FUNCȚII DE CUPLARE ȘI ANALIZA DEPENDENȚELOR: O ABORDARE MATEMATICĂ A FENOMENELOR PSIHOLOGICE

Student: Ciulavu George-Adrian
Îndrumător: Prof.univ.dr. Vernic Raluca-Ileana

Facultatea de Matematică și Informatică
Universitatea ,,Ovidius" din Constanța
Sesiunea de Comunicări Matematice

7.decembrie.2024

Cuprins

- Introducere.
- Noțiuni introductive: funcții de cuplare.
  - Funcția de cuplare bidimensională.
- Familii de funcții de cuplare.
  - Funcții de cuplare eliptice.
  - Funcții de cuplare Arhimediene.
- ANXIETATE VS STIMA DE SINE
- Aplicație
- Rezultatele obținute
- Concluzii
- Bibliografie.

Introducere

Multe situații din viața reală pot fi modelate printr-un număr mare de variabile care joacă un rol semnificativ și astfel de variabile nu sunt în general independente.

În teoria probabilităților și statistică, o funcție de cuplare este o funcție de repartiție cumulativă multivariată pentru care distribuția de probabilitate marginală a fiecărei variabile este uniformă pe intervalul $[0, 1]$.

Funcțiile de cuplare sunt folosite pentru a descrie/modela dependența dintre variabilele aleatoare.

Funcțiile de cuplare sunt populare în aplicațiile statistice, deoarece permit modelarea și estimarea cu ușurință a repartiției vectorilor aleatori prin estimarea separată a marginalelor și a funcțiilor de cuplare.

Funcția de repartiție

$F_X : \mathbb{R} \to [0,1]$ se numește **funcție de repartiție** asociată variabilei aleatoare X dacă $F_X(x) = P(X < x) \overset{\text{not.}}{=} P(X^{-1}(-\infty, x))$

Funcția de repartiție pentru variabile aleatoare de tip continuu

O funcție de repartiție F este de tip **absolut continuu** dacă $\exists f : \mathbb{R} \to \mathbb{R}$ măsurabilă Borel astfel încât $F(x) = \int_{-\infty}^{x} f(u)du$. Atunci f se numește **densitate de repartiție**.

Funcția de repartiție multidimensională $F_X : \mathbb{R}^d \to [0,1]$ este:

$$F_X(a_1, ..., a_d) = P(X_1 < a_1, ..., X_d < a_d)$$

Definiția matematică a funcției de cuplare

Fie vectorul aleatoriu $(X_1, X_2, ..., X_d)$ având marginalele continue; adică funcția de repartiție cumulativă (FRC) $F_i(x)$ este o funcție continuă.

Se știe că vectorul aleatoriu $(U_1, U_2, ..., U_d) = (F_1(X_1), F_2(X_2), ..., F_d(X_d))$ are marginalele repartizate uniform pe intervalul $[0, 1]$.

Funcția de cuplare a lui $(X_1, X_2, ..., X_d)$ este definită ca funcția de repartiție cumulativă comună a $(U_1, U_2, ..., U_d)$:
$C(u_1, u_2, ...u_d) = P(U_1 \leq u_1, U_2 \leq u_2, ..., U_d \leq u_d)$
$= P(X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), ..., X_d \leq F_d^{-1}(u_d))$.

1. Funcția de cuplare C conține toate informațiile despre structura de dependență dintre componentele lui $(X_1, X_2, ..., X_d)$, în timp ce funcția cumulativă F_i conține toate informațiile despre repartiția marginală a lui X_i.
2. Funcția de cuplare se poate folosi și pentru a genera eșantioane pseudo-aleatoare din clasele generale de repartiții de probabilitate multivariate: se generează un eșantion $(U_1, U_2, ..., U_d)$ din funcția de cuplare, din care se construiește eșantionul necesar astfel: $(X_1, X_2, ..., X_d) = (F_1^{-1}(U_1), F_2^{-1}(U_2), ..., F_d^{-1}(U_d))$.

