

Souřadnicová reprezentace kompozičních tabulek

Kamila Fačevicová, Karel Hron

Univerzita Palackého v Olomouci

Kompoziční tabulky - definice

Kompoziční tabulky

$$\mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1} & \cdots & x_{IJ} \end{pmatrix}, \quad \text{pro } x_{ij} > 0, \quad \forall i, j.$$

představují

- zobecnění $I \cdot J$ -složkových vektorových kompozičních dat,
- spojitou analogii kontingenčních tabulek.

	Násilný tr. čin	Loupež	Vloupání	Krádež auta	Krádež
Centrum	146	229	467	243	1380
Předměstí	425	610	1078	691	3001
Okrajové části	193	183	399	251	2032

Tabulka 1: Struktura kriminálních činů, které byly spáchány v Houstonu v lednu roku 2014.

Kompoziční tabulky - definice

Kompoziční tabulky

$$\mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1} & \cdots & x_{IJ} \end{pmatrix}, \quad \text{pro } x_{ij} > 0, \quad \forall i, j.$$

představují

- zobecnění $I \cdot J$ -složkových kompozičních dat,
- spojitou analogii kontingenčních tabulek.

Vyjádření s pevným součtem (v proporcích, %, ...) získáme použitím operace uzávěru

$$\mathcal{C}(\mathbf{x}) = \begin{pmatrix} \frac{\kappa \cdot x_{11}}{\sum_{i,j=1}^{I,J} x_{ij}} & \cdots & \frac{\kappa \cdot x_{1J}}{\sum_{i,j=1}^{I,J} x_{ij}} \\ \vdots & \ddots & \vdots \\ \frac{\kappa \cdot x_{I1}}{\sum_{i,j=1}^{I,J} x_{ij}} & \cdots & \frac{\kappa \cdot x_{IJ}}{\sum_{i,j=1}^{I,J} x_{ij}} \end{pmatrix}.$$

Výběrovým prostorem je $I \cdot J$ -složkový simplex

$$\mathcal{S}^{IJ} = \left\{ \mathbf{x} = (x_{11}, \dots, x_{1J}, \dots, x_{IJ}) \mid x_{ij} > 0, \quad i = 1, \dots, I, j = 1, \dots, J; \sum_{i,j=1}^{I,J} x_{ij} = \kappa \right\}.$$

Základní operace

Stejně jako vektorová kompoziční data se kompoziční tabulky řídí Aitchisonovou geometrií.

- Perturbace:

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C} \begin{pmatrix} x_{11}y_{11} & \cdots & x_{1J}y_{1J} \\ \vdots & \ddots & \vdots \\ x_{I1}y_{I1} & \cdots & x_{IJ}y_{IJ} \end{pmatrix}$$

- Mocninná transformace:

$$\alpha \odot \mathbf{x} = \mathcal{C} \begin{pmatrix} x_{11}^\alpha & \cdots & x_{1J}^\alpha \\ \vdots & \ddots & \vdots \\ x_{I1}^\alpha & \cdots & x_{IJ}^\alpha \end{pmatrix}$$

- Skalární součin:

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{2IJ} \sum_{i,j} \sum_{k,l} \ln \frac{x_{ij}}{x_{kl}} \ln \frac{y_{ij}}{y_{kl}}$$

K analýze kompozičních tabulek proto nelze použít standardní metody a je potřeba je nejprve vyjádřit v ortonormálních souřadnicích.

Rozklad kompozičních tabulek

Pomocí projekcí mohou být kompoziční tabulky rozděleny na nezávislou a interakční část.

$$\mathbf{x} = \mathbf{x}_{ind} \oplus \mathbf{x}_{int}.$$

- **Nezávislá tabulka:**

$$\mathbf{x}_{ind} = \left(x_{ij}^{ind} = \left(\prod_{k=1}^I \prod_{l=1}^J x_{kj} x_{il} \right)^{\frac{1}{IJ}} \right)_{i,j=1}^{I,J}.$$

- **Interakční tabulka:**

$$\mathbf{x}_{int} = \left(x_{ij}^{int} = \left(\prod_{k=1}^I \prod_{l=1}^J \frac{x_{ij}}{x_{kj} x_{il}} \right)^{\frac{1}{IJ}} \right)_{i,j=1}^{I,J}.$$

