

Stanisław Heilpern

Akademia Ekonomiczna we Wrocławiu

ELIPTYCZNE FUNKCJE ŁĄCZĄCE

1. Wstęp

Funkcje łączące są ważnym narzędziem umożliwiającym badanie zależności między zmiennymi losowych. Są łącznikiem między rozkładem łącznym a rozkładami brzegowymi. Ostatnio można spotkać coraz więcej prac poświęconych badaniom zależności. W ich zastosowaniach, np. w ubezpieczeniach czy finansach, często odchodzi się od klasycznych modeli opartych na niezależności rozpatrywanych zmiennych losowych, zakładając ich zależność.

Przedstawiona praca poświęcona jest eliptycznym funkcjom łączącym, często używanym w praktyce rodzinom funkcji łączących generowanych przez rozkłady eliptyczne. Ma ona charakter przeglądu i można ją traktować jako kontynuację referatu autora z poprzedniej XLI Konferencji Statystyków, Ekonometryków i Matematyków Akademii Ekonomicznych Polski Południowej, która odbyła się w Osieczanach w 2005 r. [Frees, Voldez 1998]. W pracy przedstawiono najważniejsze własności eliptycznych funkcji łączących, podkreślając własności wyróżniające je spośród innych rodzin funkcji łączących, głównie archimedesowych. Omówiono podstawowe rodzaje rozkładów eliptycznych, eliptycznych funkcji łączących, ich identyfikację, zagadnienie symulacji oraz kierunki ich uogólnienia.

2. Funkcje łączące

Funkcje łączące (*copula*) [Schweitzer, Sklar 1999; Nelsen 1999] można zdefiniować jako n -wymiarowe dystrybuanty skupione na $[0, 1]^n$ o jednostajnych na $[0, 1]$ dystrybuantach brzegowych.

Podstawowym twierdzeniem teorii funkcji łączących jest twierdzenie Sklara mówiące, że dla n -wymiarowej dystrybuanty F z dystrybuantami brzegowymi $F_1, \dots, F_n$ istnieje funkcja łącząca C spełniająca warunek

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (1)$$

dla każdych $x_1, \dots, x_n$. Ponadto dla każdej n -wymiarowej funkcji łączącej C i jedno-wymiarowych dystrybuant $F_1, \dots, F_n$ funkcja F określona wzorem (1) jest n -wymiarową dystrybuantą, a $F_1, \dots, F_n$ jej dystrybuantami brzegowymi.

Funkcję łączącą można więc traktować jako łącznik między rozkładami brzegowymi a rozkładem łącznym. Jest ona niezależna od rozkładów brzegowych tak, że doskonale nadaje się do modelowania zależności zachodzących między zmiennymi losowymi, wydobywając najważniejsze cechy charakteryzujące dany typ zależności. Analiza tej zależności podzielona jest na dwie składowe: na badanie rozkładów brzegowych oraz na analizę funkcji łączącej. Mówimy wtedy, że funkcja łącząca wyznacza strukturę zależności.

W przypadku ciągłych dystrybuant brzegowych $F_1, \dots, F_n$ funkcja łącząca C jest jednoznacznie określona wzorem

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)).$$

W przeciwnym wypadku jest ona jednoznacznie wyznaczona na iloczynie kartezyjskim zbiorów wartości dystrybuant brzegowych.

Do najważniejszych własności funkcji łączących można zaliczyć to, że jest ona niezmiennicza ze względu na rosnące przekształcenia zmiennych brzegowych.

3. Miary zależności

Funkcje łączące opisują zależność zachodzącą między zmiennymi losowymi w sposób globalny jako funkcje. Natomiast współczynniki korelacji charakteryzują tę zależność w sposób punktowy, podając pojedynczą wartość. Do najbardziej znanych i powszechnie stosowanych współczynników korelacji należy klasyczny współczynnik Pearsona. Jednakże nie jest on związany z funkcją łączącą, ponieważ zależy istotnie od rosnących przekształceń zmiennych brzegowych. Opisuje on jedynie zależności liniowe i jest niezmienniczy ze względu na tego typu przekształcenia [Embrechts, Lindskog, McNeil 2001].

Inaczej wygląda sytuacja w przypadku rangowych współczynników korelacji Kendalla τ i Spearmana ρ . Są one niezmiennicze ze względu na rosnące przekształcenia zmiennych brzegowych i zależą jedynie od funkcji łączących. Współczynniki te można przedstawić wzorami:

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad \text{ i } \quad \rho(X, Y) = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3,$$

gdzie C jest funkcją łączącą opisującą zależność zachodzącą między zmiennymi losowymi X oraz Y .