Funcția de cuplare bidimensională

Funcția de cuplare bidimensională

$$C(x, y) = P(U_1 \leq x, U_2 \leq y)$$

Din această funcție rezultă:

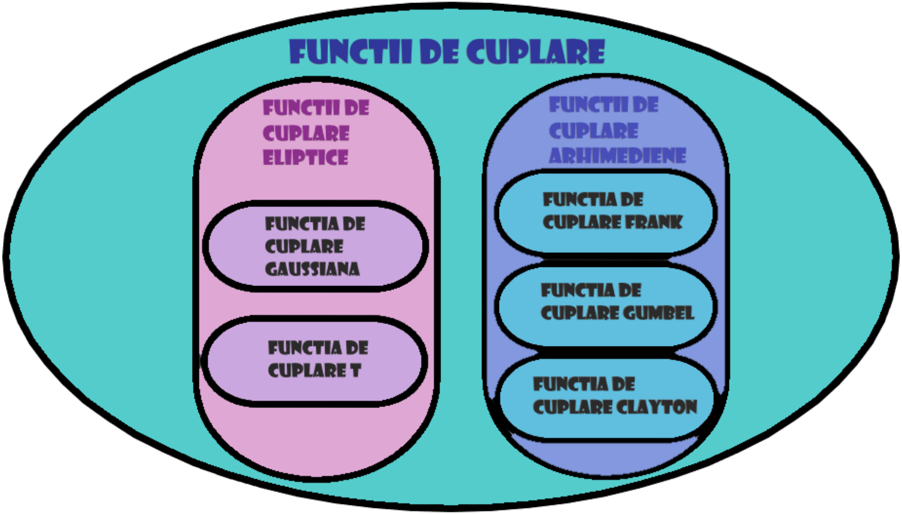
- $C(x, 0) = P(U_1 \leq x, U_2 \leq 0) = 0$
- $C(0, y) = P(U_1 \leq 0, U_2 \leq y) = 0$
- $C(x, 1) = P(U_1 \leq x, U_2 \leq 1) = P(U_1 \leq x) = x$
- $C(1, y) = P(U_1 \leq 1, U_2 \leq y) = P(U_2 \leq y) = y$

Se demonstrează că:

$$max(x + y - 1) \leq C(x, y) \leq min(x, y)$$

fiind un caz special pentru cele mai importante rezultate ale statisticii bidimensionale.

Familii de funcții de cuplare



ANXIETATE VS STIMA DE SINE

În ultimele decenii, relația dintre diferiți parametri psihologici a fost studiată în detaliu.
Cu toate acestea, structura de dependență a parametrilor corelați a fost rareori studiată.
Cunoașterea structurii de dependență ajută la găsirea matricei de probabilitate a interacțiunii dintre parametri.

În această prezentare vom analiza stima de sine și anxietatea a 141 de studenți, datele fiind extrase folosind ,,Inventarul de stimă de sine Coopersmith” și ,,Scala de anxietate Zang”.

Introducere

Tulburarea de anxietate este al treilea cel mai frecvent tip de tulburare mentală după tulburarea depresivă majoră și dependența de alcool.
În ultimele decenii, au fost efectuate mai multe studii privind anxietatea studenților.
Aceste studii au arătat că experiența anxietății în timpul școlarizării creează probleme în următoarele etape ale vieții studenților care, pe lângă discomfortul personal, pot avea un impact negativ asupra locului de muncă și a performanței zilnice.

Studiile de cerecetare au arătat că autoevaluarea negativă, teama de a evalua pe alții și evitarea situațiilor care implică evaluarea sunt printre cele mai evidente caracteristici ale fobiei sociale.
În ultimii ani, cercetătorii au citat stima de sine ridicată ca o sursă importantă de sprijin pe care oamenii o pot folosi împotriva evenimentelor din viața de zi cu zi.

