3.) Kopuly extrémních hodnot

Příklad (motivační)

Nechť $(X_1, Y_1)^T, \dots (X_n, Y_n)^T$ je náh. výběr z dvourozměrného rozdělení se sdruženou distribuční funkcí $F(x,y)$ a marginálními dist. funkcemi $F_1(x)$ a $F_2(y)$.

označme $M_n^X = \max\{X_1, \dots, X_n\} \quad \dots \quad F_1^n(x)$

$M_n^Y = \max\{Y_1, \dots, Y_n\} \quad \dots \quad F_2^n(y)$

Sklar $\rightarrow F(x,y) = C(F_1(x), F_2(y))$

Jak vypadá sdružená distribuční funkce $(M_n^X, M_n^Y)^T$?

$$F^{max}(x,y) = P(M_n^X \le x, M_n^Y \le y) = P(X_1 \le x, \dots, X_n \le x, Y_1 \le y, \dots, Y_n \le y) = P((X_1, Y_1)^T \le (x,y)^T, \dots, (X_n, Y_n)^T \le (x,y)^T) = P((X_i, Y_i)^T \le (x,y)^T)^n = F^n(x,y)$$

$$\tilde{C}(F_1^n(x), F_2^n(y)) \qquad\qquad C^n(F_1(x), F_2(y))$$

$$\tilde{C}(F_1^n(x), F_2^n(y)) = C^n(F_1(x), F_2(y)) = C^n\left((F_1^n(x))^{\frac{1}{n}}, (F_2^n(y))^{\frac{1}{n}}\right)$$

$$\tilde{C}(u,v) = C^n(u^{\frac{1}{n}}, v^{\frac{1}{n}})$$

$$\underline{\tilde{C}(u^n, v^n) = C^n(u,v)}$$

Kopula extrémních hodnot je taková kopula, pro kterou platí $C(u^n, v^n) = C^n(u,v)$, $\forall\, 0 \le u \le 1,\ 0 \le v \le 1$ a $\forall n = 1, 2, \dots$. (max-stabilita).

Poznámka

Předchozí vlastnost říká, že kopula příslušná $(M_n^X, M_n^Y)^T$ je stejná jako u $(X_i, Y_i)^T$.

a) Gumbelova kopula

$$C(u,v) = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{\frac{1}{\theta}}\right\}$$, kde $\theta \geq 1$ je parametr.

$$C(u^n, v^n) = \exp\left\{-\left[(\overbrace{-\log u^n}^{-n\log u})^\theta + (\overbrace{-\log v^n}^{-n\log v})^\theta\right]^{\frac{1}{\theta}}\right\} = \exp\left\{-\left[n^\theta(-\log u)^\theta + n^\theta(-\log v)^\theta\right]^{\frac{1}{\theta}}\right\} = \exp\left\{-n\left[(-\log u)^\theta + (-\log v)^\theta\right]^{\frac{1}{\theta}}\right\} = \exp\left\{-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{\frac{1}{\theta}}\right\}^n$$
$$= C^n(u,v).$$

b) Galambosova kopula

$$C(u,v) = u\cdot v\cdot \exp\left\{\left[(-\log u)^{-\theta} + (-\log v)^{-\theta}\right]^{-\frac{1}{\theta}}\right\}$$, $\theta > 0$ parametr.

c) Tawnova kopula

$$C(u,v) = u^{1-\alpha}\cdot v^{1-\beta}\cdot \exp\left\{-\left[(-\alpha\log u)^\theta + (-\beta\log v)^\theta\right]^{\frac{1}{\theta}}\right\}$$, $\theta \geq 1$, $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ jsou parametry.

Jak generovat náhodné veličiny (jejich realizace) z daného rozdělení?

- jednorozměrný případ

předpokládejme, že umíme generovat z $Rs(0,1)$.

- zamítací metoda

- metoda inverzní transformace - chceme generovat realizaci n.v. X s distribuční funkcí F

platí: $U \sim Rs(0,1)$, pak $F^{-1}(U) \sim$ rozdělení s distribuční funkcí F

nejprve generujeme realizaci n.v. z $Rs(0,1)$, označme ji u, pak $F^{-1}(u)$ je realizace n.v. X.

