

*Stanisław Wanat**

MODELOWANIE STRUKTUR ZALEŻNOŚCI ZA POMOCĄ FUNKCJI POŁĄCZEŃ W ANALIZIE RYZYKA UBEZPIECZYCIELA

Słowa kluczowe: funkcja połączenia, struktura zależności, ryzyko, miara zależności, rozkład wielowymiarowy.

1. Wprowadzenie

Działalność zakładu ubezpieczeń narażona jest na wiele rodzajów ryzyka specyficznego dla rynku ubezpieczeniowego i finansowego oraz systematycznego związanego z funkcjonowaniem gospodarki. W związku z tym, budowa kompleksowego systemu zarządzania nimi wymaga zastosowania rozkładów wielowymiarowych, których jednowymiarowe rozkłady brzegowe są różne lub należą do różnych rodzin. W takich przypadkach nie można wykorzystać np. wielowymiarowych rozkładów normalnych, Studenta, czy też ogólnie wielowymiarowych rozkładów eliptycznych. Obecnie szeroko stosowanym narzędziem, za pomocą którego można konstruować tego typu rozkłady są funkcje połączenia.

Przedmiotem niniejszej pracy jest prezentacja funkcji połączeń, jako narzędzia dającego możliwość modelowania „bogatszych” struktur zależności niż ma to miejsce w przypadku korzystania z korelacji liniowej oraz umożliwiającego konstrukcje rozkładów dwuwymiarowych w sytuacji, gdy znane są rozkłady brzegowe (nie koniecznie takie same) i korelacja między nimi. W pracy podano definicję i podstawowe własności funkcji połączenia. Omówiono jej związek z podstawowymi typami zależności wykorzystywanymi w zarządzaniu ryzykiem tj. współmonotonicznością, przeciwmotonicznością, niezależnością i zależnością liniową oraz z podstawowymi miarami zależności (współczynnikiem korelacji Pearsona, τ -Kendalla, ρ -Spearmana, Blomqvista oraz współczynnikami zależności dolnego i górnego ogona). Następnie scharakteryzowano wybrane rodziny funkcji połączeń a na koniec podano przykład konstrukcji rozkładu dwuwymiarowego z zadanymi rozkładami brzegowymi i zadanym współczynni-

* Katedra Statystyki, Uniwersytet Ekonomiczny w Krakowie.

kiem korelacji τ -Kendalla między nimi. Obliczenia i wykresy wykonano wykorzystując pakiet „R”.

2. Definicja funkcji połączenia

Funkcja połączenia¹ C (ang. *copula*²) jest to niemalejące i prawostronnie ciągłe odwzorowanie k -wymiarowej kostki jednostkowej $[0,1]^k$ w przedział $[0, 1]$, spełniające następujące własności:

$$(1) C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_k) = 0, \text{ dla } i = 1, 2, \dots, k;$$

$$(2) C(1, \dots, 1, u_i, 1, \dots, 1) = u_i, \text{ dla } i = 1, 2, \dots, k;$$

$$(3) \text{ dla wszystkich } \mathbf{a}, \mathbf{b} \in [0, 1]^k \text{ takich, że } a_i < b_i, i = 1, 2, \dots, k \text{ i każdego}$$

$\mathbf{u} \in [0, 1]^k$, $\Delta_{a_1, b_1} \Delta_{a_2, b_2} \dots \Delta_{a_k, b_k} C(\mathbf{u}) \geq 0$, gdzie

$$\Delta_{a_i, b_i} C(\mathbf{u}) = C(u_1, \dots, u_{i-1}, b_i, u_{i+1}, \dots, u_k) - C(u_1, \dots, u_{i-1}, a_i, u_{i+1}, \dots, u_k).$$

Jest to więc dystrybuenta wektora losowego (U_1, \dots, U_n) , dla którego rozkłady brzegowe są jednostajne na przedziale $[0, 1]$ ($U_i \sim \text{Uni}(0,1)$, $i = 1, \dots, k$).

Szybki rozwój metod modelowania zależności stochastycznej wykorzystujących funkcje połączenia i wzrost obszarów ich aplikacji szczególnie w finansach i ubezpieczeniach rozpoczął się na przełomie XX. i XXI w. i nadal trwa. Jednak korzenie tej metody sięgają co najmniej lat czterdziestych ubiegłego wieku i prac H. Eyrauda [1934], w której rozważana jest będąca w istocie funkcją połączenia “fonction de corrélation” oraz W. Hoeffdinga [1940], w której “standardized version of a random pair” też jest funkcją połączenia, przy czym rozkłady jednostajne są w niej określone na przedziale $[-1/2, 1/2]$ zamiast na $[0, 1]$. Zaproponowane w tych pracach pojęcie było później szeroko wykorzystywane przez wielu naukowców (głównie matematyków) w badaniu własności rozkładów wielowymiarowych z ustalonymi rozkładami brzegowymi; analizie miar zależności wyprowadzonych z aksjomatyki A. Rényi’ego [1959] i M. Scarsini’ego [1984]; analizie typów zależności zaproponowanych przez E.L. Lehmana [1966] oraz analizie funkcji zależności dla rozkładów wartości ekstremalnych zaproponowanych przez M. Sibuya [1959].

¹ W literaturze polskiej stosuje się również nazwę *funkcja łącząca* lub *kopuła*.

² Słowa „*copula*” (kopuła) użył po raz pierwszy A. Sklar [1959]. Wywodzi się ono od łacińskiego rzeczownika oznaczającego połączenie, łączenie lub wiązanie rzeczy.

Przełomowe w rozwoju teorii funkcji połączeń okazały się prace A. Sklára [1959] oraz B. Schweizera i A. Sklára [1983]. W drugiej z nich wykorzystano je do badania własności probabilistycznych przestrzeni metrycznych (zaproponowanych przez K. Mengera [1942]). Dała ona podstawy archimedesowym funkcjom połączenia, które pod koniec lat 80. ubiegłego wieku zaczęto powszechnie wykorzystywać w modelach wielowymiarowych. Na przykład A.W Marshall i I. Olkin [1988] zastosowali je w analizie mieszanek modeli, natomiast D. Oakes [1989], pokazał związek tej klasy funkcji połączeń z uogólnionym na dwa wymiary proporcjonalnym modelem hazardu Coxa. Z kolei, szczególną rolę we wspomnianym szybkim rozwoju metod wykorzystujących funkcje połączenia w ubezpieczeniach i finansach odegrały książki [Joe, 1997] i [Nelsen, 1999] oraz m.in. prace [Frees, Valdez, 1998] (w ubezpieczeniach) i [Embrechts, McNeil, Straumann, 2002] (w finansach).

Znaczenie funkcji połączeń w modelowaniu rozkładów wielowymiarowych tłumaczy następujące twierdzenie, na podstawie którego można przedstawić rozkład wielowymiarowy za pomocą rozkładów brzegowych i funkcji połączenia.