Cinci funcții de cuplare arhimediene, precum și versiunile Normal și Plackett, au fost folosite pentru modelarea structurii de dependență a stimei de sine și a anxietății.
Proprietățile acestor funcții sunt prezentate în tabelul de mai jos:

Copula	Domain of Dependence Parameter θ	Copula Function $C(u,v)$	Kendall's Tau (τ)
Clayton	$[-1,\infty)\backslash\{0\}$	$\left\{max\left(u_1{}^{-\theta}+u_2{}^{-\theta}-1,0\right)\right\}^{\frac{-1}{\theta}}$	$[-1, 1]$
Frank	$(-\infty,\infty)\backslash\{0\}$	$-\frac{1}{\theta}log\left[1+\frac{\{\exp(-\theta u_1)-1\}\{\exp(-\theta u_2)-1\}}{exp(-\theta)-1}\right]$	$[-1, 1]$
Gumbel–Hougaard	$[1,\infty)$	$exp\left[-\left\{(-logu_1)^{\frac{1}{\theta}}+(-logu_2)^{\frac{1}{\theta}}\right\}^{\theta}\right]$	$[0, 1]$
Joe	$[1,\infty)$	$1-\left\{(1-u_1)^{\theta}+(1-u_2)^{\theta}-(1-u_1)^{\theta}(1-u_2)^{\theta}\right\}^{\frac{1}{\theta}}$	$[0, 1]$
Ali–Mikhail–Hag	$[-1, 1]$	$\frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)}$	$[-0.18, 0.33]$
Normal	$[-1, 1]$	$\int_{-\infty}^{\Phi^{-1(u_1)}}\int_{-\infty}^{\Phi^{-1(u_1)}}\frac{1}{2\pi\sqrt{1-\theta^2}}exp\left\{\frac{2\theta st-s^2-t^2}{2(1-\theta^2)}\right\}dsdt$	$[-1, 1]$
Plackett	$[0,\infty)$	$\frac{\{1+(\theta-1)(u_1+u_2)\}-\sqrt{\{1+(\theta-1)(u_1+u_2)\}^2-4u_1u_2\theta(\theta-1)}}{2(\theta-1)}$	$[-1, 1]$

Pentru a selecta cea mai bună funcție de distribuție cumulativă pentru datele de stimă de sine și anxietate, au fost folosite cinci distribuții:

Distribution	**CDF**
Normal	$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)\ 0 < x < \infty$ Φ is the CDF of standard normal distribution
Log-normal	$F(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)\ 0 < x < \infty$
Inverse Gaussian	$F(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right) + \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right)e^{\frac{2\lambda}{\mu}}\ 0 < x < \infty$ Φ is the CDF of standard normal distribution
Generalized Pareto	$F(x) = 1 - \left(1 + k\frac{(x-\mu)}{\sigma}\right)^{\frac{-1}{k}}\ \mu \leq x < \infty$
Gamma	$F(x) = \frac{\Gamma_{\frac{x}{\beta}}(\alpha)}{\Gamma(\alpha)}\ 0 < x < \infty$

Rezultatele obținute:

Parametrii estimați ai distribuțiilor marginale adaptate la datele privind stima de sine și anxietatea.