- dvourozměrný případ

nechť náh. vektor $(U,V)^T$ má kopulu $C(u,v)$, $U \sim Rs(0,1)$ a $V \sim Rs(0,1)$. Nechť C je absolutně spojitá

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \, \partial v}$$ je hustota kopuly.

hustota n.v. $U|V=v$ je $c(u|v) = \dfrac{c(u,v)}{\underbrace{c(v)}_{=1}} = c(u,v)$, tj. funkce $c(\cdot,v)$, neboli $u \mapsto c(u,v)$. a podobně

hustota n.v. $V|U=u$ je $c(v|u) = \dfrac{c(u,v)}{\underbrace{c(u)}_{=1}} = c(u,v)$, tj. funkce $c(u,\cdot)$, neboli $v \mapsto c(u,v)$.

distribuční funkce n.v. $V|U=u$ je $v \mapsto \dfrac{\partial C(u,v)}{\partial u}$.

hustota n.v. $V|U=u$

$$\frac{\partial C(u,v)}{\partial u} = \frac{\partial}{\partial u} \int_0^u \left(\int_0^v c(s,t)\,dt \right) ds = \int_0^v \left(\frac{\partial}{\partial u} \int_0^u c(s,t)\,ds \right) dt = \int_0^v \underbrace{c(u,t)}_{f_{V|U=u}(t)}\,dt = P(V \leq v \,|\, U=u) = F_{V|U=u}(v)$$

.. distribuční funkce n.v. $V|U=u$.

podobně distribuční funkce n.v. $U|V=v$ je $u \mapsto \dfrac{\partial C(u,v)}{\partial v}$.

Nyní chceme generovat realizaci náh. vektoru $(X,Y)^T$ s distribuční funkcí $F(x,y)$, označ. marg. dist. fce $F_1(x)$ a $F_2(y)$

Sklar $F(x,y) = C\left(F_1(x), F_2(y)\right)$, ozn. $F_1(x)=u$, $F_2(y)=v$

$$C(u,v) = F\left(F_1^{-1}(u), F_2^{-1}(v)\right)$$

(i) generujeme realizaci n.v. U z $Rs(0,1)$, označme ji u.

(ii) generujeme realizaci n.v. V z podmíněného rozdělení $V|U=u$ s distribuční funkcí $\dfrac{\partial C(u,v)}{\partial u}$, označme ji v.

Tedy (u,v) je realizace kopuly C

(iii) hledanou realizaci (x,y) dostaneme aplikací kvantilových funkcí: $(x,y)^T = \left(F_1^{-1}(u), F_2^{-1}(v)\right)^T$.

Modelování dvourozměrných dat

$(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ je náh. výběr z dvourozměrného rozdělení se sdruženou distr. fcí $F(x,y)$ a marginálními $F_1(x)$ a $F_2(y)$.

Sklar: $F(x,y) = C(F_1(x), F_2(y))$

parametrický přístup

– modelujeme marginální distr. fce $F_1(x)$ a $F_2(y)$ → přidáme neznámé parametry → $F_1(x, \theta_1)$, $F_2(y, \theta_2)$, θ_1 a θ_2 jsou neznámé parametry

– zvolíme vhodnou kopulu C, označme její parametry jako θ_3 → $C_{\theta_3}(u,v)$.

$F(x,y) = C_{\theta_3}(F_1(x,\theta_1), F_2(y,\theta_2))$, kde $\theta = (\theta_1^T, \theta_2^T, \theta_3^T)^T$ je vektor neznámých parametrů.

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \, \partial y} = c_{\theta_3}(F_1(x,\theta_1), F_2(y,\theta_2)) \cdot f_1(x,\theta_1) \cdot f_2(y,\theta_2)$$

, kde $f_1(x,\theta_1)$ a $f_2(y,\theta_2)$ jsou marginální hustoty a $c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \, \partial v}$ je hustota kopuly.

logaritmická věrohodnostní funkce: $$\ell(\theta) = \sum_{i=1}^{n} \log f(X_i, Y_i) = \underbrace{\sum_{i=1}^{n} \log f_1(X_i, \theta_1)}_{\ell_1(\theta_1)} + \underbrace{\sum_{i=1}^{n} \log f_2(Y_i, \theta_2)}_{\ell_2(\theta_2)} + \underbrace{\sum_{i=1}^{n} \log c_{\theta_3}(F_1(X_i, \theta_1), F_2(Y_i, \theta_2))}_{\ell_3(\theta_1, \theta_2, \theta_3)}$$

maximálně věrohodný odhad parametru Θ je $\hat{\Theta} = \underset{\Theta}{\arg\max}\ \ell(\Theta)$.