Do badania zależności zachodzących między ekstremalnymi wartościami zmiennych losowych służy *współczynnik zależności wartości ekstremalnych (tail dependence coefficient)*. Górny współczynnik podany jest wzorem [Joe 1997]:

$$\lambda_U(X, Y) = \lim_{u \rightarrow 1} P(Y > F_2^{-1}(u) | X > F_1^{-1}(u)),$$

a dolny

$$\lambda_L(X, Y) = \lim_{u \rightarrow 0} P(Y \leq F_2^{-1}(u) | X \leq F_1^{-1}(u)).$$

Są to granice prawdopodobieństw warunkowych opartych na kwantylach, gdzie prawdopodobieństwa tworzące kwantyle dążą do jedynki lub zera. Współczynniki zależności wartości ekstremalnych odgrywają dużą rolę w finansach i ubezpieczeniach. Na przykład jednocześnie występowanie bardzo dużych szkód różnego typu, tzw. katastrofalnych, ma duży wpływ na działalność firmy ubezpieczeniowej. Ich kumulacja może spowodować poważne zakłócenia w jej funkcjonowaniu.

Współczynniki te zależą jedynie od funkcji łączących opisujących strukturę zależności zachodzącą między badanymi zmiennymi losowymi. Zachodzą bowiem zależności [Joe 1997; Embrechts, Lindskog, McNeil 2001]

$$\lambda_U(X, Y) = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \quad \text{ i } \quad \lambda_L(X, Y) = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}.$$

Jeśli współczynniki zależności wartości ekstremalnych są równe zeru, czyli $\lambda_U(X, Y) = 0$ lub $\lambda_L(X, Y) = 0$, to mówimy, że zachodzi asymptotyczna niezależność górna lub dolna, tzn. niezależność między ekstremalnymi wartościami zmiennych losowych.

4. Archimedesowe funkcje łączące

Najczęściej wykorzystywanymi w praktyce funkcjami łączącymi są *archimedesowe funkcje łączące*. Funkcje te określone są wzorem

$$C(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \dots + \varphi(u_n)),$$

gdzie $\varphi: [0, 1] \rightarrow [0, \infty]$ jest ciągłą, wypukłą, malejącą funkcją, nazywana *generatorem*, spełniającą warunek $\varphi(1) = 0$.

W zastosowaniach wykorzystuje się zwykle rodziny archimedesowych funkcji łączących indeksowane parametrem. Zaletą archimedesowych funkcji łączących jest ich prostota. Umożliwiają one addytywną separację zmiennych, co znacznie ułatwia modelowanie zależności. Jednowymiarowa funkcja, jaką jest generator, jednoznacznie wyznacza nam wielowymiarową funkcję łączącą.

Jednakże funkcje te mają poważne wady, w dużym stopniu ograniczające ich potencjalne zastosowanie. Przede wszystkim, w przypadku wielowymiarowym, dla $n > 2$, gdy struktura zależności opisana jest archimedesową funkcją łączącą, współczynniki korelacji Kendalla i Spearmana przyjmują dla każdej pary brzegowych zmiennych losowych te same wartości. Ponadto wartości tych współczynników są dodatnie. Innymi słowy, brzegowe zmienne losowe są jednakowo i dodatnio skore-

lowane. Jest to poważne ograniczenie, istotnie zmniejszające pole zastosowania tych funkcji łączących.

Dla każdej archimedesowej funkcji łączącej istnieje zmienna losowa taka, że brzegowe zmienne losowe są warunkowo niezależne dla ustalonej wartości tej zmiennej. Jest to tzw. model słabości (*frailty*) [Frees, Valdez 1998; Wong 1999, 11]. Warunkowa niezależność również mocno ogranicza zastosowania tego typu funkcji łączących.

