O hledání skrytých (ale ne hlavních) komponent

Kamila Fačevicová

Univerzita Palackého v Olomouci

(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem

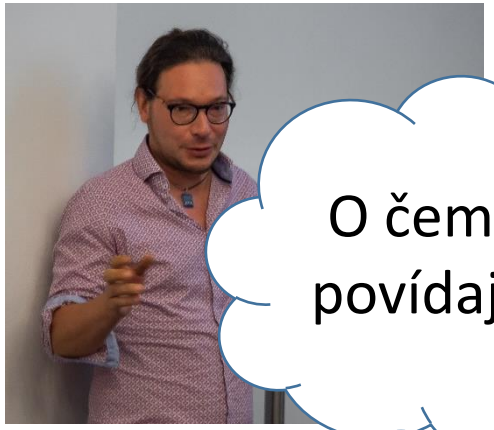
(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem

(Robustní) Cocktail Party Problem



O čem si asi povídají…???

Analýza nezávislých komponent (ICA)

Pozorovaný p-rozměrný vektor $\boldsymbol{x}$ se řídí modelem nezávislých komponent, pokud

$$\boldsymbol{x} = A\boldsymbol{z} + \boldsymbol{b},$$

kde A^{pxp} je regulární matice, $\boldsymbol{b} \in \mathbb{R}^p$ a $\boldsymbol{z}$ je p-rozměrný vektor skrytých proměnných splňující

- $E(\boldsymbol{z}) = 0$ a $\mathrm{cov}(\boldsymbol{z}) = I_p$,
- složky vektoru $\boldsymbol{z}$ jsou nezávislé a
- Normálním rozdělením se řídí nejvýše jedna z nich.

⇨ Cílem je nalézt matici W^{pxp} tak, aby vektor $\hat{\boldsymbol{z}} = W\boldsymbol{x}$ měl nezávislé složky s rozdělením co nejvíce odlišným od Normálního.

Hledaná matice je dána jednoznačně až na pořadí a znaménko řádků (→ komponent $\widehat{z_i}$).

Analýza nezávislých komponent (ICA)

Vyjděme ze standardizovaného vektoru

$$\boldsymbol{x}^{st} = \mathrm{cov}(\boldsymbol{x})^{-1/2}(\boldsymbol{x}-\mathrm{E}(\boldsymbol{x})).$$

Namísto W pak stačí nalézt ortogonální matici U takovou, že

$$U^{\mathrm{T}}\mathbf{x}^{st} = \mathbf{z} \quad \text{a} \quad W = U^{\mathrm{T}}\mathrm{cov}(\boldsymbol{x})^{-1/2}.$$

Analýza nezávislých komponent (ICA)

Metoda FOBI

- U_{FOBI} je matice vlastních vektorů $\text{cov}_4(\boldsymbol{x}^{st})$, kde

$$\text{cov}_4(\boldsymbol{x}) = \text{E}((\boldsymbol{x}-\text{E}(\boldsymbol{x}))^{\text{T}}\text{cov}(\boldsymbol{x})^{-1}\,(\boldsymbol{x}-\text{E}(\boldsymbol{x}))(\boldsymbol{x}-\text{E}(\boldsymbol{x}))(\boldsymbol{x}-\text{E}(\boldsymbol{x}))^{\text{T}}),$$

- a tedy $W_{FOBI} = U_{FOBI}^{\text{T}}\text{cov}(\boldsymbol{x})^{-1/2}$.
- Vlastní čísla $\text{cov}_4(\boldsymbol{x}^{st})$ pak reflektují špičatost komponent.

Metoda Deflation-based fastICA

- Postupně hledá vektory $\boldsymbol{u}_k$ tak, aby

$$|\text{E}[G(\boldsymbol{u}_k^T\boldsymbol{x}^{st})]| \rightarrow \max, \quad \boldsymbol{u}_k^T\boldsymbol{u}_k = 1 \quad \text{a} \quad \boldsymbol{u}_k^T\boldsymbol{u}_k = 0.$$

- Funkce $G(y)$ představuje míru nenormality a musí pro ni platit
  - $\text{E}(G(y)) = 0$ pokud $y \sim N(0, 1)$,
  - dvakrát diferencovatelná funkce.
- Typicky vycházíme z maximalizace špičatosti a $G(y) = (y^4 - 3)/4$.