Twierdzenie Sklára. Niech C będzie k -wymiarową funkcją połączenia a F_1, \dots, F_k dystrybuantami jednowymiarowych rozkładów prawdopodobieństwa. Wtedy, dla $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$,

$$F(x_1, \dots, x_k) = C(F_1(x_1), \dots, F_k(x_k)) \quad (1)$$

jest dystrybuantą rozkładu wielowymiarowego, dla którego F_1, \dots, F_k są dystrybuantami rozkładów brzegowych. Odwrotnie, dla dowolnej dystrybuanty F łącznego wielowymiarowego rozkładu o dystrybuantach brzegowych F_1, \dots, F_k istnieje funkcja połączenia spełniająca warunek (1). Ponadto, w przypadku, gdy F_1, \dots, F_k są ciągłe istnieje tylko jedna taka funkcja połączenia dana wzorem:

$$C(u_1, \dots, u_k) = F(F_1^{\leftarrow}(u_1), \dots, F_k^{\leftarrow}(u_k)), \quad (u_1, \dots, u_k) \in [0, 1]^k, \quad (2)$$

gdzie $F_i^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F_i(x) \geq q\}$, $0 \leq q \leq 1$, $i = 1, \dots, k$.

Funkcja połączenia, o której jest mowa w drugiej części tego twierdzenia jest określona w sposób jednoznaczny dla dowolnych dystrybuant brzegowych na zbiorze $\text{Ran}F_1 \times \dots \times \text{Ran}F_k$, gdzie $\text{Ran}F_i$ oznacza zbiór wartości dystrybuanty F_i . Dowód tego twierdzenia w przypadku rozkładów dwuwymiarowych

zawierają m.in. prace [Sklar, 1959] i [Nelsen, 1999], natomiast w ogólnym przypadku prace: [Moore, Spruill, 1975] i [Sklar, 1996], [Rüschendorf, 2009].

Na podstawie drugiej części twierdzenia Sklara, funkcją połączenia wektora losowego (X_1, \dots, X_k) , dla którego dystrybuanty brzegowe F_1, \dots, F_k są ciągle będziemy nazywać funkcję połączenia C związaną z dystrybuantą F łącznego rozkładu tego wektora określoną wzorem (2). Jeżeli wektor losowy (X_1, \dots, X_k) posiada ciągle rozkłady brzegowe o dystrybuantach F_1, \dots, F_k i gęstościach f_1, \dots, f_k , to jego funkcję gęstości można przedstawić w postaci:

$$f(x_1, \dots, x_k) = c(F_1(x_1), \dots, F_k(x_k)) \prod_{i=1}^k f_i(x_i), \quad (3)$$

gdzie $c(u_1, \dots, u_k) = \frac{\partial^k C(u_1, \dots, u_k)}{\partial u_1 \dots \partial u_k}$ jest gęstością funkcji połączenia wektora

(X_1, \dots, X_k) . Zatem, przy odpowiednich założeniach, wartość funkcji gęstości $f(x_1, \dots, x_k)$ rozkładu wielowymiarowego może być przedstawiona jako iloczyn składowej $\prod_{i=1}^k f_i(x_i)$ korespondującej z niezależnością zmiennych losowych i składowej $c(F_1(x_1), \dots, F_k(x_k))$, która „burzy” niezależność wprowadzając strukturę zależności. W szczególności, gdy zmienne X_1, \dots, X_k są niezależne, to c jest tożsamościowo równa 1 ($c \equiv 1$) a f jest iloczynem gęstości f_1, \dots, f_k . W tym sensie wartość $c(F_1(x_1), \dots, F_k(x_k)) = f(x_1, \dots, x_k) / \prod_{i=1}^k f_i(x_i)$ wskazuje na „proporcje między zależnością danego rozkładu a niezależnością” w punkcie (x_1, \dots, x_k) . Zatem można stwierdzić, że gęstość funkcji połączenia koduje wszystkie informacje dotyczące zależności między zmiennymi tworzącymi dany wektor losowy, z tego też powodu nazywana jest czasem funkcją zależności i interpretowana jako lokalna miara zależności (por. [Actuarial Theory for...2005]).

Funkcja połączenia dzięki twierdzeniu Sklara jest ważnym i interesującym narzędziem analizy łącznych rozkładów wektorów losowych. Pozwala ona analizując taki rozkład „oddzielić” strukturę zależności od rozkładów brzegowych. Wtedy funkcja połączenia może być rozważana jako składowa rozkładów wielowymiarowych, która modeluje zależność „uzupełniając” zachowanie się każdego z rozkładów brzegowych. Umożliwia to konstruowanie rozkładów wielowymiarowych w dwóch krokach: pierwszym polegającym na wyborze rozkładu

dów brzegowych, drugim na specyfikacji funkcji połączenia. Stanowi to ważne narzędzie, które można wykorzystać w modelowaniu i symulacji.

3. Wybrane własności funkcji połączeń

Naturalne środowisko do badania zależności stwarzają tzw. przestrzenie Fréchéta. Niech F_1, \dots, F_k będą dystrybuantami rozkładów jednowymiarowych. Przestrzenią Fréchéta $\mathbf{F}_k(F_1, \dots, F_k)$ nazywany jest zbiór wszystkich k -wymiarowych wektorów losowych \mathbf{X} (k -wymiarowych dystrybuant F), których rozkładami brzegowymi są F_1, \dots, F_k . Zatem elementami przestrzeni $\mathbf{F}_k(F_1, \dots, F_k)$ są k -wymiarowe wektory losowe o odpowiednio takich samych rozkładach brzegowych, które różnią się tylko strukturą zależności.

Można pokazać (np. Joe[1997]), że jeżeli k -wymiarowy rozkład $F \in \mathbf{F}_k(F_1, \dots, F_k)$, to dla każdego $\mathbf{x} \equiv (x_1, \dots, x_k) \in \mathbb{R}^k$ zachodzą nierówności:

$$M_k(\mathbf{x}) = \max \left\{ \sum_{i=1}^n F_i(x_i) - (k-1), 0 \right\} \leq F(\mathbf{x}) \leq \min_i \{F_i(x_i)\} = W_k(\mathbf{x}), \quad (4)$$

gdzie $M_k(\mathbf{x})$, $W_k(\mathbf{x})$ jest odpowiednio tzw. *dolnym i górnym ograniczeniem Fréchéta*. Należy zaznaczyć, że górne ograniczenie Fréchéta jest zawsze dystrybuantą, czyli należy do przestrzeni $\mathbf{F}_k(F_1, \dots, F_k)$ (por. np. Joe[1997, tw. 3.2]). Natomiast dolne ograniczenie Fréchéta jest dystrybuantą wtedy i tylko wtedy, gdy dla każdego $\mathbf{x} \in \mathbb{R}^k$ takiego, że $0 < F_i(x_i) < 1$, $i = 1, \dots, k$, zachodzi:

$\sum_{i=1}^k F_i(x_i) \leq 1$ lub $\sum_{i=1}^k \bar{F}_i(x_i) \leq 1$ (\bar{F}_i oznacza funkcję przetrwania). Dowód warunkowego koniecznego i wystarczającego jest przedstawiony np. w pracy Joe[1997, tw. 3.7]). Z warunków tych wynika, że dla $n = 2$ dolne ograniczenie Fréchéta jest zawsze dystrybuantą.