5. Wielowymiarowe rozkłady eliptyczne

Wielowymiarowe rozkłady eliptyczne możemy określić na wiele sposobów, podając ich funkcje charakterystyczną, gęstość bądź ich reprezentację stochastyczną. Niech $\mu \in \mathbb{R}^n$, Σ będzie symetryczną, dodatnio określoną macierzą wymiaru $n \times n$, a $\phi: \mathbb{R}_+ \rightarrow \mathbb{R}$. Mówimy, że wektor losowy $\mathbf{X} = (X_1, \dots, X_n)$ ma n -wymiarowy rozkład eliptyczny, jeśli jego funkcja charakterystyczna przyjmuje postać

$$\varphi_{\mathbf{X}-\mu}(\mathbf{t}) = \phi(0,5\mathbf{t}^T \Sigma \mathbf{t}).$$

Stosujemy wtedy oznaczenie $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$.

Wektor μ przedstawiający wektor wartości oczekiwanych zmiennej $\mathbf{X}$ jest określony jednoznacznie, a pozostałe parametry rozkładu: macierz Σ oraz funkcja ϕ , nazywana generatorem charakterystycznym, są wyznaczone z dokładnością do stałej. To znaczy, jeśli $\mathbf{X} \sim E_n(\mu, \Sigma, \phi) = E_n(\mu^*, \Sigma^*, \phi^*)$, to $\mu = \mu^*$ oraz istnieje $c > 0$ takie, że $\Sigma^* = c\Sigma$, $\phi^*(u) = \phi\left(\frac{u}{c}\right)$ [Embrechts, Lindskog, McNeil 2001; Smidt 2002].

Dobierając odpowiednio stałą c , możemy otrzymać $\Sigma = \text{cov}(\mathbf{X})$, czyli macierz kowariancji. Ponadto dla dowolnej reprezentacji $\mathbf{X} \sim E_n(\mu, \Sigma, \phi)$

$$R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

jest współczynnikiem korelacji Pearsona między brzegowymi zmiennymi losowymi X_i oraz X_j . Jeśli wielowymiarowa zmienna losowa ma rozkład $E_n(\mathbf{0}, \mathbf{R}, \phi)$, gdzie $\mathbf{R}$ jest jej macierzą korelacji, to mówimy, że ma *standardowy rozkład eliptyczny*.

Rozkłady brzegowe wielowymiarowego rozkładu eliptycznego mają również rozkład eliptyczny, podobnie jak rozkłady warunkowe, z tym że w ostatnim przypadku mogą być to rozkłady eliptyczne innego typu, tzn. mają inny generator charakterystyczny ϕ [Embrechts, Lindskog, McNeil 2001; Schmidt 2002]. Jednowymiarowe rozkłady eliptyczne sprowadzają się do rozkładów symetrycznych.

Gęstość wielowymiarowej zmiennej losowej $\mathbf{X}$ o rozkładzie eliptycznym, o ile taka istnieje, przyjmuje postać

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{c_n}{\sqrt{|\Sigma|}} g_n \left(\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right),$$

gdzie $g_n: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, a $c_n > 0$. Funkcja g_n może zależeć od wymiaru n i nazywa się generatorem gęstości. Natomiast c_n jest stałą normującą równą

$$c_n = \frac{\Gamma(n/2)}{(2\pi)^{n/2}} \left(\int_0^x x^{n/2-1} g_n(x) dx \right)^{-1}$$

W tym przypadku stosujemy oznaczenie $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, g_n)$. Jak łatwo zauważyć, warstwy gęstości są elipsoidami. Stąd pochodzi nazwa tak określonego rozkładu.

Wielowymiarową zmienną losową o rozkładzie eliptycznym $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \phi)$ można też scharakteryzować przez *reprezentację stochastyczną*:

$$\mathbf{X} = \boldsymbol{\mu} + Q\mathbf{A}\mathbf{U},$$

gdzie macierz $\mathbf{A}$ jest rozbiciem macierzy Σ spełniającym warunek $\mathbf{A}\mathbf{A}^T = \Sigma$, Q jest dodatnią zmienną losową, a $\mathbf{U}$ jest wektorem losowym o rozkładzie jednostajnym na sferoidzie $\{\mathbf{z} \in \mathbb{R}^n: \mathbf{z}^T \mathbf{z} = 1\}$. Zakłada się, że zmienne losowe Q oraz $\mathbf{U}$ są niezależne. Zmienna losowa Q wyznacza nam rodzaj rozkładu eliptycznego.