Analýza nezávislých komponent (ICA)

Výstupem metody jsou

- matice koeficientů W – „zátěže“
- vektory skrytých komponent $\boldsymbol{z}_i$, $i = 1, \ldots, p$.

Pro přehlednost komponenty před prezentací uspořádáváme podle špičatosti,

→ první komponenty typicky zachycují odlehlá pozorování,

→ poslední komponenty reflektují přítomnost skupin.

Srovnání PCA a ICA

- Hledá směr maximální variability.
- Lze použít pro n < p.
- Může sloužit jako první krok před ICA.
- …

- Hledá směr maximální nenormality.
- Nelze použít pro n < p.
- Lze chápat jako další rotaci řešení PCA.

- Existuje ve variantách pro iid, časové řady, prostorově závislá data.
- Existují robustní varianty metod.
- …

Srovnání PCA a ICA

Srovnání PCA a ICA

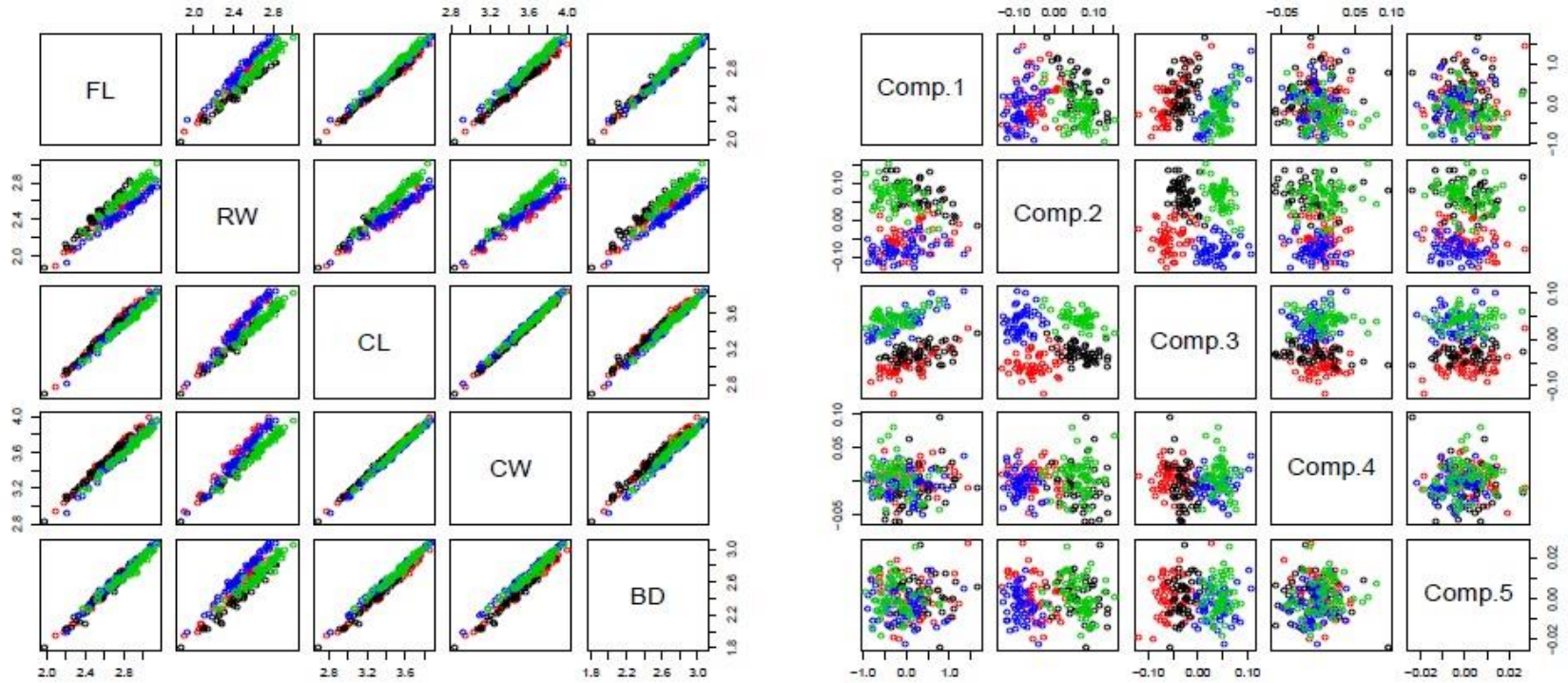


FIGURE – Crabs data with two species and two genders : Left panel original data, right panel principal components.

