona C_α, oraz wyznaczmy metodami symulacyjnymi (20 000 powtórzeń) rozkłady sumy wypłat dla wartości parametru α wynoszącego 0; 0,86; 2 oraz ∞. Pierwszy przypadek odpowiada niezależności, dwa następne wartości współczynnika korelacji Kendalla wynoszącego odpowiednio 0,3 i 0,5, a ostatni dotyczy współmonotoniczności. Wyniki symulacji przedstawione są na rys. 2. Wartości odchylenia standardowego sumy S są odpowiednio równe: 31,62;



Rys. 2. Wykresy losowych sum zależnych zmiennych losowych

Źródło: opracowanie własne.

39,57; 45,65 oraz 58,03. Widzimy, że wraz ze wzrostem stopnia zależności między zmiennymi losowymi X_i zmienia się istotnie kształt histogramów przedstawiających rozkład sumy, rosną ogony tych rozkładów i zwiększa się wariancja.

Opierając się na twierdzeniu 2 można udowodnić następujący wniosek dotyczący sum losowych.

Wniosek. Jeśli dla każdego n wektory losowe $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ oraz $\mathbf{X}'_n = (X'_1, X'_2, \ldots, X'_n)$ są uporządkowane według porządku supermodularnego, tzn. $\mathbf{X} <_{\mathrm{sm}} \mathbf{X}'_n$, to dla ich sum losowych $S = \sum_{i=1}^{N} X_i$ i $S' = \sum_{i=1}^{N} X_i'$ zachodzi porządek zatrzymania straty, czyli $S \leq_{\mathrm{sl}} S'$.

Dowód. Niech f będzie dowolną rosnącą funkcją wypukłą, $S_n = \sum_{i=1}^{n} X_i$ oraz $S'_n = \sum_{i=1}^{n} X'_i$. Na podstawie twierdzenia 2 dla każdego n otrzymujemy $S \leq_{\mathrm{sl}} S'$.
Wtedy

$$E(f(S)) = \sum_{n=1}^{\infty} p_n E(f(S_n)) \leq \sum_{n=1}^{\infty} p_n E(f(S'_n)) = E(f(S')),$$

gdzie $p_n = \Pr(N = n)$.

Jeśli struktura zależności zmiennych losowych X_i jest opisana funkcją łączącą należącą do regularnej rodziny archimedesowej i spełnione są założenia Wniosku, to odpowiednie sumy losowe są uporządkowane według porządku zachowania straty. Opierając się na rozważaniach z paragrafu 3, wiemy, że rodzina Claytona spełnia te założenia i wraz ze wzrostem wartości parametru α, czyli wraz ze wzrostem stopnia zależności, wzrasta ryzyko związane z tak zagregowaną sumą wypłat.

5. Indywidualny model ryzyka

W pracy [4] przeprowadzona została analiza indywidualnego modelu ryzyka dopuszczającego zależność występowania szkód. Wyznaczony został rozkład łączny szkód

$$S = X_1 + X_2 + \ldots + X_n,$$

gdzie $X_i = I_i B_i$, B_i jest wielkością szkody z i-tej polisy, a zmienne I_i są zmiennymi dwupunktowymi z prawdopodobieństwem wystąpienia szkody $p_i = P(I_i = 1)$.

Przyjęto również, że indykatory I_1, I_2, ..., I_n są zależne i struktura zależności opisana jest funkcją łączącą C.

W przypadku gdy zarówno indykatory I_i, jak i wielkości szkód B_i mają ten sam rozkład, a funkcja łącząca jest archimedesowa z generatorem φ, dystrybuanta łącznej sumy szkód jest kombinacją wypukłą dystrybuant splotów zmiennych B_i, tzn.

$$F_S(x) = \sum_{k=0}^{n} a_k F_B^{*k}(x).$$

Współczynniki tej kombinacji a_k spełniają warunki: $a_k \geq 0$, $\sum_{k=0}^{n} a_k = 1$ oraz zależą jedynie od funkcji łączącej C i prawdopodobieństwa wystąpienia szkody p. Są one równe

$$a_k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} M_{n-k+j},$$

gdzie $M_k = M_\Theta(-k\varphi(q)) = \varphi^{-1}(k\varphi(q))$, a Θ jest zmienną losową indukowaną przez funkcję łączącą C. Jest to uogólnienie wzoru dotyczącego warunkowego rozkładu dwumianowego. W przypadku współmonotoniczności otrzymujemy prosty wzór

$$F_S(x) = q + (1-q) F_B^{*n}(x).$$

Wariancja sumy S określona jest wzorem (1), a kowariancja

$$\mathrm{Cov}(X_i, X_j) = \mu^2 \mathrm{Cov}(I_i, I_j),$$

gdzie $\mu = E(B_i)$. Wzrost zależności między indykatorami również spowoduje zwiększenie wartości wariancji sumy S. Wariancję sumy szkód można też wyznaczyć stosując wzór [4]

$$V(S) = np(\mu^2 q + \sigma^2) + n(n-1)\mu^2(\varphi^{-1}(2\varphi(q)) - q^2).$$