Jeśli wektor losowy $\mathbf{X}$ o rozkładzie eliptycznym ma ciągle zmienne brzegowe X_i , to współczynnik Kendalla między zmiennymi losowymi X_i oraz X_j jest funkcją współczynnika korelacji Pearsona R_{ij} :

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(R_{ij}).$$

Przedstawimy teraz w dużym skrócie stosowane w praktyce podstawowe wielowymiarowe rozkłady eliptyczne (więcej informacji w [Embrechts, Lindskog, McNeil 2001; Landsman, Valdez 2003]).

a) Normalny

Najczęściej stosowany jest wielowymiarowy rozkład eliptyczny oznaczany symbolem $N_n(\boldsymbol{\mu}, \Sigma)$, gdzie $\boldsymbol{\mu}$ jest wektorem wartości oczekiwanych zmiennych brzegowych, a Σ jest macierzą kowariancji. Generatory charakterystyczne i gęstości są sobie równe i wynoszą

$$\phi(u) = g(x) = \exp(-u).$$

Natomiast zmienna losowa Q wyznaczająca reprezentację stochastyczną jest związana z rozkładem chi-kwadrat:

$$Q \sim \sqrt{\chi_n^2}.$$

Gęstość wielowymiarowego rozkładu normalnego jest równa

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

Rozkład warunkowy wielowymiarowego rozkładu normalnego jest również normalny, a niezależność rozkładów brzegowych jest równoważna nieskorelowaniu, tzn. zmienne losowe X_i, X_j są niezależne wtedy i tylko wtedy, gdy $R_{ij} = 0$. Właśność ta zachodzi wśród rozkładów eliptycznych jedynie dla rozkładu normalnego.

b) t -Studenta

Jest to wielowymiarowe uogólnienie rozkładu t -Studenta. Oprócz wektora wartości oczekiwanych $\boldsymbol{\mu}$ i macierzy Σ parametrem tego rozkładu jest również liczba stopni swobody ν . Stosuje się wtedy oznaczenie $\mathbf{X} \sim \mathbf{t}_n(\boldsymbol{\mu}, \Sigma, \nu)$. Macierz Σ spełnia warunek $\text{cov}(\mathbf{X}) = \frac{\nu}{\nu+2} \Sigma$, a generator gęstości wynosi

$$g_n(x) = \left(1 + \frac{2x}{\nu}\right)^{-0,5(n+\nu)}$$

Zmienna losowa wyznaczająca reprezentację stochastyczną związana jest z rozkładem F Snedecora. Zachodzi zależność

$$\frac{Q^2}{n} \sim F(n, \nu).$$

Gęstość wielowymiarowego rozkładu t -Studenta określona jest wzorem

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma((n+\nu)/2)}{\Gamma(\nu/2)\sqrt{(v\pi)^n |\Sigma|}} \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\nu}\right)^{-(n+\nu)/2}$$

Ponadto wielowymiarowa zmienna losowa $\mathbf{X}$ o rozkładzie t -Studenta określona może być reprezentacją stochastyczną opartą na standardowym wielowymiarowym rozkładzie normalnym:

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{W} \mathbf{A}^T \mathbf{Z},$$

gdzie $\mathbf{Z} \sim \mathbf{N}_n(\mathbf{0}, \mathbf{R})$, $\mathbf{A}\mathbf{A}^T = \Sigma$, a zmienna losowa W związana jest z odwrotnym rozkładem gamma, tzn. $W \sim \text{Ig}(\nu/2, \nu/2)$. Zmienna losowa W jest niezależna od zmiennej $\mathbf{Z}$. Oprócz tego jej odwrotność związana jest z rozkładem chi-kwadrat.

Zachodzi bowiem zależność $\frac{\nu}{W} \sim \chi_{\nu}^2$.

Granicznym rozkładem, gdy n dąży do nieskończoności, wielowymiarowego rozkładu t -Studenta jest wielowymiarowy rozkład normalny.

c) Logistyczny

Wielowymiarowy rozkład logistyczny jest rzadziej stosowanym rozkładem eliptycznym. Jego generator gęstości jest równy

$$g(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2},$$

a gęstość określona jest wzorem

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{c_n}{\sqrt{|\Sigma|}} \frac{\exp(-0,5(\mathbf{x} - \boldsymbol{\mu})^T \Sigma(\mathbf{x} - \boldsymbol{\mu}))}{(1 + \exp(-0,5(\mathbf{x} - \boldsymbol{\mu})^T \Sigma(\mathbf{x} - \boldsymbol{\mu})))^2}.$$