Srovnání PCA a ICA

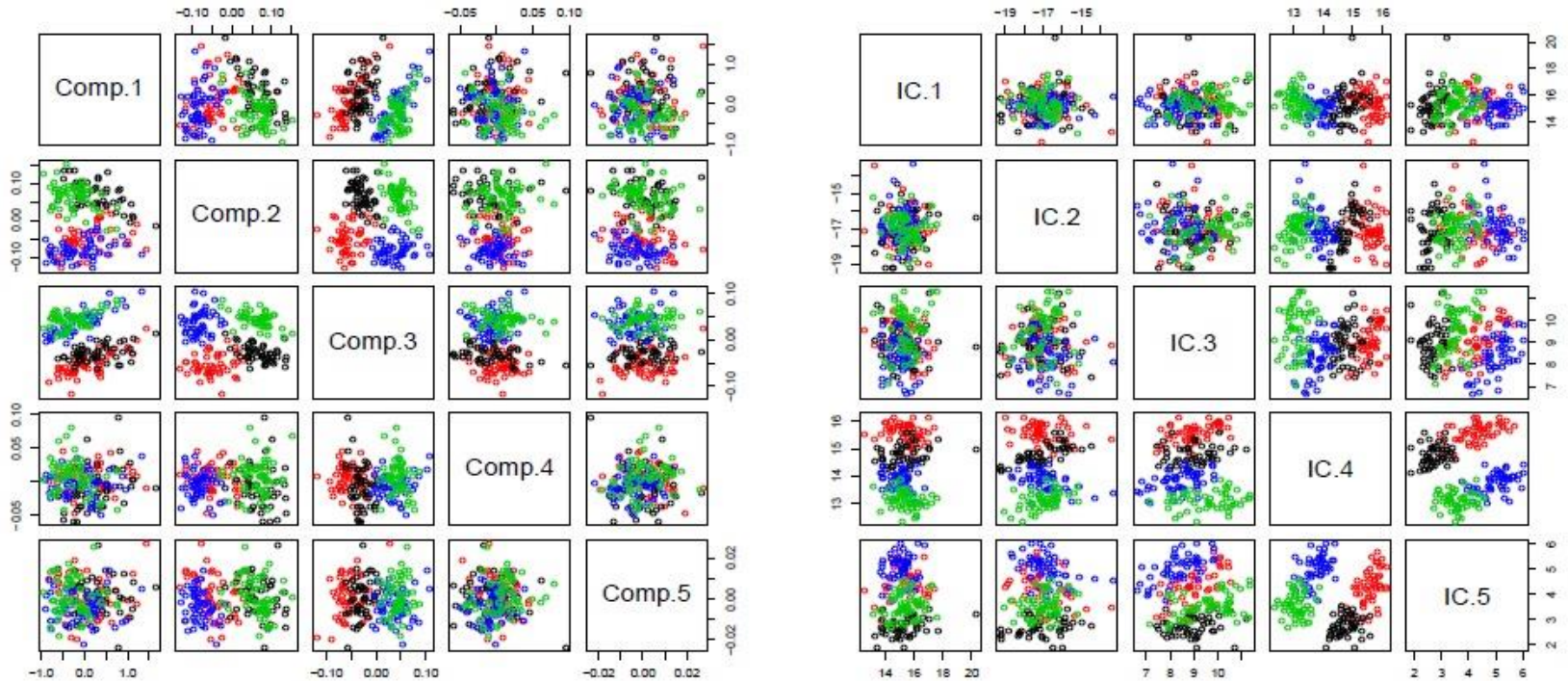


FIGURE – Crabs data with two species and two genders : Left panel principal components, right panel independent components.