Przykład 4 [10]. Rozpatrzy przykład $n = 3$ polis z prawdopodobieństwem zajścia szkody $p = 0{,}1$. Szkody mają rozkład wykładniczy z wartością oczekiwaną $\mu = 1$. Wtedy k-ty splot zmiennej losowej B ma rozkład gamma $G(k, \mu)$. W przypadku klasycznym, dotyczącym niezależności szkód, dystrybuanta sumy S przybiera postać

$$F_S(x) = 0{,}729 + 0{,}243F_B(x) + 0{,}027F_B^{*2}(x) + 0{,}001F_B^{*3}(x),$$

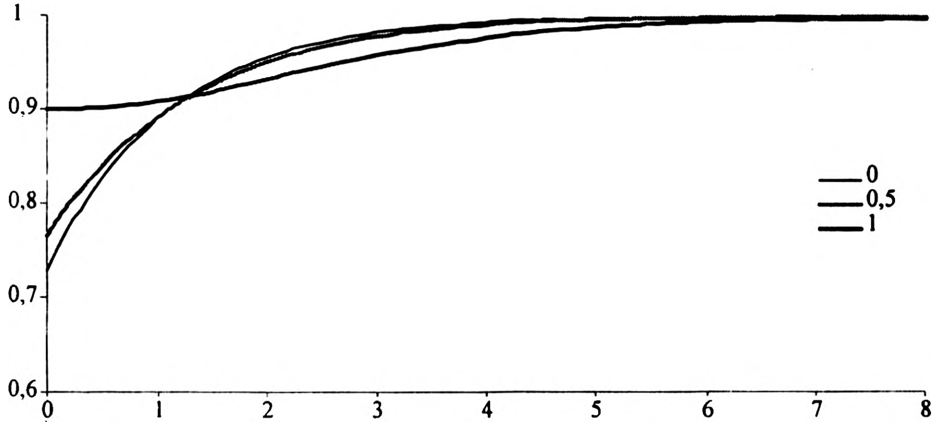
gdy struktura zależności opisana jest funkcją łączącą Claytona z parametrem $\alpha = 2$ mamy

$$F_S(x) = 0{,}766 + 0{,}177F_B(x) + 0{,}048F_B^{*2}(x) + 0{,}009F_B^{*3}(x),$$

a dla współmonotoniczności otrzymujemy

$$F_S(x) = 0{,}9 + 0{,}1F_B^{*3}(x).$$

Wykresy tych dystrybuant zamieszczone są na rys. 3. Z wykresu można zauważyć, że wraz ze wzrostem zależności rosną ogony rozkładów. Wzrasta też wariancja. W rozpatrywanych przypadkach odchylenie standardowe łącznego rozkładu szkód S jest odpowiednio równe: 0,755; 0,813 oraz 1,054.



Rys. 3. Wykresy dystrybuant sumy szkód S

Źródło: opracowanie własne.

Do wyznaczenia rozkładu łącznej sumy S dla indywidualnego modelu ryzyka można skorzystać w przypadku struktury zależności opisanej archimedesową funkcją łączącą, z aproksymacji złożonym rozkładem Poissona T [4; 5]. Wykorzystuje się wtedy warunkową niezależność i przeprowadza się aproksymację T_θ według klasycznych metod dla każdej wartości zmiennej losowej Θ indukowanej przez archimedesową funkcję łączącą. Wtedy dystrybuanta tej aproksymacji określona jest mieszanką