<u>Poznámka</u>

podmínky regularity $\rightarrow$ as. nestrannost, konzistence a as. normalita $\hat{\Theta}$.

výpočetně náročné a nestabilní (dimenze Θ může být velká) $\rightsquigarrow$ modifikace metody.

pseudo-maximálně věrohodný odhad (IFM = inference pro marginální rozdělení)

- nejprve odhadneme parametry marg. rozdělení: $\hat{\Theta}_1 = \underset{\Theta_1}{\arg\max}\ \ell_1(\Theta_1)$ a $\hat{\Theta}_2 = \underset{\Theta_2}{\arg\max}\ \ell_2(\Theta_2)$.

- pro tyto pevné hodnoty $\hat{\Theta}_1$ a $\hat{\Theta}_2$ maximalizujeme $\ell_3(\hat{\Theta}_1, \hat{\Theta}_2, \Theta_3)$: $\hat{\Theta}_3 = \underset{\Theta_3}{\arg\max}\ \ell_3(\hat{\Theta}_1, \hat{\Theta}_2, \Theta_3)$.

- výsledný odhad parametru Θ je: $\hat{\Theta} = (\hat{\Theta}_1^T, \hat{\Theta}_2^T, \hat{\Theta}_3^T)^T$.

<u>semiparametrický přístup</u>

- neparametricky odhadneme marginální rozdělení pomocí empirické distribuční funkce: $\hat{F}_1(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{X_i \le x\}$, $\hat{F}_2(y) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}\{Y_i \le y\}$

- kopulu C modelujeme parametricky $C(u,v) = C_{\Theta_3}(u,v)$. Parametr Θ_3 odhadneme pomocí metody max. věrohodnosti:

$$\hat{\Theta}_3 = \underset{\Theta_3}{\arg\max} \sum_{i=1}^{n} \log c_{\Theta_3}\big(\overbrace{\hat{F}_1(X_i)}^{U_i}, \overbrace{\hat{F}_2(Y_i)}^{V_i}\big).$$

transformace $(X_1,Y_1)^T, \dots, (X_n,Y_n)^T \rightsquigarrow (U_1,V_1)^T, \dots, (U_n,V_n)^T$

$U_i = \frac{R_i}{n+1}$, R_i je pořadí X_i mezi $X_1, \dots, X_n$.

$V_i = \frac{S_i}{n+1}$, S_i je pořadí Y_i mezi $Y_1, \dots, Y_n$.

Poznámka

Předchozí metoda bývá označována CML - kanonická metoda maximální věrohodnosti

Pro θ_3 jednorozměrný parametr lze pro jeho odhad použít následující postup založený na odhadu Kendallova τ.

Minule jsme odvodili, že pro některé kopuly umíme vyjádřit $\tau = f(\theta_3)$ jako nějakou funkci neznámého parametru.

Z dat odhadneme Kendallovo τ jako
$$\hat{\tau} = \frac{\sum\sum_{i \neq j} \operatorname{sign}(X_i - X_j) \cdot \operatorname{sign}(Y_i - Y_j)}{n(n-1)} = \frac{2(a-b)}{n(n-1)},$$
a ... počet konkordantních párů $(x_i, y_i), (x_j, y_j)$

b ... počet diskordantních párů $(x_i, y_i), (x_j, y_j)$

$\hat{\theta}_3$ jakožto odhad θ_3 řeší rovnici $\hat{\tau} = f(\hat{\theta}_3)$.

Příklad

Gumbelova kopula s parametrem $\theta \geq 1$

$$\tau = 1 - \frac{1}{\theta} \Rightarrow \theta = \frac{1}{1-\tau} \rightarrow \hat{\theta} = \frac{1}{1-\hat{\tau}}. \quad (\text{pro } \hat{\tau} \geq 0)$$

Poznámka

Podobně můžeme definovat p-rozměrnou kopulu pro p-rozměrné náh. vektory.

Jak zvolit vhodnou kopulu? Maximalizace věrohodnosti, příp. penalizační kritéria.