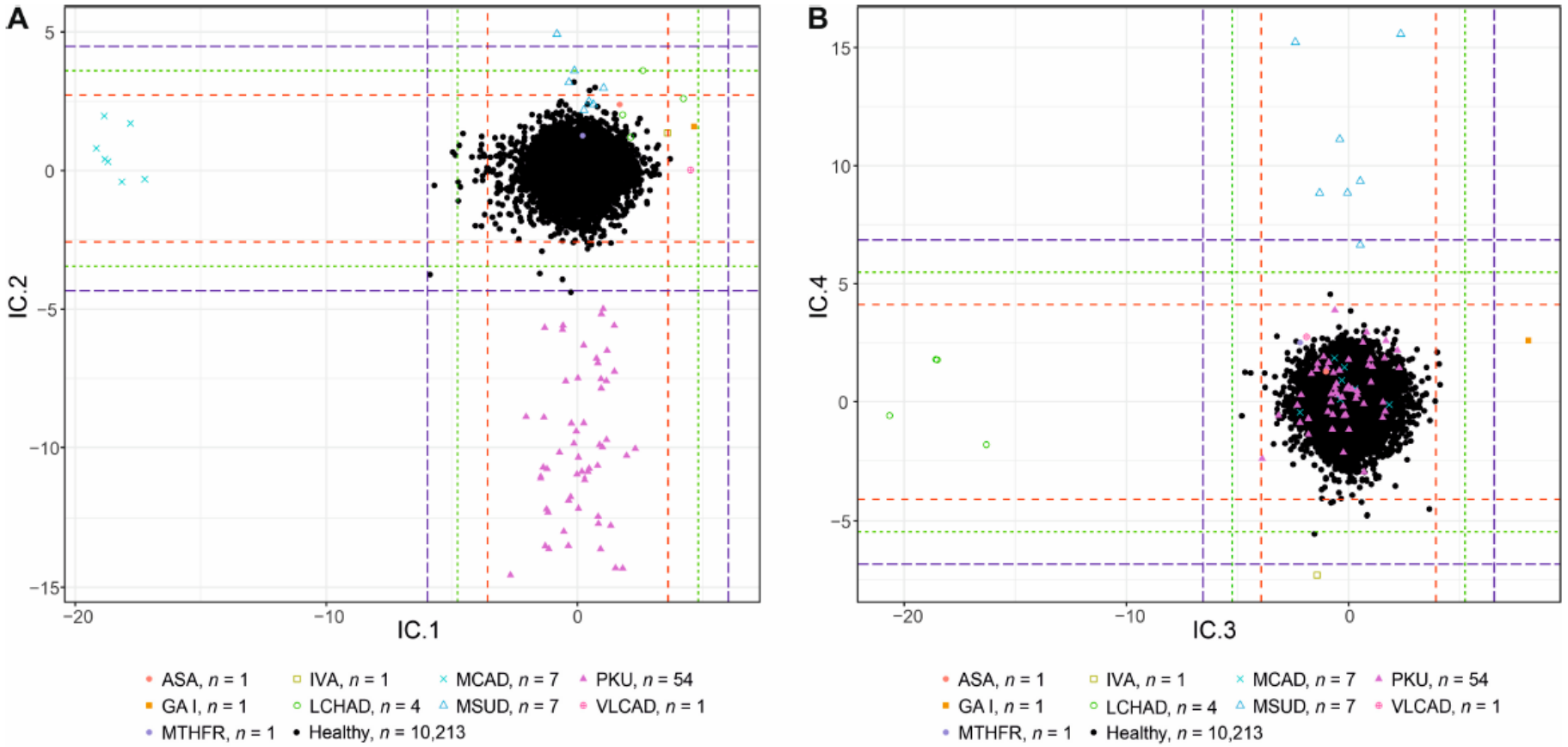
A na co jsme ICA použili my?

Novorozenecký screening (Kouřil et al, 2023)