Bezpośrednio z definicji przestrzeni Fréchéta wynika, że k -wymiarowe funkcje połączeń tworzą przestrzeń $\mathbf{F}_k(U_1, \dots, U_k)$, czyli dla każdego $(u_1, \dots, u_k) \in [0, 1]^k$ spełniony jest warunek

$$C^-(u_1, \dots, u_k) \leq C(u_1, \dots, u_k) \leq C^+(u_1, \dots, u_k), \quad (5)$$

gdzie dolne i górne ograniczenie mają odpowiednio postać:

$$C^-(u_1, \dots, u_k) = \max \left\{ \sum_{i=1}^k u_i - (k-1), 0 \right\}, \quad C^+(u_1, \dots, u_k) = \min \{u_i\} \quad (6)$$

Do przestrzeni $\mathbf{F}_k(U_1, \dots, U_k)$ należy wykorzystywana w modelowaniu niezależnych zmiennych losowych, tzw. niezależna funkcja połączenia:

$$C^\perp(u_1, \dots, u_n) = \prod_{i=1}^n u_i. \quad (7)$$

Górne ograniczenie C^+ jest zawsze funkcją połączenia, natomiast dolne C^- jest funkcją połączenia tylko dla $k = 2$. Ograniczenia te charakteryzują się następującymi własnościami (zob. [Charpentier, 2008]):

Dolne C^-

1. Dla $k = 2$ ograniczenie C^- jest dystrybuantą rozkładu wektora $(U, 1-U)$.
2. Wektor (X_1, X_2) ma funkcję połączenia C^- wtedy i tylko wtedy, gdy istnieje niemalejąca funkcja h oraz nierosnąca funkcja g , takie, że dla pewnej zmiennej losowej Z spełniony jest warunek $(X_1, X_2) \stackrel{d}{=} (h(Z), g(Z))$.
3. Wektor (X_1, X_2) ma funkcję połączenia C^- wtedy i tylko wtedy, gdy istnieje malejąca funkcja g taka, że $X_2 = g(X_1)$.

Górne C^+

1. Ograniczenie C^+ jest dystrybuantą rozkładu wektora (U, \dots, U) .
2. Wektor (X_1, \dots, X_k) ma funkcję połączenia C^+ wtedy i tylko wtedy, gdy istnieją niemalejące funkcje h_i takie, że dla pewnej zmiennej losowej Z spełniony jest warunek $(X_1, \dots, X_k) \stackrel{d}{=} (h_1(Z), \dots, h_k(Z))$.
3. Wektor (X_1, X_2) ma funkcję połączenia C^+ wtedy i tylko wtedy, gdy istnieje rosnąca funkcja h taka, że $X_2 = h(X_1)$.

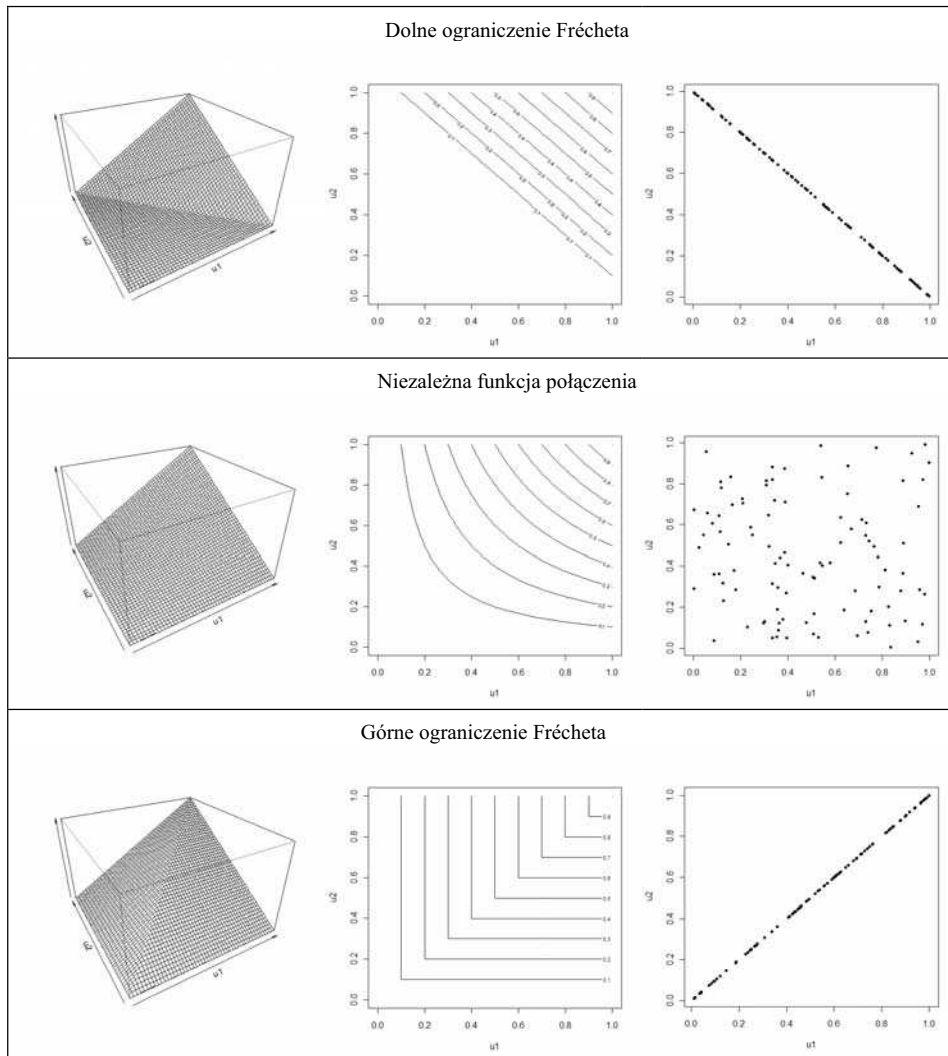
Z przedstawionych własności wynika, że w przypadku dwuwymiarowym C^- jest funkcją połączenia wektora losowego z przeciwmonotoniczną strukturą zależności, natomiast C^+ jest funkcją połączenia wektora ze współmonotoniczną strukturą zależności dla dowolnego (skończonego) wymiaru. Na rys. 1

przedstawiono wykresy C^- , C^+ i C^\perp dla $k = 2$ oraz losowe realizacje dwuwymiarowych wektorów losowych (U_1, U_2) , $U_i \sim Uni(0,1)$, $i = 1, 2$ z tymi funkcjami połączenia.