$$F_T(x) = \int_0^\infty F_{T_\theta}(x) dF_\Theta(\theta).$$

Można też w tym przypadku oszacować błąd aproksymacji

$$d_{\mathrm{TV}}(S,T)=\sup_A \left|\Pr(S\in A)\right|-\Pr(T\in A).$$

Oszacowanie to jest równe

$$d_{\mathrm{TV}}(S,T)\le n\left(\varphi^{-1}(2\varphi(q))+1-2q\right).$$

6. Zagadnienie ruiny

Na zakończenie naszych rozważań dotyczących rozkładów sum zależnych zmiennych losowych przedstawimy ich wykorzystanie w zagadnieniach związanych z teorią ruiny. Skupimy się na prostym modelu zaproponowanym przez Yuen, Guo i Wu [22]. Rozpatrywali oni model ruiny, w którym szkody należą do dwóch klas. Model ten przybiera następującą postać

$$U(t)=u+ct-\left(\sum_{i=1}^{N_1(t)}X_i+\sum_{i=1}^{N_2(t)}Y_i\right)=u+ct-S(t),$$

gdzie u jest kapitałem początkowym, a c intensywnością napływu składki. Procesy $N_i(t)$, gdzie $i=1,2$, przedstawiające liczebności szkód w poszczególnych klasach, są sumą dwóch procesów: $M_i(t)$ oraz $M_0(t)$, reprezentujących odpowiednio wpływ czynnika wewnętrznego, indywidualnego dla każdej klasy, oraz zewnętrznego, oddziałującego jednocześnie na każdą klasę, tzn.

$$N_i(t)=M_i(t)+M_0(t).$$

Istnienie wspólnego czynnika $M_0(t)$ powoduje zależność procesów $N_1(t)$ i $N_2(t)$. Ponadto przyjmuje się niezależność procesów $M_0(t)$, $M_1(t)$, $M_2(t)$ oraz zmiennych losowych przedstawiających wielkości szkód X_i, Y_i, a także jednakowy rozkład tych zmiennych losowych, których dystrybuanty będziemy oznaczać odpowiednio symbolami F_X oraz F_Y.

Chcąc znaleźć prawdopodobieństwo ruiny $\psi(u)=1-\phi(u)$, gdzie

$$\psi(u)=1-\phi(u)=P\left(U(t)\ge 0,\ \text{dla każdego } t\ge 0\right),$$

autorzy założyli, że $M_i(t)$ są procesami Poissona $\mathrm{Po}(\lambda_i)$ z parametrami λ_i, $M_0(t)$ jest procesem Erlanga $\mathrm{Erl}(2,\lambda_0)$, a rozkłady szkód mają rozkład wykładniczy z wartościami oczekiwanymi równymi odpowiednio μ_X i μ_Y. Rozpatrywali w tym celu równoważny model

$$U'(t) = u + ct - \left(\sum_{i=1}^{M(t)} X_i' + \sum_{i=1}^{M_0(t)} Y_i' \right),$$

gdzie $M(t) = M_1(t) + M_2(t)$, dystrybuanty zmiennych losowych X_i' są mieszankami dystrybuant F_X oraz F_Y:

$$F_{X'}(x) = \frac{\lambda_1}{\lambda_1 + \lambda_2} F_X(x) + \frac{\lambda_2}{\lambda_1 + \lambda_2} F_Y(x),$$

a $X_i' = X_i + Y_i$. Proces $M(t)$ jest już niezależny od $M_0(t)$, można więc stosować klasyczne metody oparte na założeniu niezależności.

Sposób wyznaczania prawdopodobieństwa ruiny dla tak określonego zadania podany jest w [22]. My natomiast skupimy się na analizie wpływu zależności na to prawdopodobieństwo. W tym celu przyjmijmy upraszczające rozpatrywane zagadnienie założenie, że proces $M_0(t)$ przedstawiający wpływ czynników zewnętrznych, jest również procesem Poissona z parametrem λ_0. Wtedy

$$S'(t) = \sum_{i=1}^{M(t)} X_i' + \sum_{i=1}^{M_0(t)} Y_i'$$

ma złożony rozkład Poissona z parametrem $\lambda = \lambda_1 + \lambda_2 + \lambda_0$ [22].