Stała normująca jest w tym wypadku równa [Londsman, Voldez 2003]

$$c_n = \frac{\Gamma(n/2)}{(2\pi)^{n/2}} \left(\int_0^\infty x^{n/2-1} \frac{\exp(-x)}{(1 + \exp(-x))^2} dx \right)^{-1} = \left(2\pi^{n/2} \sum_{j=1}^\infty (-1)^{j-1} j^{1-n/2} \right)^{-1}.$$

d) Wykładniczo-potęgowy

Rozkład ten jest uogólnieniem wielowymiarowego rozkładu normalnego. Charakteryzowany jest on przez dwa dodatkowe parametry $r, s > 0$. Jego generator jest równy

$$g(x) = \exp(-rx^s),$$

a gęstość opisana jest formułą

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{s\Gamma(n/2)r^{n/(2s)}}{\Gamma(n/(2s))(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{r}{2^s}((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))^s\right).$$

Gdy $r = s = 1$, otrzymujemy wielowymiarowy rozkład normalny.

6. Eliptyczne funkcje łączące

Funkcje łączące generowane przez wielowymiarowe rozkłady eliptyczne nazywamy *eliptycznymi*. Są one określone wzorem:

$$C(u_1, \dots, u_n) = H_n(H^{-1}(u_1), \dots, H^{-1}(u_n)),$$

gdzie H_n jest n -wymiarową dystrybuantą standardowego rozkładu eliptycznego $E_n(\mathbf{0}, \mathbf{R}, g_n)$, a H dystrybuantą jego rozkładu brzegowego $E_1(0, 1, g_1)$. Można je również przedstawić jawną formułą

$$C(u_1, \dots, u_n) = \int_{-\infty}^{H^{-1}(u_1)} \int_{-\infty}^{H^{-1}(u_n)} \frac{c_n}{\sqrt{|\mathbf{R}|}} g_n\left(\frac{1}{2} \mathbf{x}^T \mathbf{R}^{-1} \mathbf{x}\right) dx_1 \dots dx_n.$$

Niech $X_1, \dots, X_n$ będą zmiennymi losowymi z ciągłą dystrybuantą F_i , a C będzie eliptyczną funkcją łączącą, wtedy mówimy, że wielowymiarowa zmienna losowa o dystrybucie określonej wzorem

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

ma rozkład *metaeliptyczny* [Fong, Fong 2002]. Rozkład metaeliptyczny ma tę samą funkcję łączącą co odpowiadający mu rozkład eliptyczny, ale w odróżnieniu od niego może mieć dowolne rozkłady brzegowe. Współczynnik korelacji Kendalla między zmiennymi brzegowymi określony jest zależnością

$$\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(R_{ij}). \quad (2)$$

Nie zależy on od rozkładów brzegowych, więc zależność ta jest taka sama jak dla rozkładów eliptycznych. W odróżnieniu od archimedesowych funkcji łączących dla każdej pary zmiennych brzegowych współczynnik Kendalla może przyjmować różne wartości.

Scharakteryzujemy teraz pokrótce najważniejsze eliptyczne funkcje łączące indukowane przez podstawowe rozkłady eliptyczne.

a) Gaussowska funkcja łącząca

Funkcja ta jest indukowana przez standardowy wielowymiarowy rozkład normalny, tzn. $H_n = \Phi_n \sim N_n(\mathbf{0}, \mathbf{R})$ oraz $H = \Phi \sim N(0, 1)$. Oznaczać ją będziemy symbolem $C_{\mathbf{R}}^{Ga}(u_1, \dots, u_n)$. Jej parametrem jest macierz korelacji $\mathbf{R}$.

Gdy wartość współczynnika korelacji między parą brzegowych zmiennych losowych X_i i X_j jest różna od 1, to Gaussowska funkcja łącząca jest asymptotycznie niezależna. Zachodzą bowiem zależności

$$\lambda_U(X_i, X_j) = \lambda_L(X_i, X_j) = 0$$

dla każdej pary rozkładów brzegowych X_i, X_j . Dla ścisłej zależności między zmiennymi brzegowymi otrzymujemy bezwzględne wartości tych współczynników równe 1. W przypadku dwuwymiarowym jej parametrem jest współczynnik korelacji r i można ją przedstawić wzorem:

$$C_r^{Ga}(u, v) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) dx dy.$$

b) Funkcja łącząca Studenta

Funkcja łącząca Studenta jest indukowana wielowymiarowym standardowym rozkładem t -Studenta, tzn. $H_n \sim t_n(\mathbf{0}, \mathbf{R}, \nu)$ oraz $H = t_\nu$. Będziemy oznaczać ją symbolem $C_{\nu, \mathbf{R}}^t(u_1, \dots, u_n)$. Jej parametrami są: macierz korelacji $\mathbf{R}$ oraz liczba stopni swobody ν , a współczynniki zależności wartości ekstremalnych są różne od zera i wynoszą

$$\lambda_U(X_i, X_j) = \lambda_L(X_i, X_j) = 2 \left(1 - t_{\nu+1} \left(\frac{\sqrt{(\nu+1)(1-R_{ij})}}{\sqrt{1+R_{ij}}} \right) \right).$$