Niezwykle użyteczną własnością funkcji połączeń jest możliwość określenia postaci funkcji połączenia po dokonaniu ściśle monotonicznych przekształceń rozkładów brzegowych w przypadku, gdy znana jest funkcja połączenia przed transformacją. Własność ta jest wykorzystywana zarówno w procesie modelowania struktur zależności jak również podczas identyfikacji ich konkretnych postaci za pomocą metod nieparametrycznych. W szczególności funkcja połączenia jest niezmiennicza ze względu na ściśle rosnące i ciągłe przekształcenia rozkładów brzegowych. Można pokazać (por. np. [Joe, 1997]), że jeżeli wektor losowy $\mathbf{X} = (X_1, \dots, X_k)$ ma funkcję połączenia C , to tą samą funkcję połączenia ma wektor $(h_1(X_1), \dots, h_k(X_k))$, gdzie funkcje $h_i : \mathbb{R} \rightarrow \mathbb{R}$ są ściśle rosnące i ciągłe odpowiednio na zbiorach wartości zmiennych X_1, \dots, X_k . Przy czym, w przypadku, gdy rozkłady brzegowe są ciągłe wystarczy założyć, że funkcje transformujące h_i są rosnące (por. rozdz. 6 pracy [Schweizer, Sklar, 1983]). Właściwość ta sprawia, że funkcje połączenia są naturalnym narzędziem do badania własności struktur zależności niezmienniczych ze względu na rosnące transformacje rozkładów brzegowych. Z kolei, w pracy [Embrechts, Lindskog, McNeil, 2001] pokazano, że jeżeli wektor losowy (X_1, \dots, X_k) z ciągłymi rozkładami brzegowymi ma funkcję połączenia C_{X_1, \dots, X_k} , a dla dowolnego ustalonego $1 \leq j \leq k$ funkcja h_j jest ściśle malejąca, to zachodzi równość (bez straty ogólności przyjmujemy, że $j = 1$):

$$\begin{aligned} C_{h_1(X_1), \dots, h_k(X_k)}(u_1, \dots, u_k) = \\ = C_{h_2(X_2), \dots, h_k(X_k)}(u_2, \dots, u_k) - C_{X_1, h_2(X_2), \dots, h_k(X_k)}(1 - u_1, u_2, \dots, u_k), \end{aligned} \quad (8)$$

gdzie $C_{h_1(X_1), \dots, h_k(X_k)}$, $C_{h_2(X_2), \dots, h_k(X_k)}$, $C_{X_1, h_2(X_2), \dots, h_k(X_k)}$ oznaczają funkcje połączenia dla wektorów wskazanych w dolnym indeksie. Stosując rekurencyjnie wzór (8) można przedstawić funkcję połączenia $C_{h_1(X_1), \dots, h_k(X_k)}$ za pomocą funkcji połączenia C_{X_1, \dots, X_k} oraz jej rozkładów brzegowych niższego wymiaru.



Rys. 1. Wykresy funkcji połączeń C^- , C^+ i C^\perp oraz przykłady losowych realizacji wektorów losowych, których struktura zależności opisywana jest tymi funkcjami połączenia
 Źródło: opracowanie własne.

4. Funkcje połączenia a wybrane miary zależności

Jednym z głównych narzędzi służących do analizy struktur zależności są miary zależności. Ponieważ, jak to już wcześniej zauważono, funkcje połączenia modelują zależność „uzupełniając” zachowanie się każdego z rozkładów brzegowych, powstaje naturalne pytanie, czy miary zależności można wyznaczyć tylko na podstawie funkcji połączenia struktury zależności, której siłę chcemy zmierzyć. W ogólnym przypadku odpowiedź jest negatywna, gdyż np. miarą, której nie da się w ten sposób wyrazić jest współczynnik korelacji liniowej Pearsona:

$$\rho(X, Y) = \frac{1}{\sqrt{D^2(X)D^2(Y)}} \iint_{[0,1]^2} (C(u_1, u_2) - u_1 u_2) dF_X^{-1}(u_1) dF_Y^{-1}(u_2), \quad (9)$$

gdzie C jest funkcją połączenia łącznego dwuwymiarowego rozkładu zmiennych X i Y . Widać więc, że nie zależy on tylko od funkcji połączenia, ale także od rozkładów brzegowych. Dodatkowo, jak pokazuje A.W. Marshall [1996], jedyną miarą, która mierzyłaby siłę zależności między dowolnymi zmiennymi losowymi X i Y „powiązanymi” taką samą funkcją połączenia i jednocześnie byłaby wyznaczana tylko na podstawie tej funkcji połączenia, jest stała. Takie miary można jednak wskazać ograniczając się do wektorów losowych z ciągłymi rozkładami brzegowymi. Są nimi między innymi:

- współczynnik korelacji τ -Kendalla:

$$\tau(X, Y) = \tau(C) = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1; \quad (10)$$

- współczynnik korelacji ρ -Spearmana:

$$\rho_S(X, Y) = \rho_S(C) = 12 \iint_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 3 = 12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3; \quad (11)$$

- współczynnik skojarzenia Giniego:

$$\gamma(X, Y) = \gamma(C) = 2 \iint_{[0,1]^2} (|u_1 + u_2 - 1| - |u_1 - u_2|) dC(u_1, u_2); \quad (12)$$

- współczynnik korelacji Blomqvista:

$$\beta(X, Y) = \beta(C) = 4C\left(\frac{1}{2}, \frac{1}{2}\right) - 1. \quad (13)$$

W modelowaniu ryzyka w ubezpieczeniach szczególne znaczenie odgrywają także miary kwantyfikujące stopień zależności w ogonach rozkładów zmiennych, czyli opisujące zależność między zdarzeniami ekstremalnymi. Zalicza się do nich m.in. współczynnik zależności górnego ogona (*upper tail dependence parameter*)

$$\lambda_U = \lim_{q \rightarrow 1^-} \left(P\left(Y > F_Y^{\leftarrow}(q) / X > F_X^{\leftarrow}(q)\right) \right), \quad (14)$$

oraz współczynnik zależności dolnego ogona (*lower tail dependence parameter*)

$$\lambda_L = \lim_{q \rightarrow 0^+} \left(P\left(Y \leq F_Y^{\leftarrow}(q) / X \leq F_X^{\leftarrow}(q)\right) \right) \quad (15)$$

łącznego rozkładu dwuwymiarowego wektora (X, Y) , o ile odpowiednie granice istnieją. Mierzą one prawdopodobieństwo wystąpienia zdarzenia ekstremalnego dla ryzyka modelowanego za pomocą zmiennej Y pod warunkiem wystąpienia zdarzenia ekstremalnego dla ryzyka modelowanego zmienną X .

Podobnie jak wyżej, w przypadku ciągłych zmiennych losowych X i Y współczynniki (14) i (15) zależą wyłącznie od funkcji połączenia C łącznego rozkładu wektora (X, Y) :

$$\lambda_U = 2 - \lim_{q \rightarrow 1^-} \frac{1 - C(q, q)}{1 - q}, \quad \lambda_L = \lim_{q \rightarrow 0^+} \frac{C(q, q)}{q}. \quad (16)$$

Ich wartości dla wielu rodzin funkcji połączeń są podane np. w pracach [Joe, 1997] i [Nelsen, 1999].