Przykład 5. Przyjmijmy, że szkody z obydwu klas X_i oraz Y_i mają te same rozkłady wykładnicze z wartościami oczekiwanymi $\mu_X = \mu_Y = 6$, a intensywność napływu składki wynosi $c = 60$. Rozpatrzmy trzy przypadki: niezależny, mieszany i współmonotoniczny, charakteryzujące się różną wartością parametrów λ_i opisujących liczebności szkód M_i oraz jednakową wartością oczekiwaną zagregowanych szkód $S(t)$.

a) Niezależność: $\lambda_0 = 0$ oraz $\lambda_1 = \lambda_2 = 4$. Wtedy czynnik zewnętrzny nie wpływa na liczbę szkód, a proces

$$S(t) = \sum_{i=1}^{M_1(t)} X_i + \sum_{i=1}^{M_2(t)} Y_i$$

ma złożony rozkład Poissona $\text{CP}\big(8, \text{W}(6)\big)$. Prawdopodobieństwo ruiny w nieskończonym horyzoncie czasowym jest w tym przypadku równe

$$\psi(u) = 0{,}8 \exp(-0{,}033u).$$

b) Mieszanka: $\lambda_1 = \lambda_2 = \lambda_0 = 2$. Przedstawia jednakowy wpływ wszystkich czynników. Wtedy szkody X' oraz Y' rozpatrywane w równoważnym modelu U'

mają odpowiednio rozkład wykładniczy i Erlanga z dystrybuantami $F(x)$ oraz $G(x)$. Dokładnie, są to rozkłady $X_i' \sim W(6)$ oraz $Y_i' \sim \mathrm{Erl}(2,6)$. Proces zagregowanych szkód $S'(t)$ ma złożony rozkład Poissona CP(8, H), gdzie

$$H(x) = \frac{2}{3}F(x) + \frac{1}{3}G(x),$$

a prawdopodobieństwo ruiny w zależności od wartości kapitału początkowego wynosi

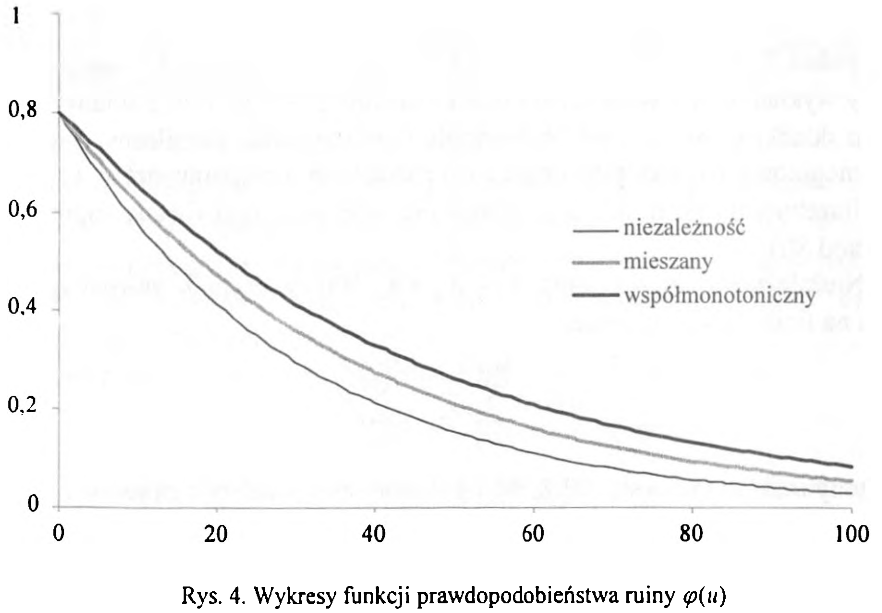
$$\psi(u) = 0{,}8085\exp(-0{,}0269u) - 0{,}0085\exp(-0{,}2064u).$$

c) Współmonotoniczność: $\lambda_0 = 4$ oraz $\lambda_1 = \lambda_2 = 0$. Jest to skrajny przypadek zakładający jedynie wpływ czynnika zewnętrznego. Wtedy zagregowana wielkość szkód

$$S(t) = \sum_{i=1}^{M_0(t)} (X_i + Y_i)$$

ma złożony rozkład Poissona CP(4, G), a prawdopodobieństwo ruiny przybiera postać

$$\psi(u) = 0{,}8221\exp(-0{,}0228u) - 0{,}0221\exp(-0{,}2439u).$$



Rys. 4. Wykresy funkcji prawdopodobieństwa ruiny $\varphi(u)$

Źródło: opracowanie własne.

Na rysunku 4 przedstawione zostały wykresy funkcji prawdopodobieństwa ruiny dla rozpatrywanych trzech przykładów. Widać na nich wzrost prawdopodobieństwa ruiny w miarę zwiększania zależności zachodzącej między liczebnościami szkód poszczególnych klas. Zależność ta spowodowana jest w tym przypadku zwiększoną wartością parametru λ_0.