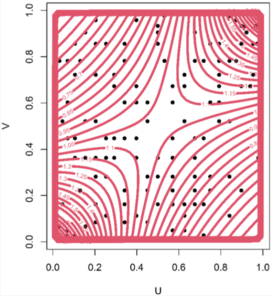
	Margin	Parameters	AIC	K–S Test	
				Statistics	p-Value
Self-esteem	**Normal**	**$(\mu, \sigma) = (26.77, 4.81)$**	853.42	0.065	0.56
	Log-normal	$(\mu, \sigma) = (3.27, 0.19)$	853.2	0.08	0.32
	Inverse Gaussian	$(\lambda, \mu) = (828.53, 26.77)$	**848.2**	0.06	0.67
	Gamma	$(\alpha, \beta) = (30.95, 0.87)$	849.4	0.07	0.47
	Generalized pareto	$(\mu, \sigma, k) = (18.47, 16.88, -1.03)$	857.2	0.082	0.28
Anxiety	Normal	$(\mu, \sigma) = (46, 9.81)$	1039.1	0.066	0.55
	Log-normal	$(\mu, \sigma) = (3.81, 0.21)$	1038.54	0.062	0.62
	Inverse Gaussian	$(\lambda, \mu) = (1012.1, 46)$	1040.2	0.07	0.47
	Gamma	$(\alpha, \beta) = (22, 2.09)$	**1037.4**	0.048	0.87
	Generalized pareto	$(\mu, \sigma, k) = (31.32, 24.54, -0.67)$	1042.3	0.076	0.36

Rezultatele optime a funcțiilor de cuplare.

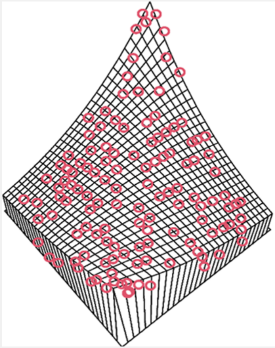
Copula	θ	Test Statistic	p-Value
Clayton	0.469	0.962	0.19
Frank	1.734	0.717	0.58
Gumbel–Hougaard	1.236	0.752	0.48
Joe	1.343	1.134	0.07
Ali–Mikhail–Hag	0.593	0.917	0.22
Normal	0.305	0.728	0.53
Plackett	2.386	0.729	0.52

Dintre toate funcțiile de cuplare, Frank a avut cea mai bună potrivire.

Schema de contur a funcției de cuplare Frank cu graficul scatter al anxietății comune și al combinațiilor de stimă de sine:



Schema de perspectivă a funcției de cuplare Frank și graficul scatter al combinațiilor comune de anxietate și stimă de sine:



Matricea de probabilitate între combinațiile de stimă de sine și anxietate.

			Anxiety			
			Normal	**Mild**	**Severe**	**Very Severe**
		Number of Students	**67**	**62**	**11**	**1**
Self-esteem	Poor	67	1.33	0.4	28	70.4
	Average	73	0.32	36	17.9	46
	Strong	1	44.3	35.6	19.8	0.33

Concluzii:

În acest studiu, a fost introdusă o nouă abordare pentru analiza psihologică folosind funcțiile de cuplare.
Pentru a costrui structura de dependență și pentru a produce matricea de probabilitate a două variabile dependente, sima de sine și anxietatea studenților, a fost folosită funcția de cuplare de tip Frank.

Matricea de probabilitate a arătat că stima de sine scăzută duce la anxietate severă cu o probabilitate foarte mare și o probabilitate neglijabilă alocată anxietății ușoare și normale.
În schimb, stima de sine ridicată este asociată cu o anxietate normală/ușoară, cu o probabilitate de 80 %.
Cu toate acestea, aproximativ 20 % din probabilitate este alocată anxietății severe, ceea ce arată că, chiar și stima de sine ridicată poate fi asociată cu anxietate severă.

Bibliografie

- https://twiecki.io/blog/2018/05/03/copulas/
- https://slideplayer.com/slide/3270700/
- https://www.slideshare.net/GiovanniDellaLunga/copule-slides
- https://cermics.enpc.fr/ lelievre/CEMRACS/slides/Lebrun.pdf
- Dehghani, E., Ranjbar, S. H., Atashafrooz, M., Negarestani, H., Mosavi, A., Kovacs, L. (2021). Introducing Copula as a Novel Statistical Method in Psychological Analysis. International Journal of Environmental Research and Public Health, 18(15), 7972.

Multumesc pentru atentie!