5. Charakterystyka wybranych funkcji połączeń

Eliptyczne funkcje połączeń

Rodzinę eliptycznych funkcji połączeń otrzymuje się z zależności (2) przyjmując, że F jest dystrybuantą wielowymiarowego standardowego rozkładu eliptycznego. Charakteryzuje się ona wieloma własnościami wielowymiarowego rozkładu normalnego, a jednocześnie można ją wykorzystać w modelowaniu

zależności w ogonach rozkładów. Do rodziny tej należą bardzo często wykorzystywane w praktyce funkcje połączenia Gaussa i Studenta.

Funkcja połączenia Gaussa jest określona następująco:

$$C_{\rho}^G(u_1, \dots, u_k) = H_{\rho}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k)), \quad (17)$$

gdzie H_{ρ} jest dystrybuantą wielowymiarowego standardowego rozkładu normalnego o macierzy korelacji ρ . Jej gęstość ma postać:

$$c_{\rho}^G(u_1, \dots, u_k) = \frac{1}{\sqrt{|\rho|}} \exp\left(-\frac{1}{2} \boldsymbol{\zeta}^T (\rho^{-1} - \mathbf{I}) \boldsymbol{\zeta}\right), \quad (18)$$

gdzie $\boldsymbol{\zeta} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k))^T$, natomiast Φ jest dystrybuantą standardowego rozkładu normalnego a \mathbf{I} macierzą jednostkową.

Jeżeli za pomocą funkcji połączenia Gaussa modelujemy strukturę zależności k rodzajów ryzyka opisywanych wektorem (X_1, \dots, X_k) , to wówczas współczynniki korelacji τ -Kendalla i ρ -Spearmana między zmiennymi X_i, X_j wynoszą:

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(\rho_{ij}), \quad \rho_s(X_i, X_j) = \frac{6}{\pi} \arcsin\left(\frac{\rho_{ij}}{2}\right). \quad (19)$$

Gdy zmienne brzegowe są normalne, otrzymujemy wielowymiarowy rozkład normalny, natomiast, gdy warunek ten nie jest spełniony rozkład metagausowski (*meta-Gaussian*). W przypadku dwuwymiarowym $C_{-1}^G = C^-$ (czyli, można modelować ryzyka z przeciwnotoniczną strukturą zależności), $C_0^G = C^\perp$ (można modelować ryzyka niezależne) oraz $C_1^G = C^+$ (można modelować ryzyka z współnotoniczną strukturą zależności). Natomiast współczynniki zależności w ogonach rozkładu z funkcją połączenia C_{ρ}^G wynoszą:

$$\lambda_L = \lambda_U = \begin{cases} 0, & \rho < 0 \\ 1, & \rho = 0 \end{cases}. \quad (20)$$

W podobny sposób, jak w przypadku gaussowskiej funkcji połączenia, określa się funkcję połączenia Studenta:

$$C_{\rho, \nu}^S(u_1, \dots, u_n) = T_{\rho, \nu}(T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_n)), \quad (21)$$

gdzie $T_{\rho, \nu}$ jest dystrybuantą wielowymiarowego standaryzowanego rozkładu Studenta o ν stopniach swobody i macierzy korelacji ρ , a T_ν jest dystrybuantą jednowymiarowego rozkładu Studenta o ν stopniach swobody. Jej gęstość ma postać:

$$c_{\rho, \nu}^S(u_1, \dots, u_n) = |\rho|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\nu+n}{2}\right) \left(\Gamma\left(\frac{\nu}{2}\right)\right)^n \left(1 + \frac{1}{\nu} \boldsymbol{\zeta}^T \rho^{-1} \boldsymbol{\zeta}\right)^{-\frac{\nu+n}{2}}}{\left(\Gamma\left(\frac{\nu+1}{2}\right)\right)^n \Gamma\left(\frac{\nu}{2}\right) \prod_{i=1}^n \left(1 + \frac{\zeta_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}}, \quad (22)$$

gdzie $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_n) = (T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_n))$.

Jeżeli wektor (X_1, \dots, X_k) ma funkcję połączenia $C_{\rho, \nu}^S$, to współczynniki korelacji τ -Kendalla i ρ -Spearmana między zmiennymi X_i, X_j są określone wzorem (19). Gdy zmienne X_1, \dots, X_k mają rozkłady Studenta z ν stopniami swobody, to wektor ten ma wielowymiarowy rozkład Studenta o ν stopniach swobody, w przeciwnym przypadku - rozkład meta-Studenta (*meta-Student*). Jeżeli liczba stopni swobody zmierza do nieskończoności, to funkcja połączenia Studenta zmierza do funkcji połączenia Gaussa. W przypadku dwuwymiarowym $C_{-1, \nu}^S = C^-$ i $C_{1, \nu}^S = C^+$, natomiast $C_{0, \nu}^S \neq C^+$. Z kolei, współczynniki zależności w ogonach rozkładu z funkcją połączenia $C_{\rho, \nu}^S$ są równe:

$$\lambda_L = \lambda_U = 2T_{\nu+1} \left(-\sqrt{\nu+1} \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right). \quad (23)$$

Wykorzystanie omówionej rodziny eliptycznych funkcji połączeń i związanych z nimi rozkładów eliptycznych w zagadnieniach aktuarialnych jest szerzej dyskutowane m.in. w pracach [Landsman, 2009], [Abdous, Genest, Rémillard, 2005].

Archimedesowe funkcje połączeń

Jedną z najważniejszych klas funkcji połączeń stanowią tzw. *połączenia Archimedeses* (*Archimedean copulas*). Zawiera ona szereg rodzin funkcji połączeń, które są stosunkowo proste w konstrukcji i nadto posiadają ważne z praktycznego punktu widzenia własności. Szczegółowy opis tej klasy zawierają m.in. następujące prace: [Genest, MacKay, 1986], [Marshall, Olkin, 1988], [Joe, 1997], [Nelsen, 1999], natomiast ich wykorzystanie w zagadnieniach aktuarialnych przedstawiono m.in. w: [Frees, Valdez, 1998], [Klugman, Parsa, 1999].