7. Zakończenie

Praca dotyczy rozkładów sum zależnych zmiennych losowych. Przedstawiono ich zastosowanie w klasycznych modelach aktuarialnych: ryzyka indywidualnego i kolektywnego oraz w teorii ruiny. Prezentowane modele były ilustrowane prostymi przykładami przedstawiającymi wpływ stopnia zależności na badany rozkład sumy. W większości przypadków nie można podać dokładnych rozkładów rozpatrywanych sum, tak że rozkłady te były otrzymane metodami symulacyjnymi. Omawiane modele, zwłaszcza dotyczący zagadnienia ruiny, były jedynie zasygnalizowane.

Literatura

[1] Bowers N.L., Gerber H.U., Hickman J.C., Jones D.A., Nesbitt C.J., *Actuarial Mathematics*, Society of Actuaries, Schaumburg 1986.

[2] Carrière J.F., *Bivariate survival models for coupled lives*, [preprint] University of Alberta 1998.

[3] Chan W.-S., Yang H., Zhang L., *Some results on ruin probabilities in a two-dimensional risk model*, „Insurance: Mathematics and Economics" 2003, no. 32, s. 345-358.

[4] Cossette H., Gaillardetz P., Marceau E., *Common mixture in the individual risk model*, „Mitteilungen der Schweiz, Aktuarvereiningung" 2002, no. 2, s. 131-157.

[5] Cossette H., Marceau E., *The discrete-time risk model with correlated classes of business*, „Insurance: Mathematics and Economics" 2000, no. 26, s. 133-149.

[6] Denneberg D., *Lectures on Non-Additive Measure and Integral*, Kluwer Academic, Boston 1994.

[7] Embrechts P., Lindskog F., McNeil A., *Modelling Dependence with Copulas and Applications to Risk Management*, [preprint], ETH, Zurich 2001.

[8] Frees E.W., Valdez E.A., *Understanding relationships using copulas*, „North American Actuarial Journal" 1998, no. 2, s. 1-25.

[9] Genest G., Marceau E., Mesfioui M., *Compound Poisson approximations for individual models*, „Insurance: Mathematics and Economics" 2003, no. 32, s. 73-91.

[10] Heilpern S., *Analiza zależnego ryzyka ubezpieczeniowego – zastosowanie funkcji łączących*, Prace Naukowe Akademii Ekonomicznej nr 1108, AE, Wrocław 2006, s. 269-285.

[11] Heilpern S., *Funkcje łączące*, AE, Wrocław 2007.

[12] Heilpern S., *Zależny rozkład dwumianowy, zastosowanie w reasekuracji i kredytach*, „Badania Operacyjne i Decyzje" 2007, nr 1, s. 45-61.

[13] Joe H., *Multivariate Models and Dependence Concepts*, Chapman & Hall, London 1997.

[14] Kolev N., Paiva D., *Multinomial model for random sums*, „Insurance: Mathematics and Economics" 2005, no. 37, s. 494-504.

[15] Müller G., *Stop-loss order for portfolios of dependent risks*, „Insurance: Mathematics and Economics" 1997, no. 21, s. 219-223.

[16] Nelsen R.B., *An Introduction to Copulas*, Springer, New York 1999.

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

[18] Pfeifer D., Neslehova J., *Modeling and generating dependent risk processes for IRM and DFA*, „ASTIN Bulletin" 2004, no. 34, s. 333-360.

[19] Rolski T., Schmidli H., Schmidt V., Teugels J., *Stochastic Processes for Insurance and Finance*, Wiley, New York 1999.

[20] Wang S., *Aggregation of correlated risk portfolios: Models and algorithms*, CAS, Proceedings 1998, s. 848-939.

[21] Wei G., Hu T., *Supermodular dependence ordering on a class of multivariate copulas*, „Statistics & Probability Letters" 2002, no. 57, s. 375-385.

[22] Yuen K.C., Guo J.Y., Wu X.Y., *On a correlated aggregate claim model with Poisson and Erlang risk processes*, „Insurance: Mathematics and Economics" 2002, no. 31, s. 205-214.

DISTRIBUTION OF SUMS OF THE DEPENDENT RANDOM VARIABLES, APPLICATION IN THE ACTUARIAL PROBLEMS

Summary

The paper focuses on an analysis of the sums of the dependent random variables. The classical but non-realistic assumption of independence is omitted. The different kinds of dependent structures, mainly connected with copula functions, are studied. The application in some classical actuarial problems: the individual model, the collective model, and the ruin theory are presented. The influence of the degree of dependence on the distribution and properties of these sums of random variables is studied.