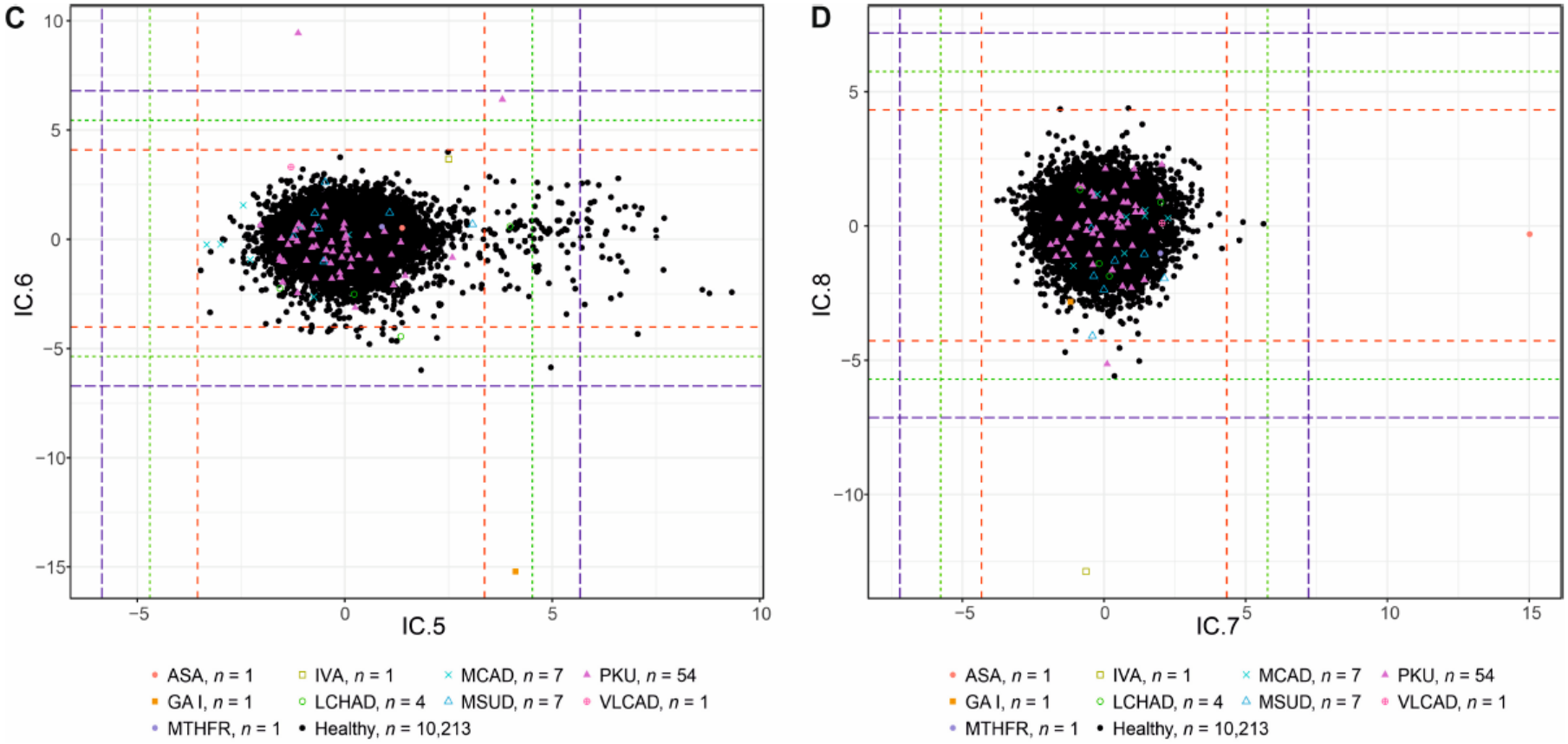
- Měření koncentrace 26 metabolitů v krvi 10 290 zdravých novorozenců. Z nichž 77 trpí metabolickou poruchou – trénovací fáze.
  Testovací fáze – 150 zdravých, 20 nemocných.
- Standardně je vyhodnocení provedeno skrze podíly koncentraci dvojic metabolitů.
- Aplikace kompoziční metodologie – ICA pro ilr transformovaná data.

→ Nalezené komponenty všechny nemocné detekovaly ve vzdálenosti ± 5 sd.

A na co jsme ICA použili my?



A na co jsme ICA použili my?



Pár užitečných (?) odkazů…

- Nordhausen, K, Oja, H. Independent component analysis: A statistical perspective. *WIREs Comput Stat*. 2018; 10:e1440.
- Miettinen, J, et al. Deflation-based FastICA with adaptive choices of nonlinearities. *IEEE Trans. Signal Process.* 62, 2014, 5716–5724.
- Scholz, M, et al. Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics, 20(15), 2004.
- Miettinen, J, Nordhausen, K, Taskinen, S. fICA: FastICA algorithms and their improved variants. *The R Journal*, 10, 2018, 148–158.
- Kouřil, Š, et al. Multivariate independent component analysis identifies patients in newborn screening equally to adjusted reference ranges. *International Journal of Neonatal Screening,* 9(4), 60 (2023).

A o čem že si povídají?

Joo, to jsem takhle jednou s mojí Karosou vezl zájezd do Chorvatska. To vám byla paráda…

WOW !!!

Dámy, už máte napečeno? Já mám zatím jen 25 druhů. Myslíte, že to bude stačit?

Kdybych aspoň nějaký měl…

Kluci tak co, těšíte se do Bardějova?