Funkcja C jest połączeniem Archimedeses, jeżeli można ją przedstawić w następujący sposób:

$$C(u_1, \dots, u_k) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_k)), \quad (24)$$

gdzie $0 \leq u_i \leq 1$, $i = 1, \dots, k$, natomiast tzw. *generator*³ ϕ jest funkcją ciągłą, ściśle malejącą i wypukłą odwzorowującą przedział $[0, 1]$ w przedział $[0, \infty]$, taką, że $\phi(1) = 0$ i $\phi(0) = \infty$ oraz pochodne rzędu n funkcji do niej odwrotnej spełniają warunek⁴:

$$(-1)^n \frac{d^n}{dx^n} \phi^{-1}(x) \geq 0, \quad \text{dla } n = 1, 2, 3, \dots \quad (25)$$

Znaczenie tej klasy, w głównej mierze wynika z faktu, że wszystkie informacje dotyczące wielowymiarowej struktury zależności wektorów losowych modelowanych za ich pomocą są zawarte w jednowymiarowym generatorze. W szczególności, w przypadku dwuwymiarowego wektora losowego (X, Y) z rozkładem opisywanym archimedesową funkcją połączenia z generatorem ϕ , współczynnik korelacji τ -Kendalla można wyrazić w postaci ([Genest, MacKay, 1986], [Nelsen, 1999]):

³ Dokładniej jest to tzw. silny generator. Gdy $\phi(0) \neq \infty$ generator nazywany jest słabym i wówczas w przypadku dwuwymiarowym funkcję połączenia Archimedeses określa się w następujący sposób: $C(u_1, u_2) = \phi^{[-1]}(\phi(u_1) + \phi(u_2))$, gdzie $\phi^{[-1]}$ jest tzw.

pseudoodwrotnością generatora ϕ zdefiniowaną jako $\phi^{[-1]}(u) = \begin{cases} \phi^{-1}(u), & 0 \leq u \leq \phi(0) \\ 0, & \phi(0) \leq u \leq \infty \end{cases}$.

W przypadku dwuwymiarowym generator nie musi spełniać warunku (25).

⁴ Funkcję spełniającą ten warunek (tzn. (25)) nazywa się całkowicie monotoniczną.

$$\tau(X, Y) = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi'(u)} du, \quad (26)$$

podobnie współczynniki zależności w ogonach rozkładu tego wektora przyjmują postać ([Joe, 1997]):

$$\lambda_L = 2 \lim_{t \rightarrow \infty} \frac{(\phi^{-1})'(2t)}{(\phi^{-1})'\phi(t)}, \quad \lambda_U = 2 - 2 \lim_{t \rightarrow 0} \frac{(\phi^{-1})'(2t)}{(\phi^{-1})'\phi(t)}. \quad (27)$$

W tab. 1 przedstawiono podstawowe własności trzech następujących rodzin funkcji połączenia Archimedesów (dla przypadku dwuwymiarowego):

- *Cook-Johnson* ([Cook, Johnson, 1981]), *Claytona*⁵ (Clayton, 1978)

$\{C_\theta^C\}_{\theta > 0}$:

$$C(u_1, \dots, u_k) = \left(\sum_{i=1}^k u_i^{-\theta} - k + 1 \right)^{-\frac{1}{\theta}}, \quad \theta > 0; \quad (28)$$

- *Gumbel-Hougaard* ([Gumbel, 1960]) $\{C_\theta^{GH}\}_{\theta \geq 1}$:

$$C(u_1, \dots, u_k) = \exp \left[- \left(\sum_{i=1}^k (-\ln u_i)^\theta \right)^{\frac{1}{\theta}} \right], \quad \theta \geq 1; \quad (29)$$

- *Franka* ([Frank, 1979]) $\{C_\theta^K\}_{\theta \neq 0}$:

$$C(u_1, \dots, u_k) = -\frac{1}{\theta} \ln \left[1 + \frac{\prod_{i=1}^k (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{k-1}} \right], \quad \theta > 0. \quad (30)$$

⁵ W przypadku dwuwymiarowym jest znana jako rodzina *Claytona*.

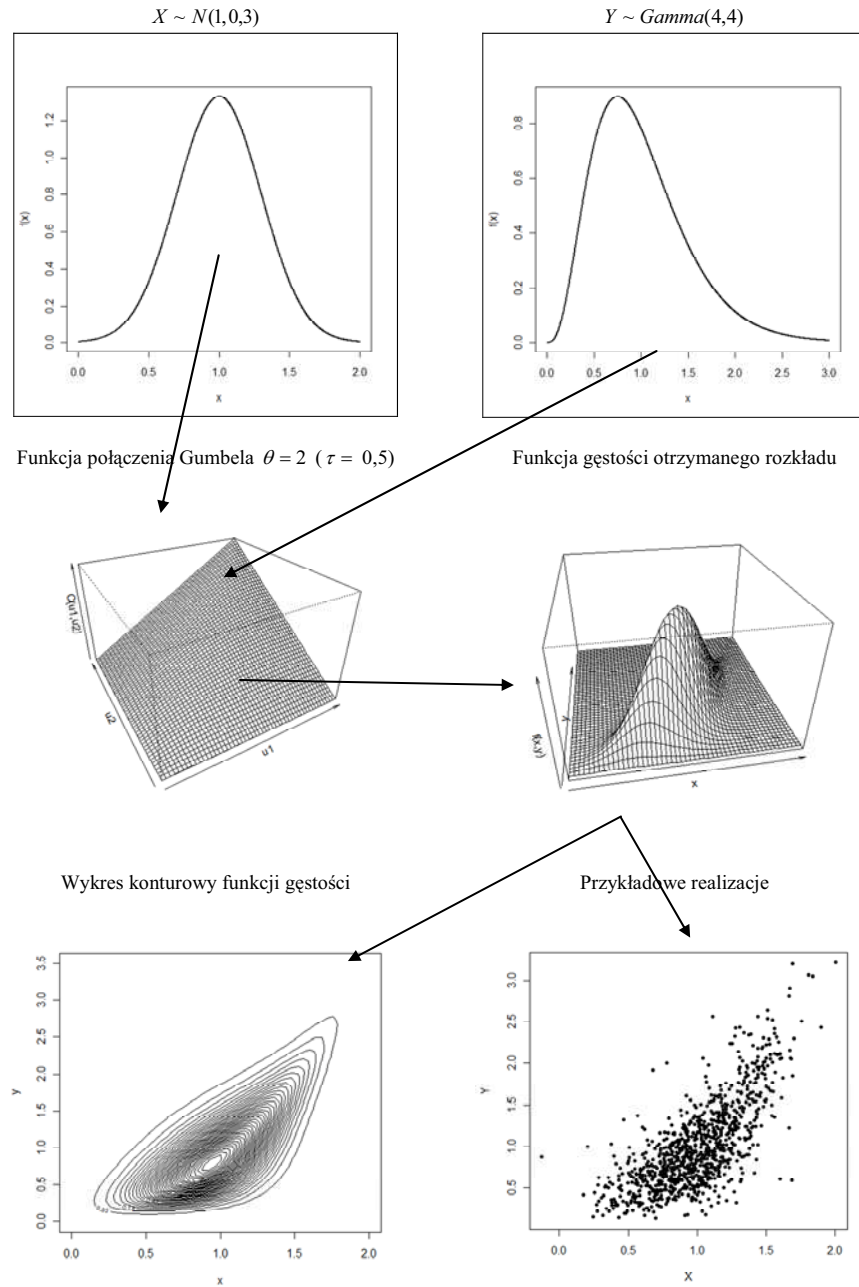
Tab. 1. Charakterystyka wybranych rodzin funkcji połączenia Archimedesesa