W przypadku dwuwymiarowym funkcja ta określona jest wzorem

$$C'_{v,\rho}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \left(1 + \frac{x^2 - 2\rho xy + y^2}{(1-\rho^2)} \right)^{-(v+2)/2} dx dy.$$

W podobny sposób określamy *logistyczne* i *wykładniczo-potęgowe* funkcje łączące. Podobnie jak gaussowska funkcja łącząca są one również asymptotycznie niezależne.

7. Identyfikacja funkcji łączącej

Wybór odpowiedniej funkcji łączącej, jej „dopasowanie” do danych z próby jest bardzo ważnym problemem spotykanym w zagadnieniach praktycznych. Dane w tym przypadku tworzą ciąg wektorów $\mathbf{x}_1, \dots, \mathbf{x}_m$, gdzie $\mathbf{x}_i = (x_{i1}, \dots, x_{in})$. Natomiast sam proces identyfikacji sprowadza się zwykle do dwóch zagadnień: estymacji parametrów przy ustalonej rodzinie funkcji łączących oraz do wyboru odpowiedniej rodziny.

Proces estymacji parametrów funkcji łączącej rozpoczyna się zwykle od estymacji rozkładów brzegowych. Estymacja ta może być parametryczna, nieparametryczna lub mieszana. Po ustaleniu rozkładów brzegowych estymację parametrów funkcji łączącej można przeprowadzić metodą największej wiarygodności (MNW). W przypadku funkcji eliptycznych parametrami są: wyrazy macierzy korelacji $\mathbf{R}$ oraz wektor dodatkowych parametrów $\mathbf{a}$, takich jak stopnie swobody ν dla funkcji Studenta czy stałe r, s dla funkcji wykładniczo-potęgowej. Stosując MNW, znajdujemy wartości parametrów maksymalizujące wyrażenie

$$\log L(\mathbf{a}, \mathbf{R}; \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_m) = \sum_{i=1}^m \log c_{\mathbf{a}, \mathbf{R}}(\hat{\mathbf{u}}_i),$$

gdzie $\hat{\mathbf{u}}_i = (u_{i1}, \dots, u_{in}) = (\hat{F}_1(x_{i1}), \dots, \hat{F}_n(x_{in}))$, $\hat{F}_j$ jest wcześniej wyestymowaną dystrybucją brzegową, a $c_{\mathbf{a}, \mathbf{R}}$ gęstością funkcji łączącej.

MNW jest zalecaną metodą estymacji parametrów. Jednakże przy dużej liczbie wymiarów mogą się pojawić kłopoty z wyznaczeniem rozwiązania powyższego problemu optymalizacyjnego. Wtedy można skorzystać z prostszej metody opartej na zależności (2). Jest to w pewnym sensie uogólnienie metody momentów. Polega na wyznaczeniu empirycznego współczynnika korelacji Kendalla

$$\hat{\tau}(X_j, X_k) = \binom{m}{2}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq m} \text{sign}(x_{i_1, j} - x_{i_2, j})(x_{i_1, k} - x_{i_2, k}),$$

a następnie, z wykorzystaniem (2), obliczeniu empirycznego współczynnika korelacji Pearsona

$$\hat{R}_{jk} = \sin\left(\frac{\pi}{2} \hat{\tau}(X_j, X_k)\right).$$

W ten sposób estymujemy macierz korelacji $\mathbf{R}$. Pozostałe parametry, czyli wektor $\mathbf{a}$, estymujemy, stosując MNW.