Vyjádření v souřadnicích - standardní přístup

Pro vyjádření kompozičních dat v souřadnicích se nejčastěji využívá **postupného binárního dělení**:

	x_1	x_2	x_3	x_4	x_5	u	v
1	+	+	-	-	-	2	3
2	+	-				1	1
3			+	-	-	1	2
4				+	-	1	1

⇒ Získáme souřadnice ve tvaru **bilancí**

$$z_i = \langle \mathbf{x}, \mathbf{e}_i \rangle_A = \sqrt{\frac{uv}{u+v}} \ln \frac{(x_{j_1} x_{j_2} \dots x_{j_u})^{1/u}}{(x_{k_1} x_{k_2} \dots x_{k_v})^{1/v}}$$

Bilance patří mezi tzv. **ilr souřadnice**, pro které platí

$$\begin{aligned} \text{ilr}(\mathbf{x} \oplus \mathbf{y}) &= \text{ilr}(\mathbf{x}) + \text{ilr}(\mathbf{y}), & \text{ilr}(\alpha \odot \mathbf{x}) &= \alpha \cdot \text{ilr}(\mathbf{x}), \\ \langle \mathbf{x}, \mathbf{y} \rangle_A &= \langle \text{ilr}(\mathbf{x}), \text{ilr}(\mathbf{y}) \rangle \end{aligned}$$

Bilance však nerespektují

- dvourozměrnou povahu tabulek
- a možnost jejich rozkladu.

Navrhovaný souřadnicový systém

Nový systém je tvořen třemi skupinami souřadnic:

- Bilance mezi úrovněmi řádkového faktoru

$$z_i^r = \sqrt{\frac{stJ}{s+t}} \ln \frac{[g(x_{j_1.}) \cdots g(x_{j_s.})]^{1/s}}{[g(x_{k_1.}) \cdots g(x_{k_t.})]^{1/t}}, \quad \text{pro } i = 1, 2, \dots, I-1.$$

- Bilance mezi úrovněmi sloupcového faktoru

$$z_j^c = \sqrt{\frac{uvI}{u+v}} \ln \frac{[g(x_{.j_1}) \cdots g(x_{.j_u})]^{1/u}}{[g(x_{.k_1}) \cdots g(x_{.k_v})]^{1/v}}, \quad \text{pro } j = 1, 2, \dots, J-1.$$

- Odds ratio souřadnice zachycující vztah mezi faktory

$$z^{OR} = \sqrt{\frac{a \cdot d}{a+b+c+d}} \ln \frac{(x_{i_1} \cdots x_{i_a})^{1/a} (x_{l_1} \cdots x_{l_d})^{1/d}}{(x_{j_1} \cdots x_{j_b})^{1/b} (x_{k_1} \cdots x_{k_c})^{1/c}}.$$

Navrhovaný souřadnicový systém






Tento souřadnicový systém:

- Umožňuje **analýzu vztahu mezi dvěma faktory** na základě souboru kompozičních tabulek.
- Představuje **analogii standardní analýzy** prostřednictvím log-lineárních modelů.
- Umožňuje **rozšíření** teorie kompozičních dat na dvourozměrné kompoziční tabulky.
- Umožňuje **zobecnění** pro kompoziční data vyšší dimenze.

Co najdete na posteru

- Definici kompozičních tabulek jako spojitě analogie kontingenčních tabulek.
- Definici navrhovaného souřadnicového systému.
- Příklad ilustrující konstrukci souřadnic a jejich interpretaci.

Reference

-  Aitchison J (1986) *The statistical analysis of compositional data*. Chapman and Hall, London.
-  Egozcue JJ, Díaz-Barrero JL, Pawłowsky-Glahn V (2008) *Compositional analysis of bivariate discrete probabilities*. In Daunis-i-Estadella J, Martín-Fernández JA, eds, Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop. University of Girona, Spain.
-  Fačevicová K, Hron K, Todorov V, Guo D, Templ M (2014) *Logratio approach to statistical analysis of 2x2 compositional tables*. Journal of Applied Statistics, 41 (2014), 5, 944-958.
-  Fačevicová K, Hron K, Todorov V, Templ M (2014) *Compositional tables analysis in coordinates*. Scandinavian Journal of Statistics, DOI: 10.1111/sjos.12223.
-  Fačevicová K, Hron K, Todorov V, Templ M (2016) *General approach to coordinate representation of compositional tables*.