	<i>Claytona</i>	<i>Gumbela-Hougaard</i>	<i>Franka</i>
C (dla dwóch wymiarów)	$\max\left(\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{\frac{1}{\theta}}, 0\right)$	$\exp\left(-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta\right)^{\frac{1}{\theta}}\right)$	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$
Związek parametru z współczynnikiem korelacji τ -Kendalla	$\theta = \frac{2\tau}{1-\tau}, \theta \geq -1, \theta \neq 0$ $\tau = \frac{\theta}{\theta+2}, -1 \leq \tau \leq 1$	$\theta = \frac{1}{1-\tau}, \theta \geq 1$ $\tau = \frac{\theta-1}{\theta}, 0 \leq \tau \leq 1$	$\tau = 1 - \frac{4}{\theta} + \frac{4D_1(\theta)}{\theta},$ $-1 \leq \tau \leq 1 \quad \theta \neq 0, \tau \neq 0$ $D_1(\theta) = \int_0^{\theta} \frac{x}{\theta(e^x - 1)} dx$
Generator	$\phi(t) = \frac{t^{-\theta} - 1}{\theta},$ dla $\theta \geq 0$ silny i całkowicie monotoniczny	$\phi(t) = (-\ln t)^\theta,$ silny całkowicie monotoniczny	$\phi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$ silny całkowicie monotoniczny
Współczynniki zależności w ogonach	dla $\theta \geq 0$ $\lambda_L = 2 - \frac{1}{\theta}$ $\lambda_U = 0,$	$\lambda_L = 0$ $\lambda_U = 2 - 2^{\frac{1}{\theta}}$	$\lambda_L = 0$ $\lambda_U = 0$
Przypadki graniczne	$C_{-1} = C^-, C_0 = C^\perp,$ $C_\infty = C^+$	$C_1 = C^\perp, C_\infty = C^+$	$C_{-\infty} = C^-, C_0 = C^\perp,$ $C_\infty = C^+$

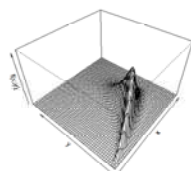
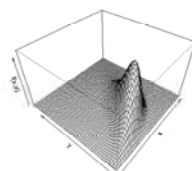
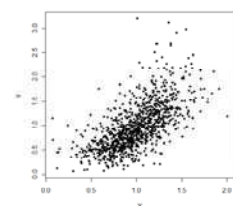
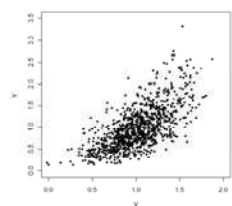
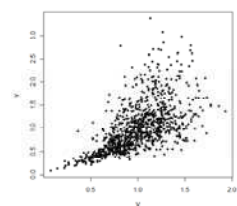
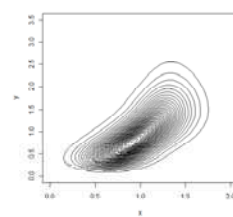
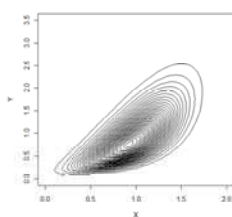
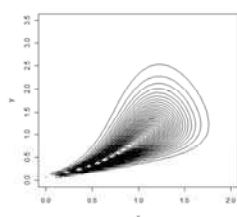
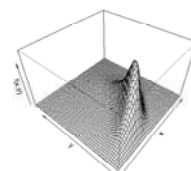
Źródło: Opracowanie własne.

6. Przykład konstrukcji rozkładu dwuwymiarowego z zadaną strukturą zależności

Opisane własności funkcji połączeń umożliwiają dowolne rozkłady jednowymiarowe „powiązać ze sobą” tą samą funkcją połączenia lub ustalone rozkłady jednowymiarowe „powiązać ze sobą” różnymi funkcjami połączenia. Ta druga możliwość jest szczególnie istotna z punktu widzenia modelowania wielowymiarowych zmiennych ryzyka, kiedy to najczęściej znane są rozkłady brzegowe a zależność między nimi jest albo nieznaną, albo tylko częściowo znaną. Metodę konstrukcji rozkładu dwuwymiarowego, w której wykorzystuje się to podejście, w przypadku gdy jedna zmienna jednowymiarowa ma rozkład normalny ($(X \sim N(1,0,3))$) a druga gamma ($(Y \sim Gamma(4,4))$), natomiast struktura zależności opisywana jest funkcją połączenia Gumbela z parametrem $\theta = 2$, schematycznie przedstawiono na rys. 2. Założona postać funkcji połączenia oznacza, że współczynnik korelacji τ -Kendalla między rozważanymi zmiennymi losowymi wynosi $\tau(X, Y) = 0,5$. Oczywiście wykorzystując różne funkcje połączenia można otrzymać różne rozkłady dwuwymiarowe z zadanymi rozkładami brzegowymi i ustalonym współczynnikiem korelacji τ -Kendalla (por. rys. 3).



Rys. 2. Schemat konstrukcji rozkładu dwuwymiarowego z zadaną strukturą zależności
 Źródło: Opracowanie własne.

a) Funkcja połączenia *Claytona*
 $\theta = 2,00$ a) Funkcja połączenia *Gausa*
 $\rho = 0,71$ a) Funkcja połączenia *Franka*
 $\theta = 5.74$ 

Rys. 3. Przykładowe rozkłady dwuwymiarowe dla zadanych rozkładów brzegowych ($X \sim N(1, 0,3)$, $Y \sim Gamma(4,4)$) i ustalonego współczynnika korelacji $\tau(X, Y) = 0,5$ otrzymane za pomocą wybranych funkcji połączenia

Źródło: Opracowanie własne.

7. Podsumowanie

Wykorzystanie funkcji połączeń w modelowaniu ryzyka w ubezpieczeniach ma swoje zalety i wady. Bez wątpienia ich zaletą jest możliwość modelowania „bogatszych” struktur zależności, niż ma to miejsce w przypadku korzystania z korelacji liniowej. Z kolei, możliwość „oddzielenia” za ich pomocą struktury zależności od rozkładów brzegowych sprawia, że są wygodnym narzędziem symulacji rozkładów wielowymiarowych zmiennych ryzyka, niezbędnych w analizie scenariuszy i testach stresu. Za ich pomocą można modelować zależności w ogonach rozkładów i zjawiska ekstremalne. Dają się także w stosunkowo prosty sposób łączyć z innymi modelami, co ma szczególne znaczenie w zintegrowanych systemach zarządzania ryzykiem. Jednak, jak z każdą meto-

da, z wykorzystaniem funkcji połączeń związane są także pewne wady. Do głównych można zaliczyć wymaganie dość dużej liczby danych empirycznych niezbędnych do oszacowania właściwej funkcji połączenia, jak również pewne trudności przy rozstrzygnięciu, którą z funkcji połączeń należy wybrać do opisu rozkładu wielowymiarowego. Pomimo tych wad, funkcje połączenia wydają się wygodnym narzędziem wprowadzania struktur zależności do wielowymiarowych modeli ryzyka w ubezpieczeniach. Świadczy o tym chociażby ich szeroki obszar zastosowań, który obejmuje m.in. modelowanie rozkładu łącznych szkód w portfelach ubezpieczeń, modelowanie zależności między wysokością szkód a związaną z nimi wysokością kosztów, szacowanie efektu dywersyfikacji, szacowanie ryzyka niewywiązania się reasekuratora ze zobowiązań, alokację kapitału ekonomicznego, modelowanie ryzyka katastroficznego. Wykorzystuje się je także w zintegrowanych systemach zarządzania ryzykiem (np. w systemie *ERM - Enterprise Risk Management*), w dynamicznej analizie finansowej (*DFA - Dynamic Financial Analysis*) oraz w modelowaniu wypłacalności ubezpieczyciela (np. w modelach wewnętrznych w „*Solvency II*”, szwajcarskim teście wypłacalności).