Wybór odpowiedniej rodziny funkcji łączącej polega na ustaleniu listy dopuszczalnych rodzin. Następnie dla każdej rodziny estymujemy parametry, stosując przedstawione wcześniej metody. Właściwą rodzinę funkcji łączących wybieramy według ustalonego kryterium, takiego jak: wartość logarytmu funkcji wiarygodności lub wartość AIC, czy porównanie funkcji łączącej C z jej wersją empiryczną. Ewentualnie można stosować metody graficzne, np. wykres kwantylowy. Więcej informacji dotyczącej problemów estymacji można znaleźć w [Frees, Valdez 1998; Joe 1997; Demarta, McNeil].

8. Symulacje

Funkcje łączące można wykorzystać do symulacji wartości z wielowymiarowego rozkładu. W naszym przypadku będą to rozkłady metaeliptyczne, indukowane przez eliptyczne funkcje łączące. Możliwość prostej symulacji należy do najważniejszych zalet funkcji łączących.

Symulacja wartości z rozkładu metaeliptycznego polega na wygenerowaniu wartości losowych z odpowiedniego rozkładu eliptycznego, nałożeniu na te wartości dystrybuant brzegowych, dzięki czemu otrzymuje się wartości wygenerowane przez eliptyczną funkcję łączącą, a następnie wyznaczeniu przeciwobrazów dystrybuant brzegowych rozkładu metaeliptycznego. Jest ona zwykle oparta na reprezentacjach stochastycznych danego standardowego rozkładu eliptycznego. Zwykle jest to reprezentacja postaci

$$\mathbf{X} = {}_d W \mathbf{Z},$$

gdzie W jest nieujemną zmienną losową, a $\mathbf{Z} \sim \mathbf{N}_n(\mathbf{0}, \mathbf{R})$. Można również stosować reprezentację bazującą na wielowymiarowej zmiennej losowej o rozkładzie jednostajnym na sferoidzie.

W przypadku Gaussowskiej funkcji łączącej $C_{\mathbf{R}}^{Ga}$ korzystamy z reprezentacji $\mathbf{X} = \mathbf{A}\mathbf{T}$, gdzie $\mathbf{T} = (T_1, \dots, T_n)$, a $T_i \sim N(0, 1)$ są niezależnymi zmiennymi losowymi. Algorytm generowania wartości losowych jest następujący:

1. Znaleźć dekompozycję Cholesky'ego $\mathbf{A}\mathbf{A}^T = \mathbf{R}$.
2. Symulacja n niezależnych wartości $t_1, \dots, t_n$ z rozkładu $N(0, 1)$.
3. Wyznaczyć $\mathbf{x} = \mathbf{A}\mathbf{t}$, gdzie $\mathbf{t} = (t_1, \dots, t_n)$.
4. Obliczyć $u_i = \Phi(x_i)$, $i = 1, \dots, n$.

Wartości $x_1, \dots, x_n$ są wygenerowane z wielowymiarowego standardowego rozkładu normalnego $\mathbf{N}_n(\mathbf{0}, \mathbf{R})$, $u_1, \dots, u_n$ z Gaussowskiej funkcji łączącej $C_{\mathbf{R}}^{Ga}$, a $y_1 = F_1^{-1}(u_1)$, $\dots$, $y_n = F_n^{-1}(u_n)$ z rozkładu metaeliptycznego o brzegowych dystrybuantach $F_1, \dots, F_n$.

Podobnie postępujemy w przypadku funkcji łączącej Studenta $C'_{v,R}$, wykorzystując reprezentację stochastyczną

$$\mathbf{Y} = \frac{\sqrt{k}}{\sqrt{S}} \mathbf{X},$$

gdzie $S \sim \chi_k^2$, a $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{R})$. Punkty 1-3 są takie same jak dla Gaussowskiej funkcji łączącej. Następne punkty algorytmu są postaci:

4. Wygenerować wartość s z rozkładu χ_v^2 niezależnie od $\mathbf{t}$.

5. Wyznaczyć $\mathbf{y} = \sqrt{\frac{v}{s}} \mathbf{x}$.

6. Obliczyć $u_i = t_v(y_i)$, $i = 1, \dots, n$.

9. Uogólnienia eliptycznych funkcji łączących

Eliptyczne funkcje łączące charakteryzują się symetrią, tzn. permutacja argumentów nie zmienia wartości funkcji łączącej. Jednakże często w zagadnieniach praktycznych symetria nie występuje. W tym celu wprowadza się pewne modyfikacje powodujące skośność rozpatrywanego rozkładu eliptycznego.

