Dirichletovo rozdělení vzhledem k Aitchisonově míře na simplexu

Poster sekce

Petra Kynčlová

Department of Statistics and Probability Theory,
Vienna University of Technology

Robust 2012
9. – 14. září 2012, Němčičky

Kompoziční data

- D-složková kompozice je vektor $\mathbf{x} = (x_1, \dots, x_D)' \in \mathbb{R}^+$, jehož složky nesou výhradně relativní informaci.
- Kompozice můžeme reprezentovat jako data s konstantním součtem.
- Výběrový prostor reprezentací kompozic při zvoleném κ je **simplex**.
- Při práci s kompozicemi se používá **Aitchisonova geometrie na simplexu**, která je reprezentována operacemi perturbace, mocnění a Aitchisonův skalární součin.
- Zavedení Aitchisonovy geometrie na simplexu zaručuje existenci ortonormální báze $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$. **Izometrická logratio (ilr) transformace** kompozice $\mathbf{x}$ představuje souřadnice jakékoliv kompozice $\mathbf{x}$ vzhledem k ortonormální bázi $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}\}$.

Číselné charakteristiky na simplexu

- **Střed kompozice**, střední hodnota náhodné kompozice, je definována pomocí geometrické interpretace střední hodnoty náhodného vektoru. Představuje kompozici $\mathrm{cen}(\mathbf{x})$, která minimalizuje výraz $\mathrm{E}[d_a^2(\mathbf{x}, \mathrm{cen}(\mathbf{x}))]$, tj.

$$\boxed{\mathrm{cen}(\mathbf{x}) = \mathcal{C}(\exp(\mathrm{E}[\ln \mathbf{x}])).}$$

- **Metrický rozptyl** udává variabilitu náhodné kompozice tj.

$$\boxed{\mathrm{Mvar}[\mathbf{x}] = \mathrm{E}[d_a^2(\mathbf{x}, \mathrm{cen}[\mathbf{x}])].}$$

Aitchisonova míra

- Nechť je dán vektorový prostor E, na kterém je zaveden skalární součin. Zde můžeme zavést pravděpodobnostní míru λ_E, jež bude se strukturou prostoru E kompatibilní, a to prostřednictvím Lebesgueovy míry na ortonormálních souřadnicích. Funkce hustoty f_E, která je definována na E, je pak dána jako Radon-Nikodýmova derivace pravděpodobnostní míry P vzhledem k míře λ_E. Míra λ_E má v prostoru E stejné vlastnosti jako Lebesueova míra v reálném prostoru.
- **Aitchisonova míra** λ_a odpovídá geometrické struktuře simplexu, je relativní a absolutně spojitá vzhledem k Lebesgueově míře λ. Vztah mezi mírami λ_a a λ je dán pomocí Jakobiánu

$$\boxed{\frac{d\lambda_a}{d\lambda} = \frac{1}{\sqrt{D}x_1 \cdots x_D}.}$$

Dirichletovo rozdělení na simplexu

Dirichletovo rozdělení vzhledem k Lebesgueově míře

Definice

Náhodný vektor $\mathbf{X} \in \mathcal{S}^D$ *má D-rozměrné Dirichletovo rozdělení (vzhledem k Lebesgueově míře na simplexu) s parametrem* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)' \in \mathbb{R}_+^D$*, jestliže jeho hustota pravděpodobnosti má tvar*

$$f(\mathbf{x}) = \frac{dP}{d\lambda}(\mathbf{x}) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^D \Gamma(\alpha_i)} \prod_{i=1}^D x_i^{\alpha_i - 1},$$

kde λ *je Lebesgueova míra,* $\alpha_+ = \sum_{i=1}^D \alpha_i$*, a* Γ *je gamma funkce. Značíme* $\mathbf{X} \sim \mathcal{D}^D(\boldsymbol{\alpha})$*.*