Literatura

- Abdous B., Genest C., Rémillard B., 2005, *Dependence properties of meta-elliptical distribution*, [w:] P. Duchesne, B. Rémillard (red.), *Statistical Modeling and Analysis for Complex Data Problems*, GERAD 25th Anniversary Series. Springer, New York, s. 1–15.
- Actuarial Theory for Dependent Risks. Measures, Orders and Models, 2005, Denuit M., Dhaene J., Goovaerts M., Kaas R., John Wiley & Sons, Ltd.
- Charpentier A., 2008, *Dependence structures and limiting results : with applications in finance and insurance*, VDM Verlag Dr. Müller, Saarbrücken, Germany.
- Clayton D.G., 1978, A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, “*Biometrika*”, vol. 65, s. 141–151.
- Cook R.D., Johnson M.E., 1981, A Family of Distributions for Modeling Non-Elliptically Symmetric Multivariate Data, “*Journal of the Royal Statistical Society B*”, vol. 43, s. 210–218.
- Embrechts P., McNeil A., Straumann D., 2002, *Correlation and Dependence in Risk Management: Properties and Pitfalls*, [w:] M. Dempster, H. Moffat (red.), *Risk Management: Value at Risk Beyond*, Cambridge University Press.
- Embrechts P., Lidskog F., McNeil A., 2001, *Modelling dependence with copulas and applications to risk management*, report, ETHZ Zurich, <http://www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf>, (pobrano: 07.04.2007)
- Eyraud H., 1934, *Sur une représentation nouvelle des corrélations continues*, „*Comptes Rendus de l'Académie des Sciences de Paris*”, vol. 199, s. 1356–1358.
- Frank M.J., 1979, *On the Simultaneous Associativity of $F(x,y)$ and $x+y-F(x,y)$* , “*Aequationes Mathematicae*”, vol. 19, s. 194–226.
- Frees E.W., Valdez E.A., 1998, Understanding Relationships Using Copulas, “*North American Actuarial Journal*”, vol. 2, s. 1–25.
- Genest C., MacKay R.J., 1986, The Joy of copulas: Bivariate distributions with uniform marginals, “*American Statistician*”, vol. 40, s. 280–283.

- Gumbel E. J., 1960, Distributions des Valeurs Extremes en Plusieurs Dimensions, "Publications de l'Institut de Statistique de l'Université de Paris", vol. 9, s. 171–173.
- Hoeffding W., 1940, Masstabinvariante Korrelationstheorie, „Schriften des Mathematischen Instituts für Angewandte Mathematik der Universität Berlin“, vol. 5, 181–233, przedruk jako "Scale-Invariant correlation theory" [w:] "The collected works of Wassily Hoeffding" (1994), Fisher and Sen eds., Springer Verlag, s. 57–107.
- Joe H., 1997, Multivariate models and dependence concepts, Chapman-Hall.
- Klugman S., Parsa R., 1999, Fitting bivariate loss distributions with copulas, "Insurance: Mathematics & Economics", vol. 24, s. 139–148.
- Landsman Z. , 2009, Elliptical families and copulas: tilting and premium; capital allocation, "Scandinavian Actuarial Journal", vol. 2009/2, s.85–103, 2009
- Lehmann E.L. , 1966, Some Concepts of Dependence, "Annals of Mathematical Statistics", vol. 37, nr 5, s. 1137–1153.
- Marshall A.W., 1996. Copulas, marginals, and joint distributions, [w:], L. Rüschendorf, B. Schweizer, M.D. Taylor (red.), Distributions with Fixed Marginals and Related Topics, IMS Lectures Notes – Monograph Series 28. Institute of Mathematical Statistics, Hayward, CA, s. 213–222.
- Marshall A.W, Olkin I., 1988, Families of multivariate distributions, "Journal of the American Statistical Association", vol. 83, s. 834–841.
- Menger K. , 1942, "Statistical metrics", "Proceedings of the National Academy of Sciences of the United States of America", vol. 28, s. 535–537.
- Moore D.S., Spruill M.C., 1975, Unified large-sample theory of general chi-squared statistics for tests of fit, "The Annals of Statistics", vol. 3, s. 599–616.
- Nelsen R.B., 1999, An Introduction to Copulas, Springer-Verlag, New York.
- Oakes D., 1989, Bivariate survival models induced by frailties, "Journal of the American Statistical Association", vol. 84, s. 487–493.
- Rényi A., 1959, On measures of dependence, "Acta Mathematica Academiae Scientiarum Hungaricae", vol. 10, s. 441–451.
- Rüschendorf L., 2009, On the distributional transform, Sklar's Theorem, and the empirical copula process, "Journal of Statistical Planning and Inference", vol. 139, s. 3921–3927.
- Scarsini M., 1984, On measures of concordance, "Stochastica", vol. 8, s. 201–218.
- Sibuya M., 1959, Bivariate extreme statistics I. Annals of the Institute of Statistical Mathematics, vol. 11, s. 195–210.
- Sklar A., 1959, Fonctions de ré partition à n dimensions et leurs marges, „Publications de l'Institut de Statistique de l'Université de Paris", vol. 8, s. 229–231.
- Sklar A., 1996, Random variables, distribution functions, and copulas—a personal look backward and forward, [w:] L. Ruschendorf , B. Schweizer, M.D. Taylor (red) Distributions with Fixed Marginals and Related Topics, Institute of Mathematical Statistics, Hayward, CA, s. 1–14.
- Schweizer B., Sklar A., 1983, Probabilistic Metric Spaces. North-Holland/Elsevier, New York.

Abstract

This paper aims to present copulas, as a modeling tool which will give the 'richer' dependency structures than the use of linear correlation. The paper presents definitions and basic properties of the copula function. Discusses its relationship with the basic types of dependencies used in risk management i.e. comonotonicity, countermonotonicity, independence and linear dependence and the basic measures of dependence (Pearson's correlation, Kendall's rank correlation, Spearman's rank correlation, Blomqvist's beta, upper (lower) tail dependence parameter). Then selected family of copulas have been characterized and is an example of construction of two-dimensional distribution, where the marginal distributions are known and the Kendall's rank correlation between them. Calculations and graphs were performed using the package "R".