W wyniku jednej z takich modyfikacji otrzymujemy skośną funkcję łączącą Studenta [Demarta, McNeil 2004]. Indukowana jest ona przez rozkład oparty na reprezentacji stochastycznej:

$$\mathbf{X} = \mu + g(W)\boldsymbol{\gamma} + \sqrt{W} \mathbf{Z},$$

gdzie $\mathbf{Z} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, $W \sim \text{Ig}(v/2, v/2)$, $g: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ oraz $\boldsymbol{\gamma} \in \mathbb{R}^n$ jest wektorem powodującym skośność rozkładu. Zwykle przyjmuje się $g(x) = x$. Tak otrzymany rozkład oznaczamy symbolem $\mathbf{X} \sim \mathbf{t}_n(v, \mu, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$ i nazywamy wielowymiarowym skośnym rozkładem Studenta. Jego wektor wartości oczekiwanych oraz macierz kowariancji określone są wzorami:

$$E(\mathbf{X}) = \mu + \frac{v}{v-2} \boldsymbol{\gamma}, \quad \text{cov}(\mathbf{X}) = \frac{v}{v-2} \boldsymbol{\Sigma} + \frac{2v^2}{(v-2)^2(v-4)} \boldsymbol{\gamma} \boldsymbol{\gamma}^T.$$

Indukowaną przez skośny rozkład Studenta funkcję łączącą będziemy oznaczać symbolem $C'_{v,R,\boldsymbol{\gamma}}$. W podobny sposób możemy otrzymać skośne eliptyczne funkcje łączące innego typu.

W przypadku wielowymiarowego standardowego rozkładu Studenta $\mathbf{X} \sim \mathbf{t}_n(v, \mathbf{0}, \boldsymbol{\Sigma})$ wszystkie rozkłady brzegowe mają tę samą liczbę stopni swobody v . Rozkład ten można uogólnić, dzieląc składowe wektora $\mathbf{X}$ na grupy $\mathbf{X}_j$ o różnej liczbie stopni swobody v_j . Dla każdej grupy wyznaczamy zmienną losową $W_j = G_{v_j}^{-1}(U)$,

gdzie G_v jest dystrybuantą rozkładu $\text{Ig}(v/2, v/2)$ oraz U ma rozkład jednostajny na $[0, 1]$. Zakładamy też, że każda zmienna brzegowa X_i należąca do j -tej grupy ma następującą reprezentację stochastyczną

$$X_i \sim \sqrt{W_j} Z_i,$$

gdzie $\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma)$. Współczynnik korelacji Kendalla jest wtedy w przybliżeniu równy

$$\tau(X_i, X_j) \approx 2\arcsin(R_{ij})/\pi.$$

Literatura

- Demarta S., McNeil A.J., *The t Copula and Related Copulas*, Technical Report, ETH Zurich 2004.
- Embrechts P., Lindskog F., McNeil A., *Modelling Dependence with Copulas and Applications to Risk Management*, ETH Zurich, preprint, 2001.
- Fang H., Fang K., *The Meta-elliptical Distributions with Given Marginals*, „J. Multivariate Anal.” 82 2002 1-16.
- Frees E.W., Valdez E.A., *Understanding Relationships Using Copulas*, „North Amer. Actuarial J.” 2 1998 1-25.
- Heilpern S., *Wybrane metody badania zależności*, Prace Naukowe AE Kraków (w recenzji).
- Joe H., *Multivariate Models and Dependence Concepts*, Chapman & Hall, London 1997.
- Landsman Z.M., Valdez E.A., *Tail Conditional Expectations for Elliptical Distributions*. „North Amer. Actuarial J.” 7 2003 55-71.
- Nelsen R.B., *An Introduction to Copulas*, Springer, New York 1999.
- Schmidt R., *Tail Dependence for Elliptically Contoured Distributions*, „Math. Methods for Operations Research” 55 2002 301-327.
- Schweitzer B., Sklar A., *Probabilistic Metric Spaces*, North-Holland, New York 1981.
- Wang S.S., *Aggregation of Correlated Risk Portfolios: Models & Algorithms*, CAS Committee on Theory of Risk, Working Paper, 1999.

ELLIPTICAL COPULAS

Summary

The paper is devoted to the elliptical copulas. These copulas are generated by the elliptical distributions. The main properties of the elliptical copulas are presented. The characteristic properties, which are distinguishable from other copulas, mainly Archimedean, are emphasized. The basic types of the elliptical distributions, the elliptical copulas: Gaussian and Student, their identification, the problem of simulation and the direction of generalization are discussed.