Dirichletovo rozdělení na simplexu

Dirichletovo rozdělení vzhledem k Aitchisonově míře

Zároveň můžeme hustotu Dirichletova rozdělení vyjádřit vzhledem k Aitchisonově míře λ_a ve tvaru

$$\boxed{f_a(\mathbf{x}) = \frac{dP}{d\lambda_a}(\mathbf{x}) = \frac{\Gamma(\alpha_+)\sqrt{D}}{\prod_{i=1}^{D}\Gamma(\alpha_i)}\prod_{i=1}^{D} x_i^{\alpha_i}.}$$

Dirichletovo rozdělení na simplexu

Číselné charakteristiky Dirichletova rozdělení

- Modus a střední hodnota vzhledem k Lebesgueově míře

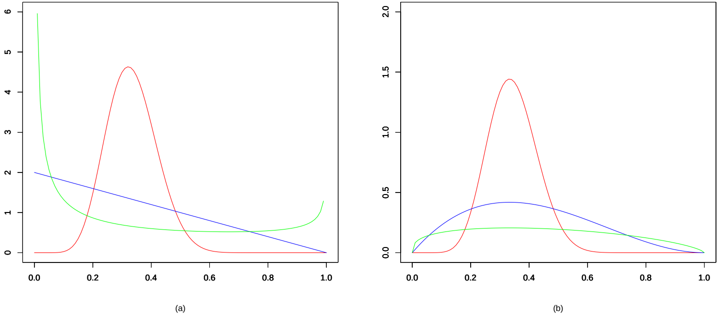
$$\boxed{\begin{aligned} \mathrm{E}(\mathbf{X}) &= \left(\frac{\alpha_1}{\alpha_+}, \frac{\alpha_2}{\alpha_+}, \ldots, \frac{\alpha_D}{\alpha_+}\right)', \\ \mathrm{modus}(\mathbf{X}) &= \left(\frac{\alpha_1 - 1}{\alpha_+ - D}, \frac{\alpha_2 - 1}{\alpha_+ - D}, \ldots, \frac{\alpha_D - 1}{\alpha_+ - D}\right)'. \end{aligned}}$$

- Modus a střední hodnota vzhledem k Aitchisonově míře

$$\boxed{\begin{aligned} \mathrm{modus}_a(\mathbf{X}) &= \left(\frac{\alpha_1}{\alpha_+}, \ldots, \frac{\alpha_D}{\alpha_+}\right)', \\ \mathrm{E}_a(\mathbf{X}) &= \mathcal{C}\left(e^{\psi(\alpha_1)}, \ldots, e^{\psi(\alpha_D)}\right)'. \end{aligned}}$$

Dirichletovo rozdělení na simplexu

$D = 2$



Obrázek 1. Hustoty Dirichletova rozdělení vzhledem (a) k Lebesgueově míře λ a (b) k Aitchisonově míře λ_a s parametry $\boldsymbol{\alpha} = (10, 20)'$ (——-), $\boldsymbol{\alpha} = (1, 2)'$ (——-) a $\boldsymbol{\alpha} = (1/3, 2/3)'$ (——-).

Použité zdroje I

- Aitchison, J.: *The statistical analysis of compositional data*, London: Chapman and Hall, 1986.
- Monti, G.S., Mateu-Figueras, G., Pawlowsky-Glahn, V., *Notes on the scaled Dirichlet distribution*, In: Pawlowsky-Glahn, V., Buccianti, A.: Compositional Data Analysis: Theory and Applications, Chichester: John Wiley & Sons, 128-138, 2011.
- Monti, G.S.,Mateu-Figueras, G.,Pawlowsky-Glahn, V., Egozcue, J.J., *The shifted-scaled Dirichlet distribution in the simplex*, In: Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I.: Compositional Data Analysis Workshop – CoDaWork'11, Proceedings, International Center for Numerical Methods in Engineering (CIMNE), Barcelona, 2011.